Institutionen för systemteknik
Department of Electrical Engineering

Examensarbete

Data Transmission over Speech Coded Voice Channels

Examensarbete utfört i Reglerteknik
vid Tekniska högskolan i Linköping
av

Andreas Tyrberg

LITH-ISY-EX--06/3843--SE

Linköping 2006

Linköpings universitet
TEKNISKA HÖGSKOLAN

Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

Linköpings tekniska högskola
Linköpings universitet
581 83 Linköping

Data Transmission over Speech Coded Voice Channels

Examensarbete utfört i Reglerteknik
vid Tekniska högskolan i Linköping
av

Andreas Tyrberg

LITH-ISY-EX--06/3843--SE

Handledare: **Janne Harju**
ISY, Linköpings universitet
Robin von Post
Sectra Communications AB
Peter Nyman
Sectra Communications AB

Examinator: **Fredrik Gustafsson**
ISY, Linköpings universitet

Linköping, 12 June, 2006

Avdelning, Institution
Division, Department

Division of Automatic Control
Department of Electrical Engineering
Linköpings universitet
S-581 83 Linköping, Sweden

Datum
Date

2006-06-12

Språk
Language

☐ Svenska/Swedish
☒ Engelska/English
☐ _____

Rapporttyp
Report category

☐ Licentiatavhandling
☒ Examensarbete
☐ C-uppsats
☐ D-uppsats
☐ Övrig rapport
☐ _____

ISBN
—

ISRN
LITH-ISY-EX--06/3843--SE

Serietitel och serienummer **ISSN**
Title of series, numbering —

URL för elektronisk version

http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-6755

Titel
Title

Datatransmission över Talkodade Kanaler

Data Transmission over Speech Coded Voice Channels

Författare
Author

Andreas Tyrberg

Sammanfattning
Abstract

The voice channel in mobile communication systems have high priority and are almost always available. By using the voice channel also for data transmissions it is possible to get the same availability as for voice calls. But due to speech codecs in the voice channel, regular modems can not be used and special techniques are needed to transmit data.

This thesis presents methods to transmit data over the voice channel in a GSM, UMTS or TETRA network. The focus has been on robust data transmission rather than high data bit rates. Approaches are introduced which improve the reliability for transmissions even for systems with low rate speech codecs and channels with some distortion.

The results of the thesis are suggestions of symbol patterns and ways to create and adapt symbols for specific application and channel conditions to achieve the desired goal for the application.

Nyckelord
Keywords

data transmission, speech codec, voice channel, GSM, UMTS, TETRA

Abstract

The voice channel in mobile communication systems have high priority and are almost always available. By using the voice channel also for data transmissions it is possible to get the same availability as for voice calls. But due to speech codecs in the voice channel, regular modems can not be used and special techniques are needed to transmit data.

This thesis presents methods to transmit data over the voice channel in a GSM, UMTS or TETRA network. The focus has been on robust data transmission rather than high data bit rates. Approaches are introduced which improve the reliability for transmissions even for systems with low rate speech codecs and channels with some distortion.

The results of the thesis are suggestions of symbol patterns and ways to create and adapt symbols for specific application and channel conditions to achieve the desired goal for the application.

Acknowledgements

I would like to thank my advisor Janne Harju at ISY and my two advisors at Sectra Communications Robin von Post and Peter Nyman. I would also like to thank the rest of the people at Sectra Communications for their help and support. Thanks also to my examiner Fredrik Gustafsson and my opponent Johan Hedborg for there help and feedback on this thesis. And thanks to Chihsheng Tsai for help proofreading.

Contents

List of Figures

List of Tables

Acronyms

AbS Analysis-by-Synthesis

ACELP Algebraic Code Excited Linear Prediction

ADC Analogue to Digital Converter

ADPCM Adaptive Differential Pulse Code Modulation

AMR Adaptive Multi-Rate

AMR-WB Adaptive Multi-Rate Wideband

ANSI American National Standards Institute

ASCII American Standard Code for Information Interchange

BER Bit Error Rate

CELP Code Excited Linear Prediction

CSD Circuit Switched Data

DAC Digital to Analogue Converter

DTX Discontinuous Transmission

EFR Enhanced Full Rate

ETSI European Telecommunications Standards Institute

FFT Fast Fourier Transform

FIR Finite Impulse Response

FR Full Rate

GSM Global System for Mobile Communications

GPRS General Packet Radio Service

HR Half Rate

ISDN Integrated Services Digital Network

ISP Immitance Spectrum Pairs

ITU International Telecommunication Union

LP Linear Prediction

LPC Linear Prediction Coding

LSP Line Spectrum Pair

LTP Long-Term Prediction

NLMS Normalized Least Mean Square

PCM Pulse Code Modulation

PCS Personal Communications Service

PMR Professional Mobile Radio

PSTN Public Switched Telephone Network

PTT Push-To-Talk

RPE Regular Pulse Excitation

SNR Signal to Noise Ratio

TETRA Terrestrial Trunk Radio

UMTS Universal Mobile Telecommunications System

VAD Voice Activity Detector

VSELP Vector-Sum Excited Linear Prediction

VoIP Voice over IP

WCDMA Wideband Code Division Multiple Access

Chapter 1

Introduction

This report document the Master of Science thesis work performed during the spring 2006 at Sectra Communications AB in Linköping.

1.1 Background

Sectra Communications AB develops and sells products for secure mobile communication systems and high-speed encryption for telecom and data lines. Since the middle of the 1990's Sectra has developed a family of products called Tiger for personal communications. Tiger is a handheld and battery powered unit for encrypted speech and data service on the highest possible security level.

Today the Tiger units use the Circuit Switched Data (CSD) service over Global System for Mobile Communications (GSM) to transmit the protocol with encoded and encrypted speech. The Tiger doesn't contain any GSM module itself; instead it connects to a GSM cellular phone via Bluetooth to acquire CSD service. Figure 1.1 shows how two Tigers connect through cellular phones. Encoding of the speech and encryption of the encoded speech or data is performed inside the Tiger before further transmission through the cellular phone.

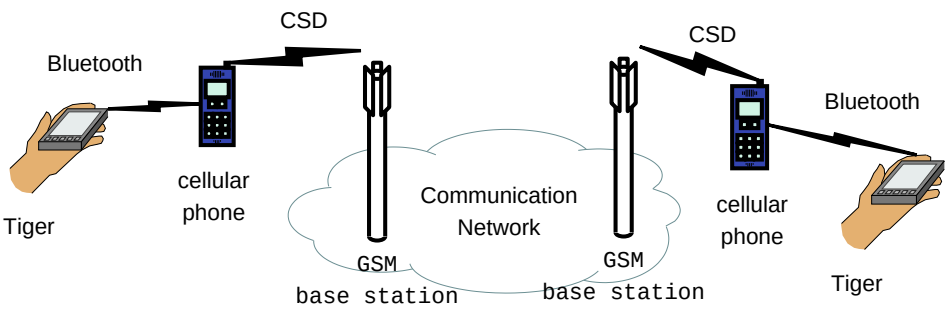

Figure 1.1. Two Tigers connected through cellular phones.

As normal users today use the voice channel for voice calls and services like General Packet Radio Service (GPRS) for data communication it has become hard to subscribe for the CSD service. Another problem is that it sometimes is uncertain if it is possible at all to connect between different systems (ISDN, PSTN, satellite) or different network providers (roaming) using CSD. In some countries (e.g. Canada), there are today no possibilities to sign up for a CSD subscription.

Normal voice calls, over the voice channel, are a basic service in the mobile systems and it is in general possible to connect between different systems and operators. If the Tiger unit can make use of the voice channel, as a complement to CSD, the accessibility and usability increase for users of the product.

The Tiger could connect to a cellular phone via handsfree, either using the handsfree profile in the Bluetooth interface or a standard wired handsfree. The Bluetooth solution has the advantage that no distortion will be introduced between the Tiger and the cellular phone.

The use of the handsfree interface also solves another problem. To establish a connection over CSD the AT-interface in the cellular phone is used. AT is short for attention and is a command set originally developed for regular modems. The AT-interfaces in different cellular phones are not unified and different functionality is supported in different cellular phones. When the handsfree interface is used, we will not have this problem and the number of cellular phones supported by the Tiger will increase.

Since the voice channel in a mobile system is speech coded it is not possible to use normal modems. Special modulation techniques are needed to be able to demodulate the signal to extract the data on the receiver side.

1.2 Goal

The main goal of this thesis is to investigate the possibility of data transmission over a speech coded digital channel. The investigation should increase Sectra's knowledge in the area to make a possible future implementation in the Tiger easier.

Even if achieving a high bit rate is a desirable goal, it is not the primary goal of the thesis. The proposed application for the voice channel data transmission, at least as a first stage, is to use it for Push-To-Talk (PTT) voice calls. Lower bit rates will introduce longer delays during a push-to-talk conversation but will not affect the functionality. A much more important criteria than short delay is that it should *always* be possible to make a call, even if a CSD connection cannot be established. This thesis focus is therefore instead on robustness and a system working for as many different speech codecs as possible supported by the GSM, Universal Mobile Telecommunications System (UMTS) and Terrestrial Trunk Radio (TETRA) networks. The data transmission should also be robust against other kind of distortions introduced during the transportation from the transmitter to the receiver.

1.3 Limitations

This thesis is only an investigation of data transmission over a speech coded voice channel. To use the findings in a real application, most likely, there need to be supporting functionality, e.g. channel coding, protocol, etc, which is not covered in the thesis.

The thesis is also limited to speech codecs found in GSM, UMTS and the TETRA networks. There exist many different speech codecs and different voice channels which make it impossible to describe and evaluate all of them. Many of the speech codecs are built around the same technologies as used in the GSM, UMTS and TETRA networks and the result should be applicable to those also. The reasons that focus is on these networks are that GSM is what the Tiger uses today and GSM and UMTS are the dominating standards for mobile communication. TETRA is a growing market for the Tiger which makes the network interesting.

1.4 Method

The work in this thesis is divided into a number of parts. The first part is a theoretical study of the problem and digital speech channels. A good understanding of the problem drastically increases the ability to solve the problem and to make a good work.

The next part of the work is to implement a framework to simulate the data transmission over speech coded voice channels. The framework is first used to repeat the method for data transmission over speech coded voice channels, developed at University of Surrey, England and some simple tests to get a deeper understanding of what impact the voice channel has on the transmitted signal.

The following step in the process is the improvements for speech channels with low rate speech codecs. The work starts with a system similar to University of Surrey which is changed iteratively to adapt to the new problem arising with low rate speech codecs. Unsuccessful efforts have been done to find out which codecs that really are used today. Since no information has been found, all codecs standardized for the GSM, UMTS and TETRA networks are investigated.

Since the goal is a system which has robust transmission, some simulations with distortion were conducted to see the effect on the Bit Error Rate (BER) for data transmission.

1.5 Disposition

The outline of the rest of this document is as follows:

Chapter 2 - Voice Channel describes the voice channel and the speech codecs used in GSM, UMTS and TETRA network.

Chapter 3 - Related work presents the work performed at University of Surrey regarding data transmission over the voice channel.

Chapter 4 - Simulation Framework introduces the implemented framework used for the simulations.

