ETSI TS 101 377-6-1 V1.1.1 (2001-03)

Technical Specification

**GEO-Mobile Radio Interface Specifications;
Part 6: Speech coding specifications;
Sub-part 1: Basic Rate Speech;
Basic Rate Speech Processing Functions;
GMR-2 06.001**

Reference

DTS/SES-002-06001

Keywords

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

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

Contents

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 101 377 V1.1.1	Digital Voice Systems Inc		US	US 5,715,365	US
TS 101 377 V1.1.1	Digital Voice Systems Inc		US	US 5,754,974	US
TS 101 377 V1.1.1	Digital Voice Systems Inc		US	US 5,226,084	US
TS 101 377 V1.1.1	Digital Voice Systems Inc		US	US 5,701,390	US
TS 101 377 V1.1.1	Digital Voice Systems Inc		US	US 5,826,222	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 101 377 V1.1.1	Ericsson Mobile Communication	Improvements in, or in relation to, equalisers	GB	GB 2 215 567	GB
TS 101 377 V1.1.1	Ericsson Mobile Communication	Power Booster	GB	GB 2 251 768	GB
TS 101 377 V1.1.1	Ericsson Mobile Communication	Receiver Gain	GB	GB 2 233 846	GB
TS 101 377 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 101 377 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 101 377 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 101 377 V1.1.1	Lockheed Martin Global Telecommunic. Inc	Cellular Spacecraft TDMA Communications System with Call Interrupt Coding System for Maximizing Traffic ThroughputCellular Spacecraft TDMA Communications System with Call Interrupt Coding System for Maximizing Traffic Throughput	US	US 5,717,686	US
TS 101 377 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 101 377 V1.1.1	Lockheed Martin Global Telecommunic. Inc	Spacecraft Cellular Communication System	US	US 5,974,314	US
TS 101 377 V1.1.1	Lockheed Martin Global Telecommunic. Inc	Spacecraft Cellular Communication System	US	US 5,974,315	US
TS 101 377 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 101 377 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 1 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: "Basic Rate Speech; Basic Rate Speech Processing Functions; GMR-2 06.001".

Introduction

GMR stands for GEO (Geostationary Earth Orbit) Mobile Radio interface, which is used for mobile satellite services (MSS) utilising 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-n 01.201.

1 Scope

The present document provides an overview of the speech processing requirements applicable to the basic rate channel of the GMR-2 system. The speech processing functions in the GMR-2 system include the following:

- speech transcoding, which includes a speech encoder that converts digitized speech samples into a compressed binary bit stream and a speech decoder that converts a compressed binary bit stream into digital speech samples;
- discontinuous transmission (DTX), which is used to reduce the transmission rate during periods of voice inactivity;
- VAD, which is used to identify periods of voice activity, as required by DTX;
- CNI, which is used to convey the characteristics of the background noise from the transmit end to the receive end of the connection, in an effort to reduce the modulation of background noise that would otherwise occur with DTX;
- lost speech frame substitution and muting, which is used to mask transmission errors and stolen frames.

Detection and regeneration of single-frequency and dual-tone multifrequency (DTMF) signals.

The high level description given in the present document relates to the Digital Voice Systems, Inc.'s (DVSI's) AMBE™ 3 600 bps voice coder/decoder [7].

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-2 01.004 (TS 101 377-01-01): "GEO-Mobile Radio Interface Specifications; Abbreviations and Acronyms".

[2] GMR-2 03.050 (TS 101 377-03-13): "GEO-Mobile Radio Interface Specifications; Transmission planning aspects of the speech services in the Public Satellite Mobile Network (PSMN) system".

[3] GMR-2 05.003 (TS 101 377-05-03): "GEO-Mobile Radio Interface Specifications; Channel Coding".

[4] GMR-2 05.005 (TS 101 377-05-05): "GEO-Mobile Radio Interface Specifications; Radio Transmission and Reception".

[5] GMR-2 05.008 (TS 101 377-05-06): "GEO-Mobile Radio Interface Specifications; Radio Subsystem Link Control".

[6] ITU-T Recommendation G.711: "Pulse Code Modulation (PCM) of voice frequencies".