Chapter 5 - Improvements for Low Rate Speech Coded Channel describes different approaches to make the data transmission more reliable against speech channels with low rate speech codecs.

Chapter 6 - Robustness against Channel Distortions contains evaluation on how a few different channel distortions effect the data transmission.

Chapter 7 - Conclusions and Further Studies contains conclusions about the results and a proposal of further studies regarding data transmissions using the voice channel.

Chapter 2

Voice Channel

All regular calls and most of the speech traffic in a telephone network go over the voice channel. The voice channel is a transmission channel with bandwidth enough to carry human voice. In a digital system with digital speech channels, like GSM, UMTS and TETRA networks, the speech must be represented in a digital form to be transmitted. This chapter describes different ways to represent speech digitally and introduces speech codecs used in the GSM, UMTS and TETRA networks a little more detailed.

2.1 Speech Coding

The goal for speech coding is to represent speech digitally with as few bits as possible and retain an acceptable speech quality level. Speech codecs for the 300-3400 Hz frequency band is called narrowband or telephone speech codecs and for the 50-7000 Hz frequency band they are called wideband speech codecs. The two most common coding paradigms for narrowband speech coding today are waveform-following coding and Analysis-by-Synthesis (AbS) methods. [25]

Waveform-following codecs attempt to reproduce the time domain speech waveform of the original signal as exactly as possible. The waveform codecs work with any input signal bounded by certain limits in amplitude and bandwidth and not only speech. The codecs usually operate on sample to sample basis. Waveform codecs work well on bit rates from 16 kbps and higher. Examples on waveform-following codecs are the Pulse Code Modulation (PCM) and Adaptive Differential Pulse Code Modulation (ADPCM), see Section 2.2.1. [31, pp. 122]

In Analysis-by-Synthesis (AbS) methods linear prediction models and a perceptual distortion measure are used to reproduce only those characteristics of the input speech that are considered to be most important. Analysis-by-synthesis methods are the most common technology in speech codecs found in the GSM, UMTS and TETRA networks.

Another approach is to divide the speech into frequency bands and then code each band separately with for example analysis-by-synthesis. [25]

2.1.1 Analysis-by-Synthesis

The basic idea behind Analysis-by-Synthesis (AbS) is that the signal can observe and represent in some form, e.g. the time or frequency domain, and that there is a model with a number of parameters which can be varied. The parameters in the model are varied in a systematic way to find a set of parameters that can produce a synthetic signal which match the real signal with a minimum error. [32, pp. 199-202]

Code Excited Linear Prediction (CELP) is the most common and most successful analysis-by-synthesis method. In CELP speech coders, a segment (a frame or subframe) of speech is synthesized using the linear prediction model along with long-term redundancy predicator for all possible excitation vectors in what is called a codebook, see Figure 2.1. A perceptually weighted error signal is calculated for each of the excitation vectors and the vector that produce the minimum error is selected for use at the decoder. The linear prediction parameters and a codeword for the selected excitation vector are sent to the receiver to decode the speech. [25]

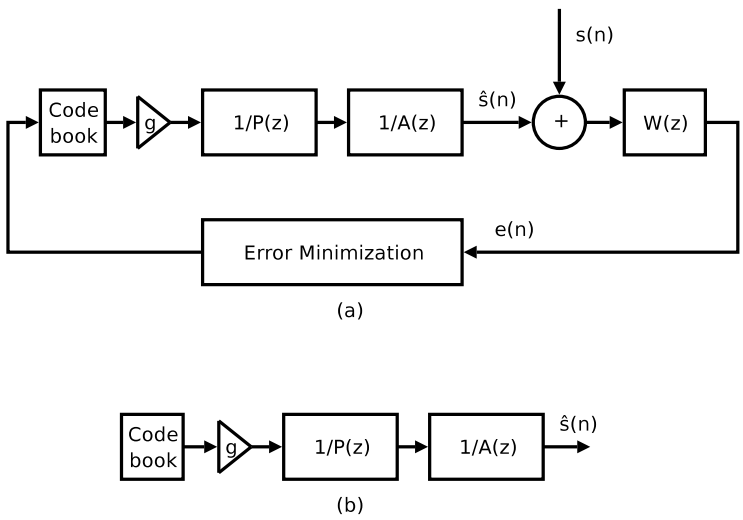

Figure 2.1. (a) Encoder for CELP coding. (b) Decoder for CELP coding.

The discovery of algebraic codebooks have reduced the computationally need for analysis-by-synthesis procedure. An algebraic codebook contains mostly zero values and only a small number of non-zero pulses are used. The pulses are positioned in interleaved tracks for efficient coding. Although there are several restrictions on each pulse position, together they are able to form most combinations necessary for adequate excitation. The pulses are usually also restricted to have the same amplitude, usually set to +1 or -1. [32, pp. 245-248]

See Section 2.2.4 for a closer explanation how the algebraic codebooks are used in speech coding.

Linear Prediction Coding

Linear Prediction Coding (LPC) analysis is one of the most powerful speech analysis methods. The technique models short-term correlations between speech samples. The aim with the analysis is to derive coefficients of a time-varying linear digital filter which models the spectral shaping of the vocal tract. The vocal tract filter is represented by the all pole transfer function

$$\frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{2.1}$$

where a_i is the LPC coefficients and p is the filter order. The coefficients are typically updated every 20 ms to 30 ms (once every frame) and the filter order is usually between 10 and 16. The filter removes the adjacent or neighboring sample correlations very efficiently. [3] [32, pp. 65-77, 202]

Long-Term Prediction

Long-Term Prediction (LTP) is sometimes also called pitch prediction (filter). The long-term prediction filter models correlations between sample that are one 'pitch' or multiple 'pitch' periods away in the speech. The aim with pitch prediction is to remove distant-sample correlations. [32, pp. 77-78]

The filter is given by

$$\frac{1}{P(z)} = \frac{1}{1 - \sum_{i=-I}^{I} b_i z^{-(D+i)}} \tag{2.2}$$

where D is a pointer to long-term correlation which usually corresponds to the pitch period or multiple periods and b_i are the pitch gain coefficients. The pitch filter is usually updated every 5-10 ms. Typically the filter has the form $I = 0$, i.e. 1 tap, and $I = 1$, i.e. 3 taps. [32, pp. 202-203]

For good speech quality of synthesized speech, a correct estimation of the pitch period is essential. The quality of the speech is seriously degraded if the pitch estimation is incorrect. Pitch period is defined as the time interval between two consecutive voiced (periodic) excitation cycles. The interval may change from cycle to cycle but usually evolves slowly, and therefore it can be estimated. Figure 2.2 shows an example of the residuals after LPC and LTP analysis. [32, pp. 149-150]

In recent years an adaptive codebook structure has become common to use to model long term memory. [25]

2.2 Speech Codecs in Telecommunication Systems

Several different speech codecs using a few different speech codings techniques are used in the GSM, UMTS and TETRA network. The following sections describe standardized codecs for these networks and PCM coding which is used for the backbone network.

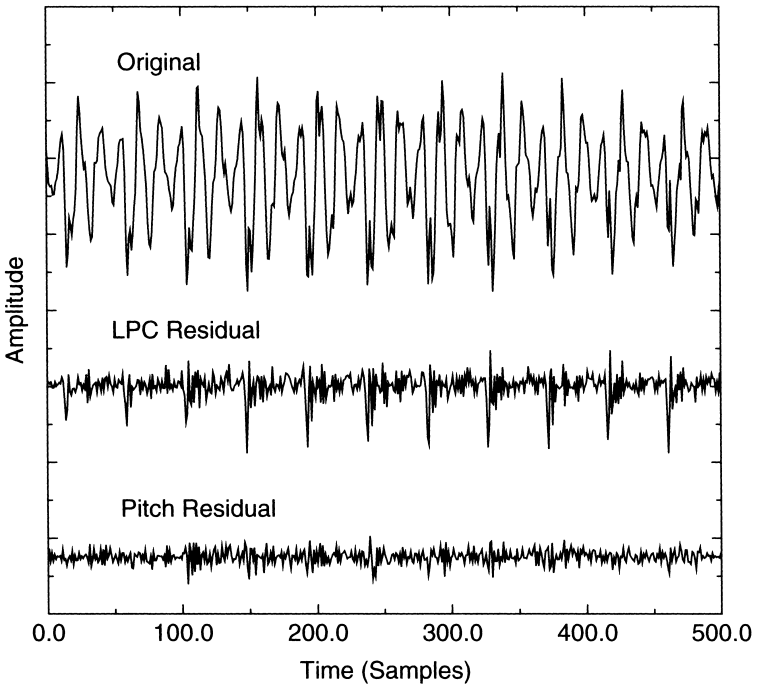

Figure 2.2. Time domain plot of the original, LPC and pitch residuals, figure from [32].

2.2.1 PCM and ADPCM

Logarithmic Pulse Code Modulation (Log-PCM) and Adaptive Differential Pulse Code Modulation (ADPCM) are two waveform codecs that have widespread applications. In long distance Public Switched Telephone Network (PSTN) Log-PCM at 64 kbps is used as speech codec at a rate of 64 kbps. [25]

ADPCM operates at 32 kbps or lower and it achieves performance comparable to log-PCM. ADPCM can by looking at more then one sample use a linear predictor to exploit short-term redundancy in the speech signal before quantization. By subtracting a predicated value from each input sample the dynamic range of the signal to be quantized is reduced. The smaller dynamic range requires fewer bits but it is still possible to achieve a good reproduction of the signal. [25]

2.2.2 GSM Full Rate

The first standard speech codec in the GSM system was the Full Rate (FR) codec, standardized in 1989. For a more detailed description see [34, pp. 156-162], [26, pp. 390-398] and the European Telecommunications Standards Institute (ETSI) standard [15]. GSM FR speech codec is based on a coding scheme called RPE-LTP which stands for Regular Pulse Excitation-Long Term Prediction. The codec works on 20 ms speech frames and has a bit rate of 13 kbps. Table 2.1 shows the bit allocation for the codec.

For every 5 ms subframe a very simple LTP filter is applied to remove distance sample correlations and for every frame an 8th order LPC filter is used for adjacent sample correlation. But instead of using a codebook the coder sample the input signal regularly (RPE, Regular Pulse Excitation) at a rate of only 8/3 kHz and send over as the excitation signal to the decoder. The decoder insert null value sample to obtain a signal sampled at 8 kHz again.

Table 2.1. GSM FR bit allocation.

Parameter	Bits
LPC filter	36
LTP filter	36
Excitation signal	188
Total	260

2.2.3 GSM Half Rate

GSM Half Rate (HR) codec [11] was standardized by ETSI in 1995 for use in the half rate channel. The codec uses the Vector-Sum Excited Linear Prediction (VSELP) algorithm, which is an algorithm belonging to the CELP algorithm class.

For every 20 ms speech frame the codec derives 18 parameters. The parameters are grouped into the following three general classes:

- energy parameters (R0 and GSP0)
- spectral parameters (LPC and INT_LPC)
- excitation parameters (LAG and CODE)

These parameters are quantized into 112 bit (Table 2.2) for transmission which, gives a bit rate of 5.6 kbps.

Once every frame, LPC coefficients are computed. The short term filter is of order 10. For all frames an overall frame energy is also computed and coded.