[7] Digital Voice Systems, Inc. http://www.dvsinc.com/

3 Definitions and abbreviations

3.1 Definitions

For the purpose of the present document, the following definitions apply:

Frame: time interval of 20 ms corresponding to the time segmentation of the basic rate speech transcoder, also used as a short term for a traffic frame

Traffic Frame: block of 72 information bits transmitted on the basic rate speech traffic channel

SID Frame: frame characterized by the SID (Silence Descriptor) code word. It conveys information on the acoustic background noise

SID Code Word: fixed bit pattern, for labelling traffic frame as a SID frame

SID Field: the bit position of the SID code word within a SID frame

Speech Frame: traffic frame that cannot be classified as a SID frame

Bad Traffic Frame: traffic frame flagged BFI = 1 (Bad Frame Indication) by the Radio Subsystem

Good Traffic Frame: traffic frame flagged BFI = 0 by the Radio Subsystem

Good Speech Frame: good traffic frame which is not an accepted SID frame

Valid SID Frame: good traffic frame flagged with SID = 1 by the DTX handler. This frame is valid for updating of comfort noise parameters at any time

Unusable Frame: bad traffic frame that is not an accepted SID frame

Lost SID Frame: unusable frame received when the RX DTX handler is generating comfort noise and a SID frame is expected (Time Alignment Flag, TAF = 1)

Lost Speech Frame: unusable frame received when the RX DTX handler is passing on traffic frames directly to the speech decoder

VAD Flag: boolean flag, generated by the VAD algorithm, indicating the presence (VAD flag = 1) or absence (VAD flag = 0) of voice activity

SP Flag: boolean flag, generated by the TX DTX handler, indicating whether the current frame of data should be transmitted by the RSS

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

BFI	Bad Frame Indication
DTX	Discontinuous Transmission
DVSI	Digital Voice Systems Inc.
GSM	Global System for Mobile communications
MES	Mobile Earth Station
PCM	Pulse Code Modulated
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
RF	Radio Frequency
RX	Receive
RSS	Radio SubSystem
SACCH	Slow Associated Control Channel
SID	Silence Descriptor
SP flag	SPeech flag
TAF	Time Alignment Flag

TX Transmit
VAD Voice Activity Detector

For abbreviations not given in this clause see GMR-2 01.004 [1].

4 General

Figure 4-1 presents a reference configuration where the various speech-processing functions are identified. In figure 4-1, the audio parts including analogue to digital and digital to analogue conversion are included to show the complete speech path between the audio input/output in the Mobile Earth Station (MES) and the digital interface of the PSTN. The detailed specification of the audio parts is considered in GMR-2 03.050 [2]. These aspects are only considered to the extent that the performance of the audio parts affects the performance of the speech transcoder.

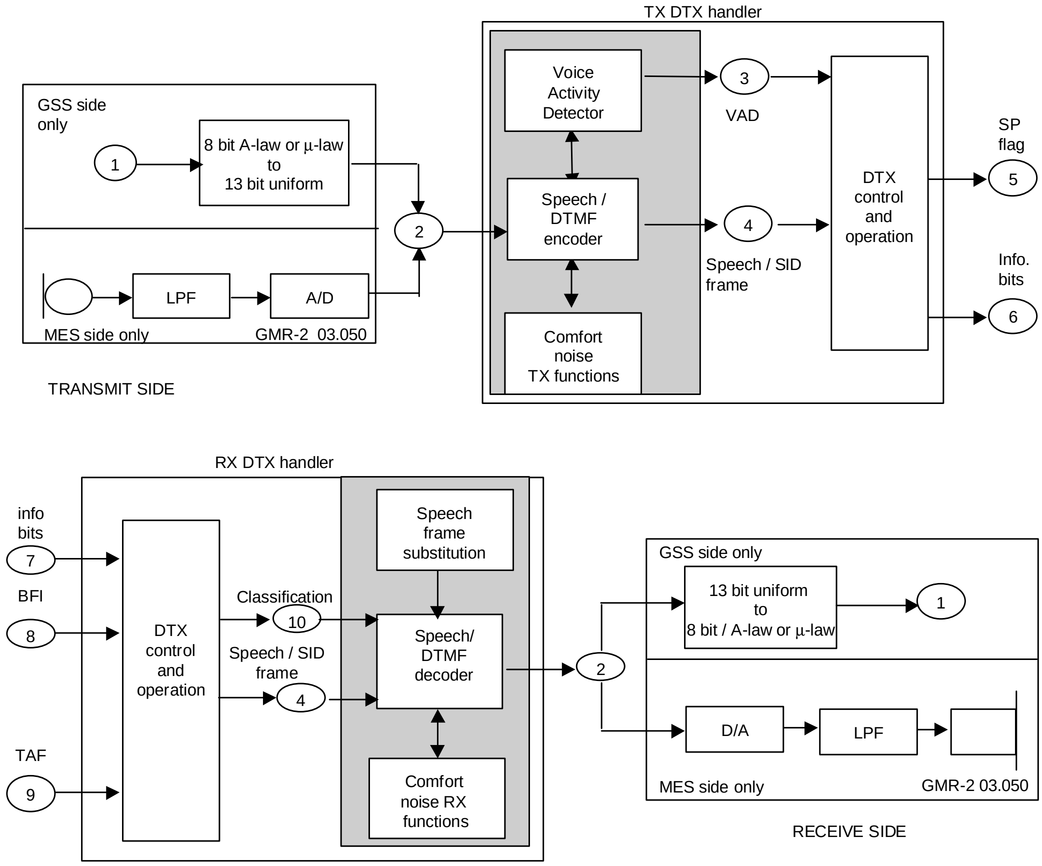

Figure 4-1: Reference Configuration

1) 8 bit A-law or µ-law PCM (ITU-T Recommendation G.711 [6]), 8 000 samples/s;
2) 13 bit uniform PCM, 8 000 samples/s;
3) Voice Activity Detector (VAD) flag;
4) Encoded speech / SID (Silence Descriptor) frame, 50 frames/s, 72 bits/frame;
5) SPeech (SP) flag, indicates whether information bits are speech or SID information;
6) Information bits delivered to the radio subsystem (72 bits/frame);
7) Information bits received from the radio subsystem (72 bits/frame);
8) Bad Frame Indication (BFI) flag;

9) Time Alignment Flag (TAF), marks the position of the SID frame within the Slow Associated Control CHannel (SACCH) cycle;

10) Frame classification flag.

5 Basic rate speech transcoding

As shown in figure 4-1, the speech encoder takes as its input a 13 bit uniform Pulse Code Modulated (PCM) signal. The PCM signals are from the audio part of the MES or, on the network side, from the Public Switched Telephone Network (PSTN) via an 8 bit/A-law or µ-law to 13-bit uniform PCM conversion. The encoder outputs 72 bits to the DTX every frame, regardless of the presence of voice. This data block contains either encoded speech, DTMF tones, or a characterization of the background noise (see clause 8.1). The encoded speech at the output of the speech encoder is delivered to the channel coding function as defined in GMR-2 05.003 [3]. The coding function produces an encoded block consisting of 120 bits that fills one complete transmission burst (exclusive of interleaving), leading to a gross bit rate of 6 kbits/s.

In the RX direction, the inverse operations take place, although the decoder processing is driven by the classification flag provided by DTX.

Input blocks of 160 speech samples in 13 bit uniform PCM format are mapped into encoded blocks of 72 bits, and these blocks are then mapped into output blocks of 160 reconstructed speech samples. The sampling rate is 8 000 sample/s leading to an average bit rate for the uncoded bit stream of 3,6 kbits/s. The coding scheme is called Advanced Multiband Excitation (AMBE™) coding.

The vocoder speech model is based on a robust speech model which is referred to as the Multi-Band Excitation (MBE) speech model. The basic methodology of the coder is to first divide a digital speech input signal into overlapping speech segment (or frames). Each segment of speech is then analysed in the context of the underlying speech model and a set of model parameters are established for that particular frame. The encoder quantizes these model parameters and transmits a bit stream at 3,6 kbits/s. The decoder receives this bit stream, reconstructs the model parameters and uses these model parameters to generate a synthetic speech signal. This synthesized speech signal is the output of the MBE speech coder as shown in figure 5-1.

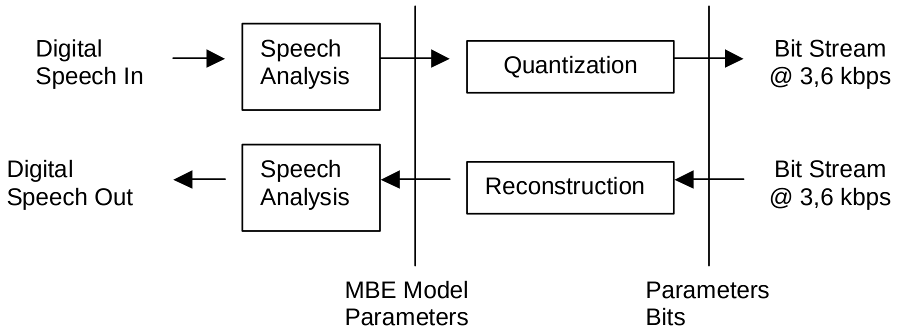

Figure 5-1: Block Diagram of the Multi-Band Excitation (MBE) Speech Coder

One defining characteristic of the speech coder is that it is a model-based coder, or vocoder, which does not try to reproduce the input speech signal on a sample-by-sample basis. Instead, the vocoder constructs a synthetic speech signal that contains the same perceptual information as the original speech signal. By using a robust speech model and sophisticated parameter estimation algorithms, the vocoder is able to achieve a low data rate while maintaining most of the quality, intelligibility and speaker recognizability found in the original speech signal.