The codec has four different modes: unvoiced (mode = 0), slightly voiced (mode = 1), moderately voiced (mode = 2) and strongly voiced (mode = 3) which is selected once per frame. If a voiced mode (mode $\neq$ 0) is selected a long term predicator is used and the pitch lag is computed for every subframe. Each frame is divided into four 5 ms subframes. A combination of open-loop and closed-loop techniques is used to find the lag. First the open-loop finds "candidate" pitch lags for each subframe and then closed-loop search is employed to select a lag for transmission in each subframe.

All possible codevectors in the codebook are synthesized and compared with the input signal to select a codevector for each subframe. To select codevector, the difference between the synthesized signal and the input signal are filtered by a spectral weighting signal (and possibly a second weighting filter). The power of

this weighted error signal is computed and the codevector with minimum weighted error power is selected. If mode = 0, two VSELP codebooks are used which are searched sequentially.

Table 2.2. GSM HR bit allocation.

Parameter	Bits
MODE	2
Frame energy	5
LPC filter	29
MODE = 1, 2 or 3	
LAG	20
Codebook	36
Gain	20
MODE = 0	
Codebook (both)	56
Gain	20
Total	112

2.2.4 GSM Enhanced Full Rate and Adaptive Multi-Rate

The Enhanced Full Rate (EFR) codec [14], [27], [35] was standardized by ETSI in 1996 for the GSM mobile communication system. The codec was also chosen as the EFR for GSM technology based US Personal Communications Service (PCS) 1900 system.

In 1999, the Adaptive Multi-Rate (AMR) codec [13], [5] was standardized for GSM. The codec is an improvement over previous GSM speech codecs in error robustness by adapting speech and channel coding depending on channel conditions. In 1999 the codec was also adapted as the default speech codec for the Wideband Code Division Multiple Access (WCDMA) 3G system. The codec was jointly developed by Ericsson, Nokia and Siemens.

AMR operates in 8 modes with bit rates of 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2 and 12.2 kbps. GSM EFR and AMR at 12.2 kbps mode is computationally the same codec and therefore will only AMR be described below.

The codec is based on the Algebraic Code Excited Linear Prediction (ACELP) coding algorithm and operates on 20 ms speech frames sampled with 8 kHz sample frequency (160 samples). Each frame is divided into four 5 ms subframes. The AMR codec can switch between different bit rates for every 20 ms speech frame.

Linear Prediction

Twice every frame, a 10th order Linear Prediction (LP) analysis is performed for the 12.2 kbps codec and once for the other modes. The two sets of LP parameters in the 12.2 kbps mode are converted to Line Spectrum Pair (LSP) and jointly

quantized using split matrix quantization and for the other modes split vector quantization is used.

Adaptive Codebook

To reduce the complexity, the adaptive codebook (CB) search or long term prediction analysis is performed in two stages - open-loop search and closed-loop search. The open-loop pitch search is performed once per frame for the 4.75 and 5.15 kbps modes and twice for the other modes. The closed-loop search is performed around the open-loop pitch lag to estimate the closed-loop pitch. The adaptive codebook operates with fractional resolution of 1/3 for all modes except the 12.2 mode which has 1/6 resolution.

Algebraic Codebook

The largest variation in bit rates of the different modes comes from the fixed algebraic codebook (CB) where the bit rates range from 7.0 to 1.8 kbps. Different bit rates are obtained by varying the number of pulses, from 10 to 2, in each subframe depending on the mode.

To restrict the number of possible pulse positions, the subframes are divided into pre-defined tracks and each pulse is located in one of these tracks. Pulse positions in Table 2.3 are used for all modes except for 10.2 kbps which use pulse positions in Table 2.4. The amplitude is preset to either +1 or -1 to simplify the search procedure. Two overlapping pulses in one track result in a single pulse with amplitude +2 or -2. For each pulse the sign bit and bits describing the pulse position within the track are transmitted together with the gain for the algebraic codebook.

Table 2.3. AMR pulse positions for each track.

Track	Position
1	0, 5, 10, 15, 20, 25, 30, 35
2	1, 6, 11, 16, 21, 26, 31, 36
3	2, 7, 12, 17, 22, 27, 32, 37
4	3, 8, 13, 18, 23, 28, 33, 38
5	4, 9, 14, 19, 24, 29, 34, 39

12.2 kbps mode - Two pulses are located in each of the five tracks in Table 2.3. The sign of the second pulse in each track is not explicitly transmitted.

10.2 kbps mode - Two pulses are located in each of the four tracks in Table 2.4. As for the 12.2 kbps mode, the sign of the second pulse is not transmitted.

7.95 and 7.40 kbps mode - One pulse is located in each of the tracks 1, 2 and 3 and one pulse located in either track 4 or 5 in Table 2.3.

Table 2.4. AMR 10.2 kbps pulse positions for each track.

Track	Position
1	0, 4, 8, 12, 16, 20, 24, 28, 32, 36
2	1, 5, 9, 13, 17, 21, 25, 29, 33, 37
3	2, 6, 10, 14, 18, 22, 26, 30, 34, 38
4	3, 7, 11, 15, 19, 23, 27, 31, 35, 39

6.70 kbps mode - One pulse is located in track 1, one pulse located in either track 2 or 4 and one pulse located in either track 3 or 5 in Table 2.3.

5.90 kbps mode - The first pulse is located in track 2 or 4 and then a second pulse is located in track 1, 2, 3 or 5 in Table 2.3.

5.15 and 4.75 kbps mode - Two pulses are located in two different tracks in Table 2.3.

An iterative, non-exhaustive search of pulse positions is performed for the modes at 6.7 kbps and higher and an exhaustive search is performed for lower bit rate modes.

Bit allocation

Table 2.5 shows the bit allocation per frame for the different AMR modes.

Table 2.5. AMR bit allocation.

	Mode kbps							
Parameter	12.2	10.2	7.95	7.4	6.7	5.9	5.15	4.75
LSP	38	26	27	26	26	26	23	23
Adpt. CB	30	26	28	26	24	24	20	20
Alg. CB	140	124	68	68	56	44	36	36
Gains	36	28	36	28	28	24	24	16
Total	244	204	159	148	134	118	103	95

2.2.5 Adaptive Multi-Rate Wideband

The Adaptive Multi-Rate Wideband (AMR-WB) speech codec [22], [2] was standardized in 2001 for the WCDMA 3G and GSM system. Unlike the other GSM codecs, AMR-WB is a wideband speech codec, i.e. working on the 50-7000 Hz frequency band. The codec operates on signals sampled at a rate of 16 kHz divided into 20 ms speech frames.

Nine different speech coding modes exist, with bit rates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 and 6.6 kbps, which the codec can operate in. The 8.85 and 6.6 kbps modes are intended to be used only temporarily during severe radio channel conditions or during network congestion. All other modes offer a high quality wideband speech.

AMR-WB is based on the Algebraic Code Excited Linear Prediction (ACELP) technology. To decrease the complexity and to focus the bit allocation into the subjectively most important frequency range, two frequency bands, 50-6400 Hz and 6400-7000 Hz, are coded separately. Before the ACELP algorithm is applied, the signal is down-sampled to 12.8 kHz and pre-processed using a high-pass filter and a pre-emphasis filter.

Linear Prediction

Once every 20 ms a Linear Prediction (LP) analysis is performed. The set of 16 LP parameters is converted into an Immitance Spectrum Pairs (ISP) and vector quantized using split-multistage vector quantization. Each frame is divided into four subframes of 5 ms (64 samples at 12.8 kHz).

Adaptive Codebook

The pitch search is performed in three stages. The first stage performs an open-loop pitch lag search twice every subframe for all modes except the 6.6 kbps. For 6.6 kbps mode the open-loop search is only performed once per frame.

The second and third stages are closed-loop searches to find the pitch lag. In the first closed-loop search the lag is performed for integer lags around the estimated open-loop pitch lag. Last stage, second closed-loop search, searches through the fraction around the optimum closed-loop integer. The resolution of the pitch lag is between 1/4 to 1 depending on the sample range, subframe number and codec mode. The adaptive codebook parameters sent to the decoder are the delay and gain of the pitch filter.

A frequency-dependent pitch predictor is used to enhance the pitch prediction performance in wideband signals. 1 bit per subframe is used to encode if a low pass filter for filtering the pitch codevector should be used.

Algebraic Codebook

To guarantee a high subjective quality in wideband speech coding a very large codebook is needed. AMR-WB use an algebraic codebook, a codeword is searched for every 64 samples subframe.

The 64 samples position is divided into 4 tracks with each 16 positions. Each track can have from 1 to 6 pulses depending on mode. All the pulses have either the amplitude +1 or -1. Table 2.6 shows the pulse position for each track. The excited vector transmit to the decoder is built up from the pulse positions.

23.85 and 23.05 kbps modes - 6 pulses are located in each of the four tracks in Table 2.6 resulting in a total of 24 non-zero pulses.

Table 2.6. AMR-WB pulse positions for each track.

Track	Position
1	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

19.85 kbps mode - A total of 18 non-zero pulses are located in the tracks. Five pulses are located in track 1 and 2 and four pulses are located in track 3 and 4 in Table 2.6.

18.25 kbps mode - Four pulses are located in each of the four tracks in Table 2.6.

15.85 kbps mode - 12 non-zero pulses, three in each track in Table 2.6, are placed to make up the excited vector.

14.25 kbps mode - Three pulses located in track 1 and 2 and two pulses located in track 3 and 4 in Table 2.6 are placed which sums up to a total of 10 pulses.

12.65 kbps mode - Two pulses are located in each of the four tracks in Table 2.6.

8.85 kbps mode - One pulse is located in each of the four tracks in Table 2.6, summing up to a total of 4 pulses.

6.60 kbps mode - One pulse is located in either track 1 or 3 and one pulse is located in either track 2 or 4 in Table 2.6, resulting in a total of 2 pulses.

Higher Frequency Band

In the decoding, the higher frequency band 6400-7000 Hz is reconstructed using parameters from the lower band and a random excitation. The transmitted parameters don't contain any information about the higher band except for the 23.85 kbps mode where the gain is included.

Bit Allocation

Table 2.7 shows the bit allocation for a frame for the different modes of AMR-WB codec.

2.2.6 TETRA Speech Codec

Terrestrial Trunk Radio (TETRA) is an ETSI standard for a trunked digital mobile radio system for voice and data communication. The purpose of the standard is to meet the needs of traditional Professional Mobile Radio (PMR) user organizations such as police, ambulance, fire, transport and security services.

Table 2.7. AMR-WB bit allocation per frame.

	Mode kbps								
Parameter	6.60	8.85	12.65	14.25	15.85	18.25	19.85	23.05	23.85
VAD flag	1	1	1	1	1	1	1	1	1
LTP-filtering	0	0	4	4	4	4	4	4	4
ISP (LP)	36	46	46	46	46	46	46	46	46
Pitch delay	23	26	30	30	30	30	30	30	30
Alg. code	48	80	144	176	208	256	288	352	352
Gains	24	24	28	28	28	28	28	28	28
High-band	0	0	0	0	0	0	0	0	16
Total	132	177	253	285	317	365	397	461	477

The TETRA speech codec [23] employs the Algebraic Code Excited Linear Prediction (ACELP) technique. In the TETRA codec 30 ms speech frames and 7.5 ms subframes sampled at 8 kHz are used. After encoding, each frame is represented by 137 bit resulting in a bit rate of 4.567 kbps. Table 2.8 gives the bit allocation for the different parameters.