The encoder is divided into two blocks: speech analysis and parameter quantization. Similarly, the decoder is divided into parameter reconstruction and speech synthesis. Once a digital speech signal has been normalized to the correct input range, the first step performed by the encoder is speech analysis. This step involves dividing the input signal into overlapping frames using an analysis window. For each 20 ms frame, an analysis algorithm estimates a set of model parameters consisting of a fundamental frequency (inverse of the pitch), a set of voiced/unvoiced (V/UV) decisions and a set of spectral amplitudes. These parameters fully describe the speech signal and are passed to the encoder's quantization block for further processing.

The encoder quantizes each frame of model parameters using 72 bits, yielding a bit rate to speech data of 3,6 kbit/s. These bits are apportioned to the different parameters in a manner which has been found to provide high fidelity over a wide range of speech conditions. First, bits are allocated to the fundamental frequency and the V/UV decisions, leaving all remaining bits for the spectral amplitudes. The spectral amplitude quantizer employed by the vocoder combines logarithmic companding, spectral prediction, Discrete Cosine Transforms and scalar quantization to achieve high efficiency, measured in terms of fidelity per bit, with relatively low complexity (MIPS + memory).

The corresponding decoder is designed to reproduce high quality speech from this 3,6 kbit/s bit stream. The decoder uses the received bits to reassemble each parameter frame of 72 bits which is then reconstructed into the model parameters for that particular frame. The reconstructed parameters form the input to the decoder's speech synthesis algorithm which interpolates successive frames of model parameters into smooth 20 ms segments of speech. The synthesis algorithm uses a set of harmonic oscillators to synthesize the voiced speech combined with a weighted overlap-add algorithm to synthesize the unvoiced speech.

To ensure that the voice codec operates at its maximum capability a set of input/output requirements for the analogue front end of a voice codec have been established. These performance recommendations include the gain, filtering and conversion elements as depicted in figure 5-2.

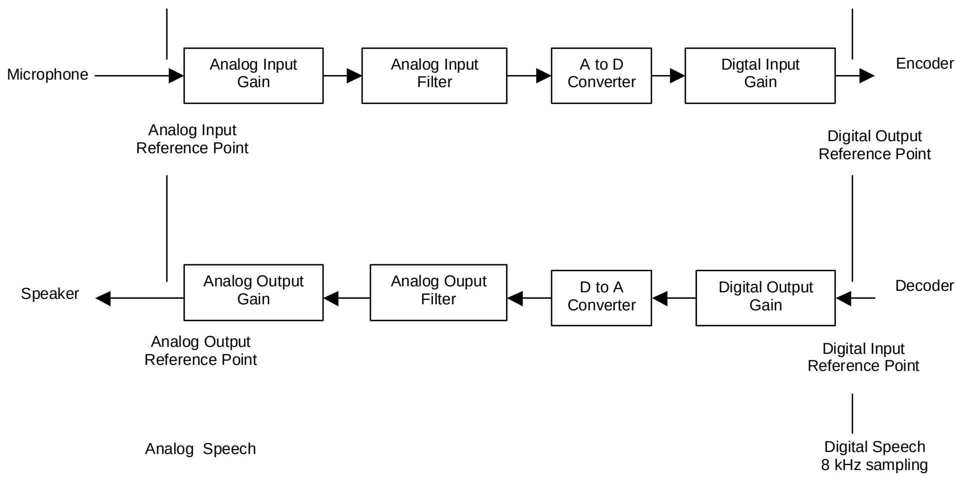

Figure 5-2: Block Diagram of the analogue Front End of a Voice Codec

It is recommended that the analogue input gain be set such that the RMS speech level under nominal input conditions is 24 dB below the saturation point of the A-to-D converter (+3 dBm0). This level, which equates to –22 dBm0, is designed to provide sufficient margin to prevent the peaks of the speech waveform from being clipped by the A-to-D converter.

The voice encoder requires the A-to-D and D-to-A converters to operate at an 8 kHz sampling rate (i.e. a sampling period of 125 microseconds) at the digital input/output reference points. This requirement necessitates the use of analogue filters at both the input and output to eliminate any frequency components above the Nyquist frequency (4 kHz).

This vocoder description assumes that the A-to-D converter produces digital samples where the maximum digital input level (+3 dBm0) is defined to be ±32 767. If a converter is used which does not meet these assumptions then the digital gain elements should be adjusted appropriately. Note that these assumptions are automatically satisfied if 16 bit linear A-to-D and D-to-A converters are used, in which case the digital gain elements should be set to unity gain. Also, note that the vocoder requires that any companding which is applied by the A-to-D converter (i.e., A-law or μ-law) should be removed prior to speech encoding. Similarly, any companding used by the D-to-A converter must be applied after speech decoding.

A final analogue recommendation addresses the maximum noise level measured at the reference points. It is recommended that the noise level for both directions should not exceed –60 dBm0 with no corresponding input.

6 Basic Rate Discontinuous Transmission (DTX)

6.1 General description

During a normal conversation, the participants alternate, such that on the average, each direction of transmission is occupied about 50 % of the time. Discontinuous Transmission (DTX) is a mode of operation where the transmitters are switched on only for those frames which contain useful information. This may be done for the following three purposes:

- in the MES, battery life will be prolonged or a smaller battery could be used for a given operational duration;
- the average interference level over the air interface is reduced, leading to better Radio Frequency (RF) spectrum efficiency;
- spacecraft power utilization on the L-band forward link is optimized.

The overall DTX mechanism is implemented in the DTX handlers (Transmit (TX) and Receive (RX) and requires the following functions:

1) a Voice Activity Detector (VAD) on the TX side;
2) evaluation of the background acoustic noise on the TX side, in order to transmit characteristic parameters to the RX side;
3) generation of comfort noise on the RX side during periods where the radio transmission is turned off.

In addition to these functions, if the parameters arriving at the RX side are detected to be corrupted by errors, the speech or comfort noise shall be generated from substituted data in order to avoid sound defects for the listener.

The transmission of comfort noise information to the RX side is achieved by means of a Silence Descriptor (SID) frame. The SID frame is transmitted at the end of speech bursts and serves as an end of speech marker for the RX side. In order to update the comfort noise characteristics at the RX side, SID frames are transmitted at regular intervals also during speech pauses. Transmissions of the SID frames are at a much lower rate than speech traffic rate, leading to minimal additional overhead. This also serves the purpose of improving the measurement of the radio link quality by the Radio Subsystem (RSS).

The DTX handlers interwork with the RSS using flags. The RSS is controlled by the transmitter keying on the TX side, which performs pre-processing functions on the RX side.

The speech flag (SP) indicates whether the information bits (voice/noise/or tones) are to be transmitted over the link. The SP flag is calculated from the VAD flag by the TX DTX handler.

6.2 Transmit (TX) Side

A block diagram of the TX side DTX functions is shown in figure 6.2-1.

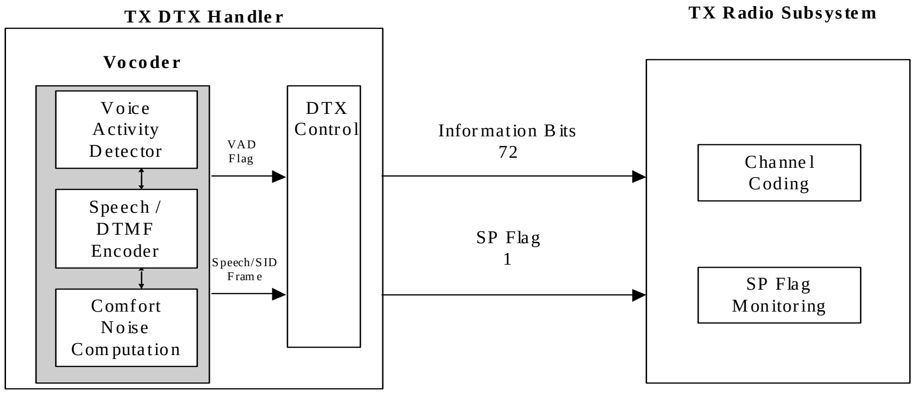

Figure 6.2-1: Block diagram of the Transmit Side DTX functions

The TX DTX handler continuously passes traffic frames, individually marked by the SP flag, to the radio subsystem (RSS). The scheduling of the frames for transmission on the air interface is controlled by the RSS alone, on the basis of the SP flag as described in clause 6.2.1. Note that the GSM SP flag indicated "Speech"; here the definition has been broadened to indicate which should be transmitted, regardless of content (i.e., speech or SID).