Table 2.8. TETRA codec bit allocation.

Parameter	Bits
LP filter	26
Pitch delay	23
Alg. codebook	64
Gains	24
Total	137

For every frame, a short term analysis (LPC analysis) is performed. The LPC filter coefficients are converted into Line Spectrum Pair (LSP) for quantization and interpolation purposes. The filter is of order 10.

A long term prediction analysis is performed for every subframe. The pitch filter is implemented using the adaptive codebook approach. A two stage search, open-loop and closed-loop, is employed for the pitch analysis.

The algebraic codebook is 16 bit and contains at most four non-zero pulses with the fixed amplitudes of +1.4142, -1, +1, -1. Allowed positions for each pulse can be found in Table 2.9. All pulse positions can simultaneously be shifted by one to occupy odd position and the sign of all pulses can simultaneously be inverted with the global sign bit. For the third and fourth pulse in Table 2.9 the last pulse positions are outside the subframe and indicate that these pulses are not present. The codebook is searched by minimizing the mean squared error between the weighted input speech and the weighted synthesis speech.

Table 2.9. TETRA codec codebook parameters.

Codebook parameters	Position of the pulses	Bit allocation
Pulse amplitude: +1.4142	0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58	5
Pulse amplitude: -1	2, 10, 18, 26, 34, 42, 50, 58	3
Pulse amplitude: +1	4, 12, 20, 28, 36, 44, 52, (60)	3
Pulse amplitude: -1	6, 14, 22, 30, 38, 46, 54, (62)	3
Global sign flag		1
Shift flag		1

2.3 Voice Activity Detector

A Voice Activity Detector (VAD) is used to detect active and inactive regions of speech in a speech codec. Compression of inactive speech regions provide benefits in speech communication systems with bandwidth limited communication channels, between other things co-channel interference reduction in cellular communication systems, power-savings for mobile terminals, reduction in packet losses when transmitting voice over packet based networks and bit rate reduction. The VAD usually produce, for every 10-20 ms long speech segment, a binary decision indicating either active or inactive region. [32, pp. 357-359]

At the decoder end, a comfort noise generator reconstruct the inactive frames and gives a natural background sound with smooth transitions between active and inactive segments. Average information on the background signal is regularly transmitted by the encoder to enhance the naturalness of the generated background signal. [32, pp. 357]

The ETSI GSM FR, HR and EFR VAD algorithms ([10], [7] and [8]) have a common structure in which the predictive residual energy is compared with an adaptive threshold. The algorithms make the assumption that the average spectral shape will be similarly to the current frame's shape if the signal is background noise only. Background noise in most environments is fairly stationary and will over time have similar spectral shape. The similar spectral shape will result in smaller residual signal energy and be marked as inactive. [32, pp. 361-362]

For the UMTS network ETSI have two options, AMR1 and AMR2 ([12]). Both the algorithms are based on spectral subband energies. AMR1 decompose the input signal with filter banks and then calculate each subband's energy and corresponding SNR estimation. The sum of the subband SNRs are compared with an adaptive threshold to make the VAD decision, followed by hangover. Hangover is a metodology to transmit a few extra speech frames after the VAD has marked a frame as inactive to avoid clipping in the end of words. AMR2 transform the signal into the frequency domain using FFT and then calculate each subband's energy and SNR on the spectra. [32, pp. 362-363]

Some speech codecs, e.g. GSM EFR, also contain the function of discontinuous

transmission (DTX). DTX allow the radio transmitter to switch off during an inactive period to save power and also reduce the overall interference in the air interface. [27]

2.4 Error Concealment of Lost Frames

Received speech frames can be erroneous and normal decoding of these frames would result in very unpleasant noise effects. Speech frames can also be lost during the transmission. To conceal the effect of these lost frames the GSM and UMTS codecs substitute the frames with either a repetition or an extrapolation of the previous good speech frame(s). In the case of subsequent lost frames a muting technique is used to gradually decrease the output level. More detailed description for specific codecs can be found in the ETSI standard [9], [6], [19], [20] and [21].

2.5 Tandem Connections

A voice call with a GSM cellular phone is encoded with a speech codec (EFR, AMR, etc.) to a digital representation that is transmitted to the base station. Across the core circuit switched network the voice call is transmitted in the form of PCM (ITU recommendation G.711) or ADPCM, which results in that the digital representation produced by the speech codec has to be transcoded to PCM or ADPCM. In a GSM-to-GSM voice call the speech may be transcoded twice, see Figure 2.3. The first transcoding is from the transmitting cellular phone's speech codec format to the PCM or ADPCM and the second transcoding is from PCM or ADPCM to a speech codec format supported by the receiving cellular phone. [34, Section 3.3.2] [25]

It is common for mobile-to-mobile calls to have asynchronous tandem of different codecs because the cellular phone support different codecs. The term asynchronous tandem refers to where the speech sample must be reconstructed and re-encoded by the next codec. Further transcodings can occur as transmissions between base stations are not necessarily over PCM links, the backbone can be a Voice over IP (VoIP) network which utilize other codecs. [25]

The speech encoding and tandem connections in a voice channel between two cellulars retain an acceptable speech quality level. While the resulting synthesized speech sounds similarly to the input speech it may have a fairly different waveform sample-by-sample. This prevents most data modems to be used over the voice channel. [28]

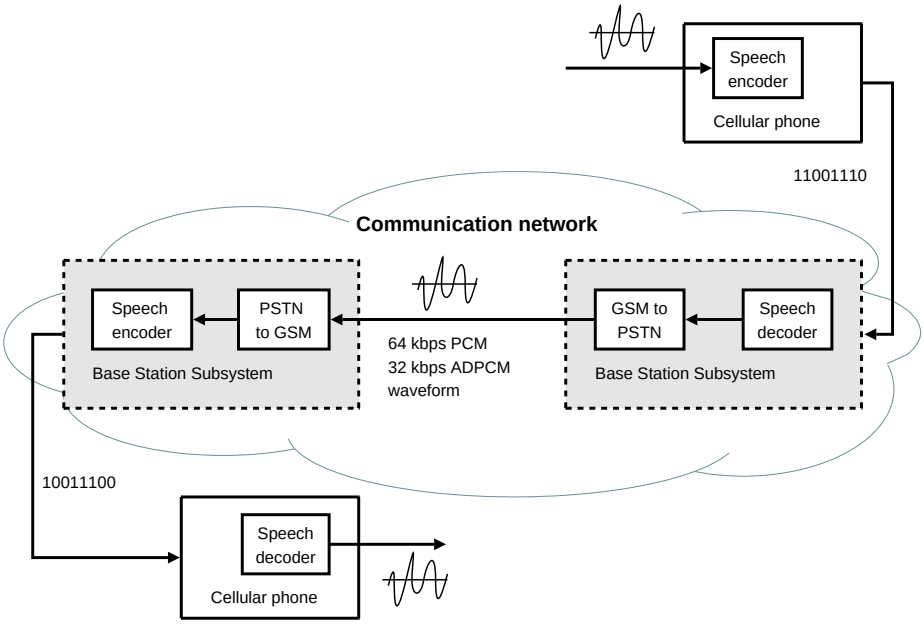

Figure 2.3. GSM tandem connection.

Chapter 3

Related work

As far as I know, the only related work using a speech coded voice channel to transmit data have been done by Katugampala, Al-Naimi, Villette and Kondoz at University of Surrey, United Kingdom. Their work is described in [30], [28] and [29] and in their patent [33].

They modulate digital data onto speech-like waveforms. The waveforms are transmitted over the GSM voice channel and demodulated at the receiver. Their system achieves a throughput of 3 kbps with 2.9% Bit Error Rate (BER) and with addition of error correction code a net throughput of 1.2 kbps and bit error rate of 0.03%. To avoid potential problem with analogue interfaces, a digital interface was emulated on the transmitter side. The receiver side used an analogue interface. They also performed data transmissions with analogue interface both on the receiver and transmitter side which resulted in higher bit error rates.

Their implementation of the system runs on two laptop PCs which are connected to cellular phones via the handsfree cable from the sound cards. When a digital interface was used, the modulated signal was transferred as a data file to the cellular phone and played while the phone was on a call with the second phone.

The goal of the system is a real time, end to end secure voice communication. To accomplish this goal they use a very low bit rate speech codec (1.2 kbps) to not exceed the available bandwidth.

While they have the goal of real time voice communication the goal in this thesis is to achieve a robust communication that works well in many different condition.

3.1 The Surrey Way

Figure 3.1 shows an overview of the data transmission parts of the system built by Katugampala et al. The first service access point is arranged to transmit voice over a voice communication network and the second is arranged to receive voice over the network.

The following sections will describe their system. Since this thesis only investigates the functionality from modulation to demodulation, only this part will be

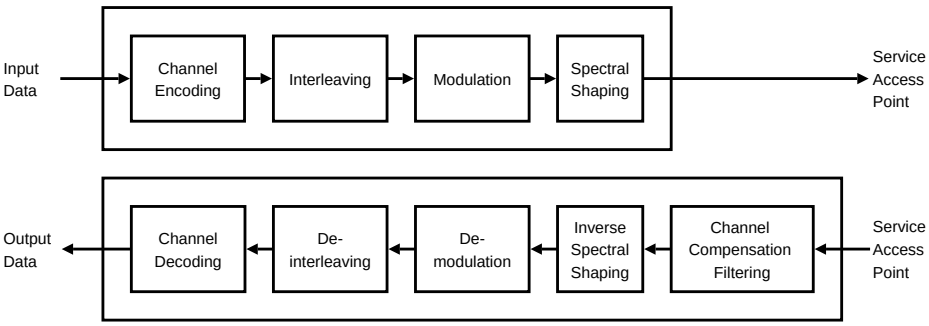

Figure 3.1. Overview of the University of Surrey's system, system from [33]

described. More information about used channel coding, interleaving and other details about their research can be found in earlier mentioned references.

3.1.1 Transmitter

Modulation

The modulation converts the data into 5 ms waveform symbols at 8 kHz sampling rate, therefore each symbol is 40 samples. The 40 samples are divided into five tracks, each consisting of 8 sample positions.

Each symbol carries 15 bits of data. The 15 bits are divided into five groups of 3 bits. Each of the 3 bit groups are allocated to one of the tracks and the data bits defines a sample position of a pulse in the corresponding track. Table 3.1 shows which sample position that belongs to each track and which sample position of a pulse that corresponds to a certain data bit pattern. The sample positions are divided into the same tracks as for the algebraic codebook for GSM EFR and most modes for AMR (see Section 2.2.4 and Table 2.3).

Table 3.1. Pulse positions for each track and corresponding data bit pattern.

Track 1:	0	5	10	15	20	25	30	35
Track 2:	1	6	11	16	21	26	31	36
Track 3:	2	7	12	17	22	27	32	37
Track 4:	3	8	13	18	23	28	33	38
Track 5:	4	9	14	19	24	29	34	39
Data bits:	000	001	010	011	100	101	110	111

After all pulse positions have been defined, the sign of the pulses have been defined to alternate through the symbol, every second pulse is negative.

Example 3.1

If the 15 bits `110010110101011` of data, the first 3 bits group is `110` and therefore the pulse should be on sample position 30, the next 3 bits, `010`, place a pulse on position 11, and the remaining pulses will be located on sample position 32, 28 and 19. Figure 3.2 shows the symbol when all 5 pulses have been placed and the signs of the pulses have been alternated through the symbol.

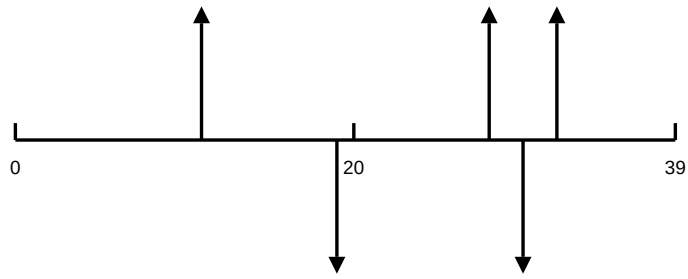

Figure 3.2. Symbol example, the symbol encode the 15 bits 110010110101011 of data.

Finally the complete waveform is multiplied with a preferred gain to make the signal suitable for onward transmission. The symbols are transmitted in a sequence to produce a continuous waveform signal.

As a modification, the symbols close or similar to each other are reorganized to also have a similar data bit pattern, i.e. short hamming distances. By assigning similar data bit patterns to similar symbols the bit errors are minimized if a symbol is wrongly demodulated, since it is more likely to be confused with another symbol that is similar.

Spectral Shaping

The function of the spectral shaping module is to ensure that the spectral shape of the signal varies over time. The signal needs to vary over time so that any Voice Activity Detector (VAD) on the voice channel will not identify any parts of the transmitted signal as a no-speech and cut them out of the transmission. Only a 20 ms section of 80 ms section of the waveform is modified and the remaining part is left unchanged (Figure 3.3).

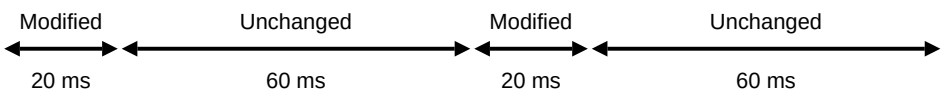

Figure 3.3. Spectral shaping section modification, the filter is applied 20 ms of every 80 ms section.

The spectral shaping module applies a gain that varies from frequency to frequency component of the signal wave. The gain varies in a sinusoidal manner between a minimum of 1 at 0 Hz, 4000 Hz and -4000 Hz and a maximum of 4 at 2000 Hz and -2000 Hz, Figure 3.4 shows the principle shape of the spectral function. After the spectral shaping function has been applied, the spectrum of the signal is significantly different from the unmodified signal and it is enough to ensure that a VAD will not cut the signal off.

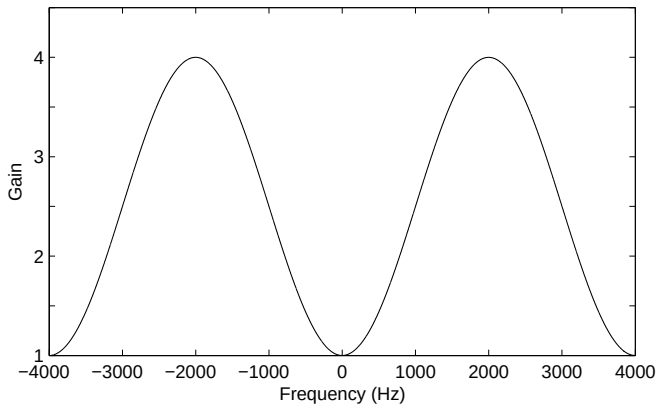

Figure 3.4. Principle shape of the spectral shaping function.

The shaping function changes the pulse in the original symbol by replacing them by a feature having in principle the shape shown in Figure 3.5. The central peak has the same position and sign as the original pulse.

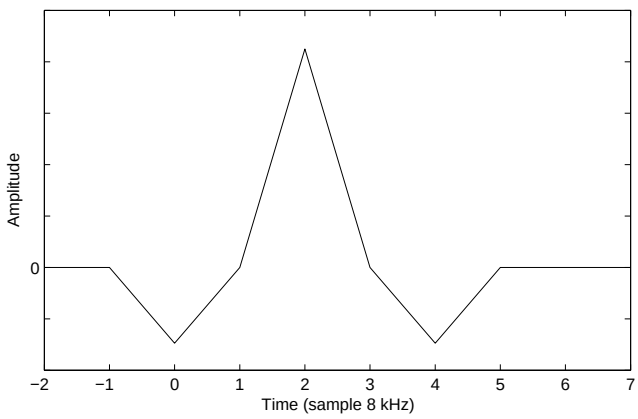

Figure 3.5. Principle shape of the spectral shaping function in time domain.

The spectral shaping can be performed in the time domain by a filter or convolution operation such that each single pulse is replaced by a feature having the shape like the principle shape in Figure 3.5.

After the waveform signal has been shaped by the filter, the signal is the output to the first service access point.

3.1.2 Receiver

Channel Compensation Filter

First step on the receiver side is to apply the channel compensation filter on the waveform signal coming from the second service access point. The functionality of the filter is to counteract the response of the entire communication link between the modulator and demodulator. The filter is arranged to have the inverse of the response of the telecommunication network which improves the demodulation result.

The filter is an adaptive filter. A predetermined training sequence is in a first stage transmitted by the modulator to adapt the coefficient (P3 in Figure 3.6). In the second stage the adaptation is suspended (P2) or the continuous adaptation mode (P1) is used. Suspend mode may be suitable if the voice channel response is time invariant.

In the continuous adaptation mode, data from the output stream is used to generate the reference signal. The output data stream is modulated in modules operating in the same way as for modulator to generate the reference signal, see Figure 3.6.

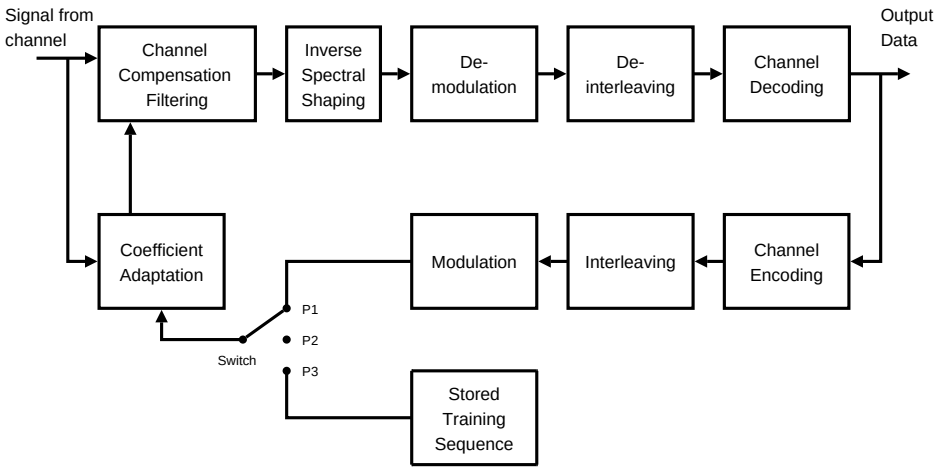

Figure 3.6. Filter coefficient adaptation, system from [33].

Inverse Spectral Shaping

The inverse spectral shaping module performs a function that is the inverse of that performed by the spectral shaping function. This removes the added modification and makes the input signal to the demodulation module as close as possible to the output from the modulation module in the transmitter.

Demodulation

The demodulation module performs essentially the inverse functionality of the modulation module. The received modulated waveform signal is compared with the reference waveform of all possible symbols and a matching metric is determined. Each symbol carries 15 bits so there are $2^{15} = 32768$ reference symbols.

The matching metric between the best symbol waveform and one or more other symbol waveforms may be small. Therefore the channel decoder performance is improved if soft decisions are considered rather than hard decisions of each bit being one or zero.

Each symbol has a unique corresponding data bit pattern. A weight for each of the 15 bits in a symbol is estimated as the soft decision. The estimation for the j bit is given by

$$w_j = \sum_{i=0}^{i=32767} n_{i,j}\ s_i \quad (0 \leq j \leq 14) \tag{3.1}$$

where $n_{i,j}$ is $+1$ if the jth bit in the ith symbol is 1 and -1 if it is 0. s_i is the similar metric between the received symbol and the ith reference symbol.

The weights are then input to the channel decoder which uses the values to estimate the best possible data output bit stream.

3.1.3 Synchronization

At the start of any communication a synchronization signal is sent from the modulator to the demodulator. The synchronization signal is a signal with pulses on predetermined time intervals. On the demodulator side the signal is passed through the channel compensation filter using a fixed set of filter coefficients that representing the average inverse of the target voice communication system before the receiver try to recognize the signal and synchronize.

3.1.4 Lag Correction

Analogue links within the telecommunication system and/or an analogue interface to the service access point may cause two problems. Due to constant phase shift in the Digital to Analogue and Analogue to Digital Converter (DAC and ADC) the sample received by the demodulator may be a fraction of a sample delayed compared to those sent by the modulator. The second problem is that the frequencies of the DAC and ADC might be slightly different. A slight difference in the frequencies will result in stretching or shrinking the received signal. The frame synchronization in the demodulator will be lost due to this.

The channel compensation filter can compensate and realign up to a few samples and synchronize the filter output to the exact sample location with respect to the reference signal. Since this effect is time invariant once a voice channel has been established this cause no adverse effect or degradation of the performance of the channel compensation filter.

By measuring the lag between the reference signal and the preprocessed signal and correct after the measured lag, larger mismatch is prevented due to different clock frequencies. The lag is estimated by cross correlation and the correction is performed by upsampling either the channel signal or the reference signal, correct the lag and then down sampling with the correct lag.

Chapter 4

Simulation Framework

A framework to simulate data transmission over a speech coded voice channel has been implemented. The framework has been used to find ways to increase the robustness of a data transmission over the channel. This chapter describes the framework, its different modules and the most important configuration capabilities.

4.1 Overview

The framework can be divided into five subsystems that perform different tasks of the simulation process. All subsystems consist of one or more programs which take one or more input files (except the random input program) and generate one or more output files. Communication between programs is through files.

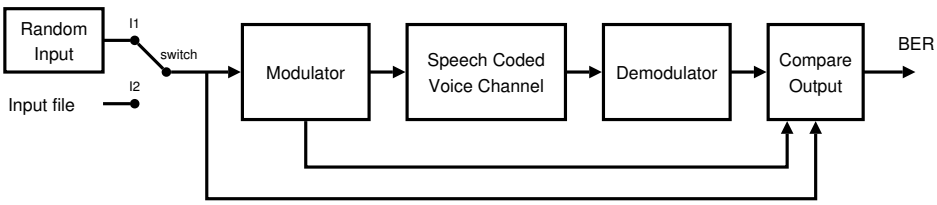

Figure 4.1. Overview of the simulation framework.

The first stage is used to select a data source to transmit. It is possible to specify a file or generate new random binary data and use that as source for the transmission.