6.2.1 Function of the TX DTX handler

The VAD shall operate continuously in order to assess whether the input signal contains speech or not. The output is a binary flag (VAD flag = 1 or VAD flag = 0, respectively) on a frame-by-frame basis.

Regardless of the state of the VAD flag, the DTX handler shall pass all frames to the RSS.

The VAD flag controls indirectly, via the TX DTX handler operations described below, the overall DTX operation on the TX side. During normal speech segments (VAD = 1), the frames are passed to the RSS and transmitted over the link. At the end of a speech burst (transition from VAD flag = 1 to VAD flag = 0) a new updated SID frame is immediately available. The first two frames are transmitted as an end of speech marker to signal the decoder that a silence period will be ensuing. (Two frames are sent to increase the probability that at least one frame gets through to the decoder.) Subsequent SID frames are made available to the RSS every frame, but are only transmitted at a reduced duty cycle.

To accomplish this process, the DTX handler shall compute the SP flag, and pass it to the RSS every frame as follows (see figure 6.2-2). The value of the SP flag shall be set to 1 for frames meeting the following conditions:

a) those marked with VAD flag = 1;

b) the first two with VAD flag = 0, after one or more frames with VAD flag = 1;

c) those marked with VAD flag = 0 and aligned with the Satellite-Slow Associated Control Channel (S-SACCH) cycle as described in GMR-2 05.008 [5].

All other frames shall be marked with SP flag = 0.

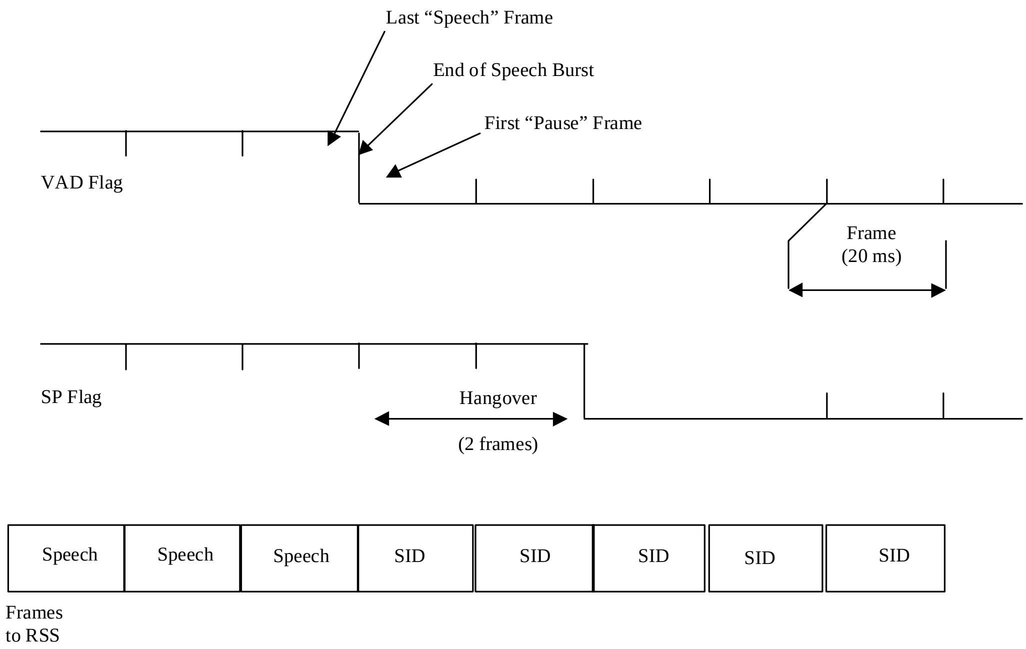

Figure 6.2-2: VAD and SP Flag Relationship at the End of a Speech Burst

6.2.2 Functions of the TX radio sub-system

All traffic frames marked with SP flag = 1 shall be scheduled for transmission. This has the overall function that the radio transmission is turned off after the transmission of a pair of SID frames when the speaker stops talking. During speech pauses the transmission is resumed at regular intervals for transmission of one SID frame, in order to update the generated comfort noise on the RX side (and to improve the measurement of the link quality by the RSS).

If a SID frame scheduled for transmission is stolen for Satellite-Fast Associated Control Channel (S-FACCH) signalling purposes, then the subsequent frame shall be scheduled for transmission instead.

6.3 Receive (RX) Side

A block diagram of the RX side DTX functions is shown in figure 6.3-1.