The last stage in a simulation is the compare stage where the input data to the modulator is compared to the output data from the demodulator. The outputs from the compare program are:

- number of incorrect bits
- bit error rate BER

- a list of where the incorrect bits occurred in the transmitted bit stream
- distances between bit errors

The modulator and the demodulator also generate lists of transmitted symbols and received demodulated symbols which are used to get a ratio measurement of the correct number of transferred symbols.

The stages between the input and comparing are the modulation of data, simulation of the channel and demodulation of the transmitted signal. The modulator and demodulator are the modules that a final system needs to implement.

4.2 Modulator

The modulator transforms the input data stream into PCM waveform symbols which are transmitted in a sequence over the voice channel.

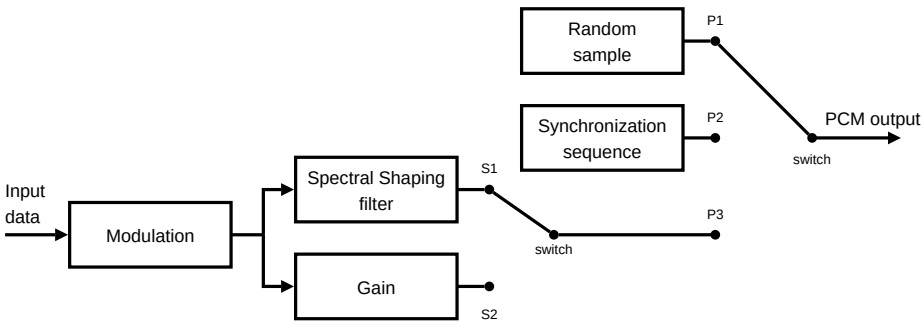

Figure 4.2. Modulator framework.

4.2.1 Random Samples

Before the data is transmitted, a random number of samples is generated. The samples are generated to assure that the modulator generated symbols are not (always) in synchronization with the codecs frames/subframes in the communication channel. There is no simple approach in a real system to guarantee such synchronization. By adding a random number of samples we avoid special case behavior to occur due to synchronization in the simulation.

4.2.2 Synchronization Sequence

After the random number of samples has been written to the PCM output stream, a synchronization sequence of samples is written to the stream. The predefined synchronization sequence of samples is known to both the modulator and demodulator. Delays are introduced in the modulator, the voice channel and the demodulator which make some kind of synchronization between the modulator

and demodulator necessary. The actual synchronization is performed by the demodulator.

4.2.3 Modulation

The modulation module in Figure 4.2 is responsible for iteratively reading bits of data from the input data stream and generating PCM waveform symbols.

The implementation supports several different symbol patterns which have different characteristics and different number of bits encoded per symbol. A typical symbol has a length of 40-80 samples (5-10 ms) and all symbols in a symbol pattern have the same length. Most symbol patterns encode data in a similar approach as used in the University of Surrey system, i.e. encode the data in pulse positions, but have also some difference, among other things different number of pulses and different distances between the pulses.

4.2.4 Spectral Shaping

Before the symbols are written to the output PCM stream they can optionally be filtered by the spectral shaping filter. The filter is implemented as a FIR filter and is applied to the signal in the time domain. Figure 4.3 shows the time domain shape of the filter which has almost the same shape as the filter used in Surrey (compare with Figure 3.4 and 3.5). By applying the filter to the signal the frequency spectrum of the signal is more conformed to the frequency band which the narrowband speech codecs encode.

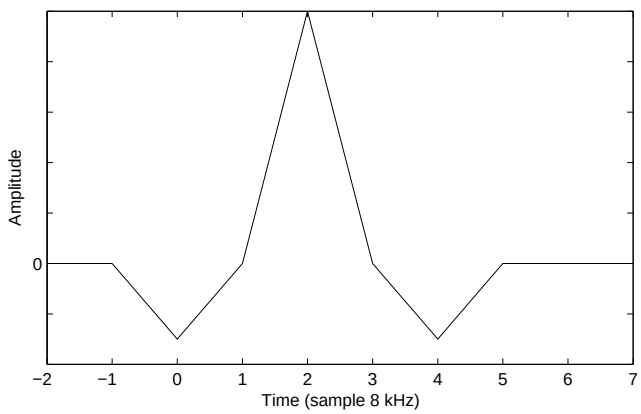

Figure 4.3. The spectral shaping filter in the time domain.

As a difference to the University of Surrey's system, the spectral shaping filter in the framework is either applied or not applied. The Surrey system applies the filter 20 ms of 80 ms, see Section 3.1.1, with the main reason to avoid cut outs caused by the Voice Activity Detector (VAD). During the simulations I have in

general case not experienced any greater degradation caused by cut outs for most of the different VAD in the reference implementations. Using the filter only 20 ms of 80 ms increases the complexity of the framework and therefore this approach was selected instead. In some situations, using the spectral shaping filter increases the performance while in other cases the filter decreases the performance.

The signal is also multiplied with a gain prior to further transmission. When the spectral shaping filter is used, the gain is combined with the filter and when no spectral shaping filter is used, the gain is multiplied directly with the signal.

4.3 Speech Coded Voice Channel

The speech coded voice channel subsystem is responsible for simulating the communication channel between the modulator and the demodulator, i.e. cellular phone to cellular phone. Distortion is applied to the speech waveform signal during the transmission. Main source of distortion in the framework is the speech codecs in the channel.

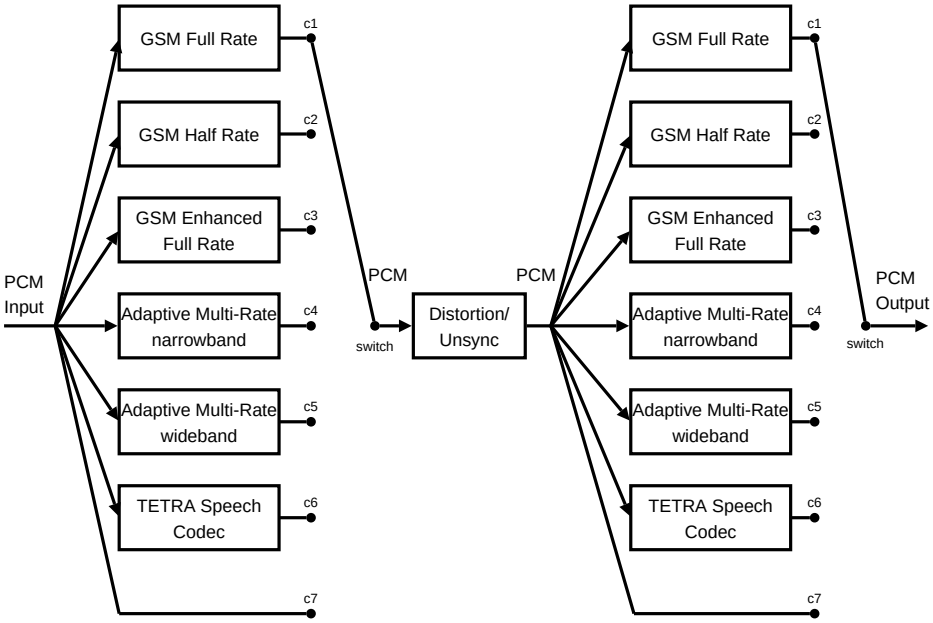

Figure 4.4. Framework voice channel.

4.3.1 Speech Codecs

All standardized codecs in the GSM, UMTS and TETRA network can be used with the framework. The PCM waveform signal can be encoded and decoded twice to simulate the tandem transcodings that can occur in the GSM network.

To simulate the speech codecs, fixed-point ANSI-C implementations from ETSI are used for the GSM HR [18], GSM EFR [17], AMR [16], AMR-WB [24] and TETRA [23] speech codec. Simulation of GSM FR speech codec is performed with the C library from [4]. The correctness of the FR C library has been verified with ETSI test patterns provided with the ETSI standard [15]. All codec modules in Figure 4.4 consist of one encoder of a PCM waveform signal and one decoder which generates a new PCM waveform output signal. Output from the speech encoding is the extracted parameters from the speech signal. The parameters are different for different codecs which make it hard to use them in some general way.

The implementations of GSM HR, GSM EFR, AMR and AMR-WB have support for VAD which can be activated to simulate the effect of some symbols sequence and patterns may be classified as inactive region and cut out from the stream. Both AMR1 and AMR2 VAD algorithms are implemented for AMR.

AMR-WB uses input waveform signals and generates output waveform signals sampled at 16 kHz while all the other codecs are narrowband and works with signals sampled at 8 kHz. No consideration is taken to this fact. The signal is not upsampled to 16 kHz, and is just fed to the codec as if it was sampled at 16 kHz. This has the result that the bit rate will be twice as high when using AMR-WB compared to the other codecs working at 8 kHz with the same symbol pattern. To achieve the same bit rate, "stretched" symbols, that have twice as many samples, can be used.

There is no guarantee that, as with the modulator symbols and speech codec frames, the speech codec frames in two base station are synchronized. In the case of double transcoding between the modulator and demodulator, the speech frames processed by the two speech codecs may not be synchronized. To simulate the effect introduced by unsynchronized codecs, a random number of samples can be inserted before the signal is passed to the second codec.

4.3.2 Additional Distortion

Additional distortion can be applied to the PCM-waveform signal between the codecs in form of random noise, a DC offset and/or bit errors. The noise and the offset are just added to the signal which may overflow and the 16 bit value wrap around. Bit errors are introduced to the stream by flipping bits in the stream with a specified probability.

4.4 Demodulator

The demodulator, Figure 4.5, tries to demodulate the data transmitted by the modulator. To improve the demodulation result the input PCM signal can be preprocessed. The demodulator is much more computational heavy than the modulator due to the number of calculations needed for correlations.

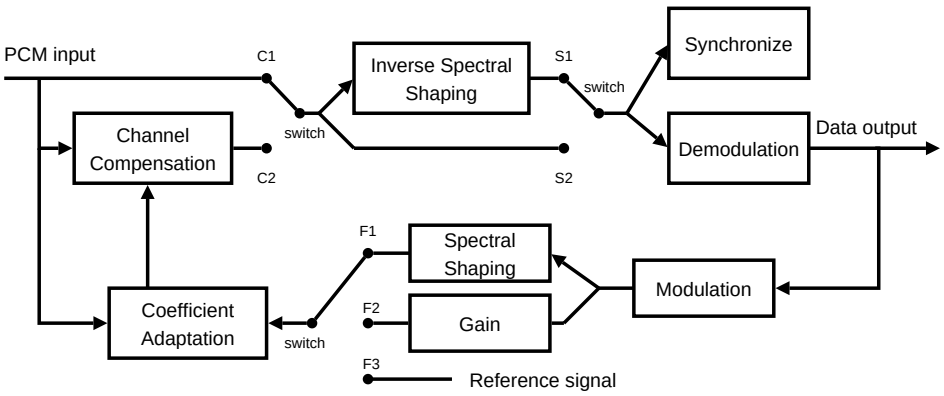

Figure 4.5. Demodulator framework.

4.4.1 Channel Compensation

There are two optional preprocessing steps of the input signal. Step one is to apply a channel compensation filter. The purpose of the filter is to counteract as much as possible of the distortion introduced in the channel.