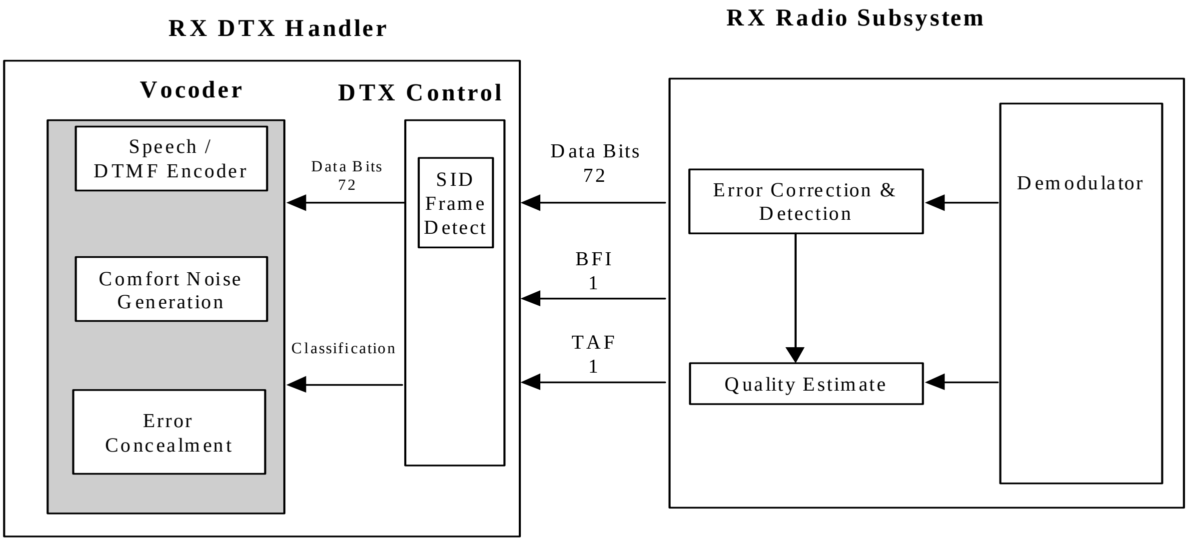

Figure 6.3-1: Block Diagram of the Receive Side DTX Functions

Whatever their context (speech, SID, S-FACCH or none), the RSS continuously passes the received traffic frames to the RX DTX handler, individually marked by various pre-processing functions with two flags. These are the BFI flag and the TAF described in clause 6.3.1 and table 6.3-1, which serve to classify the traffic frame according to the list of terms defined in clause 3.1. These flags, in conjunction with the SID flag which is defined in clause 6.3.2 and computed locally by the RX DTX control process, allow the RX DTX handler to determine how the received frame is to be handled.

6.3.1 Functions of the RX radio subsystem

The binary BFI flag (see GMR-2 05.005 [4]) indicates whether the traffic frame is considered to contain meaningful information bits (BFI flag = 0) or not (BFI flag = 1). In the context of the present document, an S-FACCH frame is considered not to contain meaningful bits and shall be marked with BFI flag = 1. The BFI flag shall fulfil the performance requirements of GMR-2 05.005 [4].

The binary Time Alignment Flag marks with TAF = 1 those traffic frames that are aligned with the S-SACCH cycle as described in GMR-2 05.008 [5].

6.3.2 Functions of the RX DTX handler

The RX DTX handler shall be responsible for the overall DTX operation on the RX side.

Received data frames shall be passed directly to the speech decoder, regardless of the BFI flag.

DTX shall perform SID frame detection. This function has been moved from the RSS to ensure that the RSS is independent of speech frame format. The SID frame detector compares, bit by bit, the relevant bits of the received traffic frame (the SID field) and gives back the binary SID flag. This flag is used internal to the DTX handler so that it knows the current state of the speech decoder.

DTX shall compute the 4-level classification flag (VALID, REPEAT, CNI, MUTE) as per table 6.3-1. For the purposes of table 6.3-1, the following abbreviations are used:

a) BSFC (bad speech frame count) = number of consecutive bad speech frames including this frame;

b) BCNFC (bad comfort noise frame count) = number of consecutive bad SID frames including this frame. Note that SID frames are transmitted at a reduced duty cycle as per GMR-2 05.008 [5], so this counter is only updated when a BFI occurs for a frame where a SID is expected (TAF=1).