The channel compensation filter is an adaptive filter. First the filter needs to be trained or have some initial coefficients. During the data transmission, demodulated data can be used to create a reference signal for the adaptation of the coefficients. When the data stream is used as reference signal for the filter coefficients adaptation, the data is first modulated and can then be filtered with the spectral shaping filter to create the waveform signal. The spectral shaping filter is only used if the modulator applies the filter. As an alternative to use demodulated data for adaptation, a reference file of the signal from the modulator can be used. This can be to prefer in testing purpose, since if there are lots of bit errors in the demodulated data, the adaptation of the coefficient will be bad and the whole demodulator performance will degrade. If the coefficients are not adapted in a correct way, the channel compensation filter may introduce more distortion and has the opposite effect of the intended purpose.

The adaptive filter implemented in the framework is a Normalized Least Mean Square (NLMS) filter. Least mean square adaptive filters are based on gradient techniques and are flexible, robust and easy to design [1]. There is also a NLMS filter implemented in the framework which updates the coefficients on blocks of samples instead of single samples. In this specific application no obvious degradation of the performance has been noticed when updating on blocks instead of samples and the computation requirement is reduced.

4.4.2 Inverse Spectral Shaping

Step two of the preprocessing is to apply the inverse function of the spectral shaping filter in the modulator. The filter should only be used if the spectral

shaping filter is used in the modulator. The functionality of the filter is to reverse the effect from the spectral shaping filter. A imperfect inversion of the spectral shaping filter is implemented as a FIR filter.

4.4.3 Synchronization

Before transmission of data occur, the synchronization procedure take place. Since it in the simulation is known that there will be a synchronization sequence in the input signal, the synchronization module cross-correlate a fixed predefined number of input samples in the beginning of the transmission with the predefined synchronization sequence. The sample sequence that has best match is used to synchronize the demodulator. The synchronization resolution is on sample level.

If the spectral shaping filter is used in the modulator, this is taken into account when synchronizing. The filter introduces some delays which need to be considered to avoid ending up in an unsynchronized state. Also the inverse spectral filter introduces some additional delays when used, which the synchronization operation also take into account.

The synchronization sequence could also be used as a training sequence for the channel compensation filter. In this case the compensation filter needs some initial coefficients that work for average channel conditions. This option is not implemented in the framework.

4.4.4 Demodulation

The demodulation module reads iteratively a number of samples from the input signal to be demodulated. Equal number of samples as the length of the symbol are read every iteration. The samples are correlated with all possible symbols in the symbol pattern and the best matching symbol is considered as the correct symbol. Bits corresponding to the symbol are written to the output data stream. The number of possible symbols varies with used symbol pattern. All possible reference symbols are pre-computed for faster correlations.

If the spectral shaping filter has been used in the modulator, there are two options for demodulation. First option is to use the inverse spectral shaping filter to restore the symbols. Second option is to apply the spectral shaping filter to the reference symbols used for correlation. Since the spectral shaping filter spreads out the symbols in the time domain, two sequential symbols may interfere with each other when some symbol patterns are used. In cases of symbol patterns with symbols that may interfere with each other it is usually an advantage to apply the inverse spectral shaping filter. In all other cases it is usually an advantage to apply the spectral shaping filter to the reference symbols when they are pre-computed. The inverse spectral shaping filter can not restore the signal perfectly.

4.5 Program Structure

The framework is implemented to make it easy to simulate and investigate different aspects of the data transmission. No regards have been made to implement a

computational efficient framework, focus has been on flexibility.

For simpler implementation the modulator works with 16 bit fixed point data, while the demodulator use floating point values. The modulator output is a 16 bit PCM stream which makes 16 bit data easy to work with while the demodulator using floating point value to eliminate the need of fix point scaling.

Filter coefficients for the spectral shaping filter and initial values for the channel compensation filter are read from ASCII-text files in runtime to make it easy to change the coefficients without recompiling the program.

Chapter 5

Improvements for Low Rate Speech Coded Channel

This chapter investigates different approaches to increase the performance of the data transmission for voice coded channels with low rate codecs. In Section 5.2 it is shown that the symbol pattern used in the work performed by University of Surrey gives a very high bit error rate if the codecs in the channel have low bit rate. The remaining of the chapter proposes different approaches to lower the bit error rate.

5.1 Simulations

All simulations have been performed with the implemented simulation framework described in Chapter 4. As input to the simulations random data has been used and the simulations have been repeated to get different random number of samples added to the streams. Presented results are the average of the repeated simulations and in some cases accompanied of the standard deviation of the series of simulations. If nothing is mentioned, all simulations have been performed without the spectral shaping filter and with the block NLMS channel compensation filter.

In all simulations the PCM output signal from the modulator is always encoded and decoded twice with the same codec, to simulate tandem connections, before the signal is passed on to the demodulator. The same codec is used twice since there would otherwise be to many different cases to simulate all combinations of codecs. In a call between two TETRA phones, the signal will in general only be encoded and decoded once, but to make the result more comparable with the other codecs the TETRA channel is also simulated with two transcodings. Calls with TETRA phones could also be encoded and decoded more than once if the call are between a TETRA phone and a phone on another network.

When the AMR and AMR-WB codecs have been evaluated in the simulations not all modes have been simulated. Simulating all modes would add rather limited information due to the fact that in general all simulations performed with the

codecs, a mode with higher bit rate gives a lower bit error rate compared to if the codec use a mode with lower bit rate. This is the case apart from a few exceptions.

The results presented are only simulation results which may not correspond to results from real conditions. In any case, the simulation results give some indication of how well a approach works and can be used to compare different approaches.

5.2 Surrey Symbol Pattern

The symbol pattern used at University of Surrey, see Section 3.1, works well for GSM EFR/AMR 12.2 kbps which can be seen in Table 5.1. Simulations with the AMR-WB codec has been performed with symbols that are stretched by adding a null sample between every original sample and also gives a low bit error rate for many of the AMR-WB codec modes. But Table 5.1 also shows that this symbol pattern is not very good for GSM FR and speech codecs with bit rates below 12 kbps. This symbol pattern will in the rest of the thesis be referred to as the Surrey pattern.

Table 5.1. BER for Surrey symbol pattern.

Codec (used twice)	BER	Standard deviation
GSM FR	28.6%	0.8%
GSM HR	44.4%	0.4%
GSM EFR/AMR 12.2 kbps	1.2%	0.4%
AMR 10.2 kbps	14.3%	0.3%
AMR 4.75 kbps	42.9%	0.5%
AMR-WB 23.85 kbps	0.3%	0.1%
AMR-WB 15.85 kbps	1.6%	0.2%
AMR-WB 12.65 kbps	5.0%	0.8%
AMR-WB 8.85 kbps	23.9%	2.0%
TETRA Codec	34.0%	1.0%

University of Surrey achieved a bit error rate of 2.9% [29] which is worse than the 1.2% for GSM EFR/AMR 12.2 kbps in the simulation and better than the simulation results of all the other GSM codecs (GSM FR, HR and AMR 10.2 kbps and lower). Most likely their cellular phones used the GSM EFR/AMR 12.2 kbps speech codec during their test as none of the other GSM codecs are close to a bit error rate of 2.9%. An explanation to the higher bit error rate for University of Surrey can be that they implemented a real system while this is simulation results. They had an analog connection between one computer and one cellular phone and may have experienced other distortions of the signal in the telecommunication channel not taken into account in the simulations. A further reason can be that no Voice Activity Detector (VAD) was used during the simulations which could degrade the result if frames were marked inactive.

If the bit error rate is plotted against the codec bit rate, Figure 5.1, it is easy to see that lower codec bit rate in general result in higher bit error rate. This result is expected since the ratio between the codec bit rate and the data bit rate decreases with lower codec bit rate. Fewer bits in the speech frames should encode the same amount of data bits which means that the data bits can not be represented equally good.

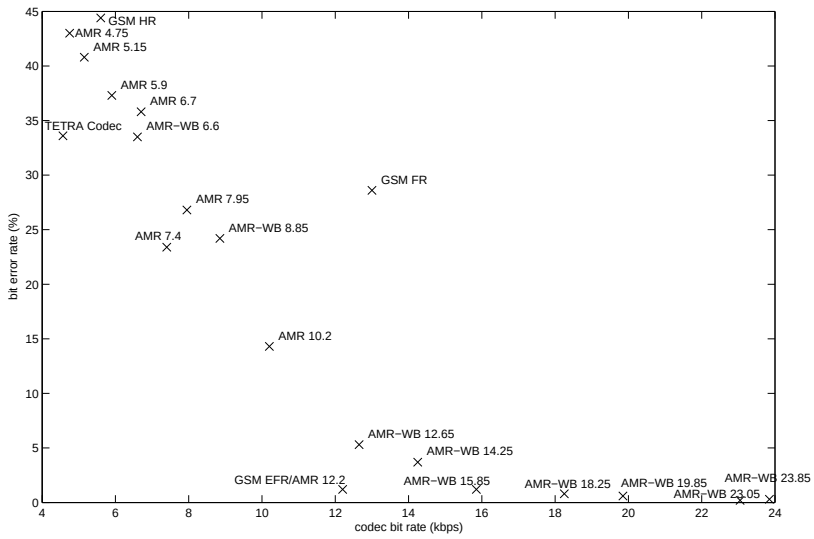

Figure 5.1. Bit error rate as a function of codec bit rate.

The following sections investigate different approaches to improve the bit error rate for speech codecs with lower bit rates. The high bit error rate experienced for some of the codecs makes it very hard to make any robust application that use the speech channel.

5.3 Pulse Position Data Encoding

The Surrey pattern encodes the data in the pulse positions. There are several changes that can be made to the symbol pattern to increase the robustness for speech codecs with lower bit rates. Most of the proposed changes lower the bit error rate but encode also fewer bits of data per symbol, resulting in reduced data transmission bit rate.

Since the Surrey pattern already gives a low bit error rate for GSM EFR/AMR 12.2 kbps and many of AMR-WB modes, only codecs with lower codec bit rates will be investigated. The applied changes will, in general, result in even lower bit error rates for these codecs also.

5.3.1 Number of Pulses

Most of the investigated speech codecs are based on ACELP technology and have an algebraic codebook containing a few pulses per speech frame. The Surrey pattern encodes data into 5 pulses which gives a low bit error rate for, among other codecs, AMR 12.2 kbps. AMR 12.2 kbps has a codebook containing 10 non-zero pulses for every subframe. The subframe length has the same length as a symbol contain 5 pulses. The AMR 4.75 kbps codec, on the other hand, has only 2 pulses per subframe in the algebraic codebook to encode the 5 pulses in each symbol. The two pulses are in most cases not enough for the codec to represent the five pulses which encode data and result in that the demodulator can not retrieve the data correctly.

Table 5.2 gives a list of ACELP codecs investigated in the thesis and the number of pulses in the algebraic codebook per subframe. The number of pulses is one of the most important parameters in the speech codecs to vary the codec bit rate.

Table 5.2. Algebraic codebook non-zero pulses per subframe in ACELP coders.