Table 6.3-1: Classification of Traffic Frames

	Current State (based upon SID flag of last valid frame)	
BFI	**Speech**	**Comfort Noise**
0	set Classification = VALID reset lost frame counters update current state from SID detect	set Classification = VALID reset lost frame counters update current state based upon SID

1	increment BSFC
	If (BSFC < 3) Class = REPEAT
	ELSEIF (4<BSFC<100) Class = CNI
	ELSEIF (BSFC> 100) Class = MUTE
	END
		IF (TAF = 1) increment BCNFC
		If (BCNFC <2) Class = CNI
		ELSEIF (BCNFC > 3) Class=MUTE
		END

The net effect of the classification flag in table 6.3-1 is as follows:

1) whenever a good speech frame or SID frame is detected, the DTX handler shall pass it directly on to the speech decoder. and the decoder will synthesize voice/noise accordingly. The decoder will also store this frame for use in subsequent comfort noise insertion or frame replacement.

2) when only a few (<4) speech frames in a row are lost, the decoder replaces it with a stored copy of the last valid speech frame. If more than this, but less than about 2 seconds of data are lost, the decoder replaces the frame with a stored comfort noise estimate. If greater than about 2 seconds of data are lost, then the output is completely muted to signal that the link has been lost.

3) lost SID frames spanning less than about a 2 second window (assuming 1 second between SID updates; refer to GMR-2 05.008 [5] for exact values) result in continued comfort noise generation based upon the last stored set of noise parameters. Outages greater than this value result in muted output to signal loss of link.

7 Basic rate Voice Activity Detection (VAD)

The VAD flag is computed and passed to the DTX handler every frame. The input to the VAD is a set of parameters computed by the basic rate speech encoder. The VAD uses this information to decide whether each 20 ms speech coder frame contains speech or not. DTX then computes the SP flag, which is used by the Radio Subsystem to control the transmitter keying.

NOTE: The VAD flag is an Input to TX DTX handler and does not control the transmitter keying directly. For the purposes of the present document, DTMF tones are considered to be "speech" (i.e., VAD indicates "voice" during tone transmissions).

8 Basic rate comfort noise insertion

When switching the transmission on and off during DTX operation, the effect would be a modulation of the background noise at the receiving end, if no precautions were taken. When transmission is on, the background noise is transmitted together with the speech to the receiving end. As the speech burst ends, the connection is off and the perceived noise would drop to a very low level. This step modulation of noise may be perceived as annoying and reduce the intelligibility of speech if presented to a listener without modification.

This "noise contrast effect" is reduced in the GMR-2 system by inserting artificial noise, termed comfort noise, at the receiving end when speech is absent. The comfort noise processes are as described in the following clauses.

8.1 Transmit functions

The encoder always outputs a frame of data every speech frame, and presents it to the DTX control process along with a VAD flag. Comfort noise processing on the transmit side is a continuous background process during speech periods. Whenever the VAD flag indicates no speech, the 72-bit data block contains a characterization of the background noise rather than speech. The encoder computes only those parameters necessary for the characterization of the background noise, and loads them into the traffic frame in their normal location. All of the other parameters are unneeded and are set to default values. These dummy bits distinguish the SID frame from a normal speech frame, and are known as the SID codeword. The SID frames are transmitted on a reduced duty cycle. When the VAD flag indicates speech is present, normal speech processing functions resume.

8.2 Receive functions

The situations under which comfort noise shall be generated in the receiver may be started or updated whenever a valid SID frame is received. The decoder then utilizes the relevant parameters of the SID frame to compute the comfort noise. When updating the comfort noise parameters, these parameters shall be interpolated over the SID update period to obtain smooth transitions.

When instructed to do so by the classification flag provided by DTX, the decoder ignores the current frame of data and uses these stored parameters to generate comfort noise. Comfort noise-processing ceases immediately upon receipt of a valid speech frame. Normal speech processing then resumes.

9 Basic rate lost speech frame substitution and muting

In the receiver, frames may be lost due to transmission errors or frame stealing. The BFI flag is used by DTX to compute a classification flag that is in turn passed to the speech decoder to define the processing that is to be employed for this frame. The classification flag takes on one of the following values:

1) VALID, indicating that the 72 bit data block is valid (either speech or noise), and that the decoder should respond to the contents of the frame. The frame must also be stored for use in concealment of speech errors (see 2)) or in comfort noise processing (see 3));

2) REPEAT, indicating that the current frame is invalid, but that the decoder should replay the last stored speech frame in memory to conceal the effects of the errors in this particular speech frame;

3) CNI, indicating that the current frame is invalid, and that it should be replaced with the noise frame stored in the decoder memory. This is used for silence periods, or for moderate duration (e.g., a few seconds) dropouts of speech;

4) MUTE, indicating that the current frame is invalid and that the output should be muted for this frame to alert the user that the link was probably lost.

History

Document history		
V1.1.1	March 2001	Publication