Codec	Number of pulses per subframe
GSM EFR/AMR 12.2 kbps	10
AMR 10.2 kbps	8
AMR 7.95/7.40 kbps	4
AMR 6.70 kbps	3
AMR 5.90/5.15/4.75 kbps	2
AMR-WB 23.85/23.05 kbps	24
AMR-WB 19.85 kbps	18
AMR-WB 18.85 kbps	16
AMR-WB 15.85 kbps	12
AMR-WB 14.25 kbps	10
AMR-WB 12.65 kbps	8
AMR-WB 8.85 kbps	4
AMR-WB 6.60 kbps	2
TETRA speech codec	4

Table 5.3 gives a modification of the Surrey pattern where 1 to 4 of the pulse positions tracks defined in Table 3.1 with belonging pulses are not used in the symbols. Fewer pulses per symbols result in lower data bit rate since each symbol carries less data.

Figure 5.2 shows the result of simulations with 1, 2, 3, 4 and 5 pulses for a few different codecs. By reducing the number of pulses in the symbol, the bit error rate is also reduced and it is possible to achieve a more robust transmission. All the codecs show a clear trend that fewer pulses give lower bit error rate. Some of the codecs still give a high bit error rate, but the error rates have been considerably improved by reducing the number of pulses. For some applications and codecs this change is enough to achieve a low enough bit error rate while some codecs (modes) still need improved bit error rate.

Table 5.3. Surrey pattern with reduced number of pulses.

Number of pulses	Bits per symbol	Number of symbols	Data bit rate	Tracks
5 (Surrey)	15	32 768	3.0 kbps	1, 2, 3, 4, 5
4	12	4 096	2.4 kbps	1, 2, 4, 5
3	9	512	1.8 kbps	1, 3, 5
2	6	64	1.2 kbps	1, 3
1	3	8	0.6 kbps	1

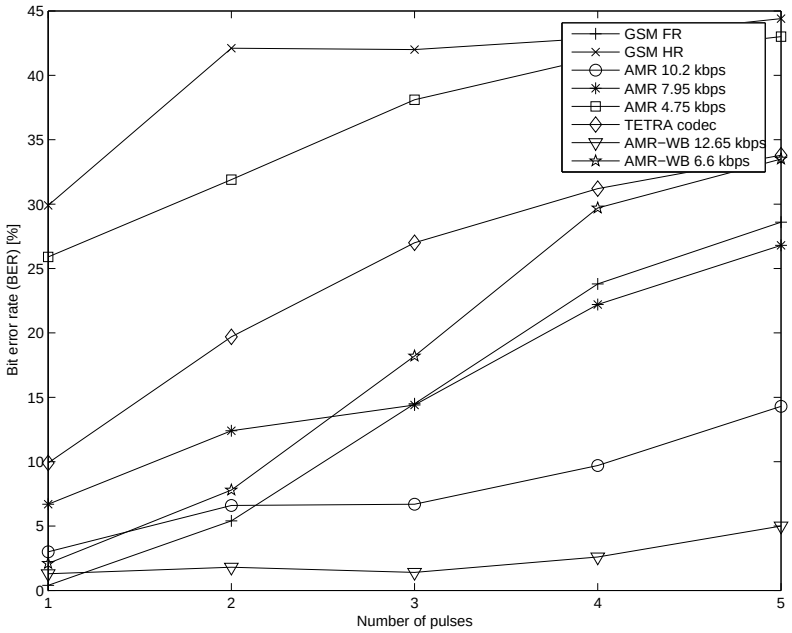

Figure 5.2. Bit error rate as a function of the number of pulses in the symbols.

An additional advantage, as a side effect from the reduction of bits per symbol, is that much fewer correlations is needed for the transmission of a certain amount of data since there are fewer different symbols to be considered. This makes the requirements of the demodulation hardware lower.

A disadvantage of using fewer pulses in the symbols is that the activity in the frames may be lower. The lower activity in the frame increases the possibility that the Voice Activity Detector (VAD) in the codec mark the frame as inactive, cut it out and the data is lost.

5.3.2 Distance between Pulse Positions

Fewer pulses are not the only reason to the improved result in the previous section. If the same symbol length is kept and in the same time reducing the number of pulses the, average distance between the pulses increase. If fewer than five pulses are used in the Surrey symbol pattern there are sample positions in the symbol which are never used to place pulses. For example, if only one or two pulses are used, there are no neighboring pulses in the signal at all; there is always at least one null sample between two pulses.

When the distance between the pulse positions increase, at least to a certain limit, the risk of that wrong symbols are demodulation is reduced. With longer distance between possible pulse positions, it is less likely that wrong symbols have high correlations with the input signal.

There are several other different ways to increase the average distance between two pulses in the symbols. All changes to the symbol also change the symbol in some other way which makes it hard to see only how the distance affects the performance of the transmission. For example if the symbol length is increased the distance between the pulse positions will increase but also the number of speech codec frames used to encode each symbol.

5.3.3 Number of Pulse Positions

Another way to improve the bit error rate by changing the Surrey symbol pattern is to change the number of pulse positions in each track. This change will, as reducing the number of pulses, increase the minimum distance between pulses and is part of the reason why this change improve the robustness of the data transmission.

Two possible changes to the Surrey pulse positions which creates two new symbol patterns are to only keep every second pulse position (2 bit per pulse) and to keep only every fourth pulse position (1 bit per pulse) of the 8 pulse positions in each of the five tracks, as shown in Table 5.4.

Table 5.4. Surrey pattern with reduced number of pulse positions.

Bits per pulse	Pulse positions per track	Bits per symbol	Number of symbols	Data bit rate
3 (Surrey)	8	15	32 768	3.0 kbps
2	4	10	1 024	2.0 kbps
1	2	5	32	1.0 kbps

It is possible to see positive trends for the bit error rates for all codecs except AMR 10.2 in Figure 5.3 when the number of pulse positions is reduced. The improvement of the bit error rate is not as effective as reducing the number of pulses. Most codecs give higher bit error rate for 1 bit per pulse symbol pattern which gives a data bit rate of 1 kbps compared to the 2 pulses symbol pattern with a data bit rate of 1.2 kbps. This mean that it in first hand is better to reduce

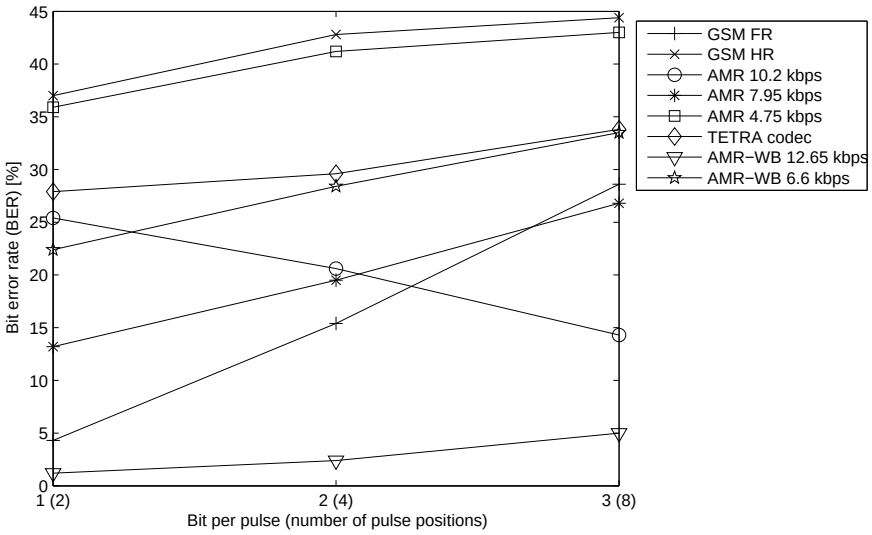

Figure 5.3. Bit error rate as a function of the number of pulse positions.

the pulses instead of reducing the pulse positions as it in most cases gives both lower bit error rate and higher data bit rates. There is a limit in how many pulses low rate codecs are capable of encode and represent in an accurate way. This also reflects that the pitch periods in speech are important and are encoded accurate by the codecs.

For the AMR 10.2 kbps codec the result from reducing the number of pulse positions is the opposite of the desired goal. This is most likely due to the different mapping of pulse positions onto tracks for this mode, see Table 2.4. AMR 10.2 kbps use four tracks in the algebraic codebook compared to five for the other AMR modes. If the number of positions are reduced in the same fashion as in the simulation, only two respective one of the four tracks in the AMR 10.2 kbps codec are used. This is a result of the fact that every fourth sample is in each track and if for example only every fourth sample position is used only one track will be used. With more pulses in the tracks, the codec is not able to represent the symbols equally accurate and the bit error rate increases.

Even with the original Surrey pattern all five pulses can be placed in one of AMR 10.2 kbps codec's track. But the probability that all five pluses are placed in the same track for random input data is much lower and therefore the impact on the average bit error rate is lower. This problem can be overcome by assigning the pulses to different codec tracks.

The number of pulses and the number of pulse positions per track are independent (at least to some degree) and changes to both can be combined. An even greater robustness can be achieved if the number of pulses is reduced as well as the number of sample positions.

5.3.4 Spectral Shaping Filter

Some combinations of symbol patterns and codecs benefit from always applying the spectral shaping filter to the sequence of symbols in the modulator and to the reference symbols in the demodulator. Applying the filter has in this case nothing to do with avoiding cut outs caused by the VAD. By applying the spectral shaping filter, more frequencies will be in the frequency band the codecs encode and not disappear during the transmission in the channel.

The spectral shaping filter replaces all the pulses by a feature. Filtered symbols will have more non-zero samples which make the correlation in the demodulator less dependent of a few non-zero samples to be transmitted correctly.

A symbol pattern that achieve better, if the spectral shaping filter is applied for some of the codecs, is a pattern taking advantage of the earlier mentioned improvements for low rate speech coded channels. Each symbol, 40 samples long, encode 4 bits in two pulses with each four pulse positions. Pulse 1 will always have a positive sign and pulse 2 will always be negative. Pulse positions and signs are given in Table 5.5. The symbol pattern will have a data bit rate of 800 bps.

Table 5.5. Pulse tracks and signs for a two pulses symbol pattern.

Pulse	Pulse positions	Sign
Pulse 1	0, 10, 20, 30	positive
Pulse 2	5, 15, 25, 35	negative

Example 5.1

If the data `1101` should be transmitted on a narrowband channel the first pulse should encode the data `11`, which place a positive pulse on pulse position 30 and the second pulse should encode the data `01`, which place a negative pulse on pulse position 15. Figure 5.4 shows the symbol.

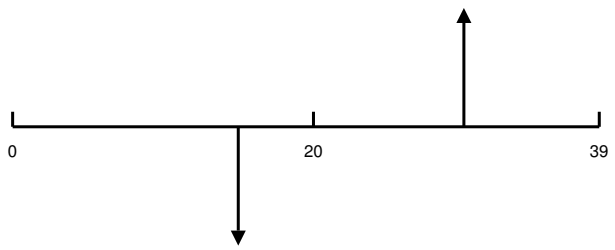

Figure 5.4. Example of the two pulses symbol.

Using the symbol for wideband codecs as AMR-WB extra null samples can be inserted between the pulse positions. Pulse positions for AMR-WB are the pulse positions specified in Table 5.5 multiplied with 2.

As Table 5.6 shows not all codecs benefit from the usage of the spectral shaping filter. The benefits vary very much from codec to codec. In some of the AMR modes, the bit error rates are reduced significantly, while for other codecs, like the GSM FR, the spectral shaping filter degrades the performance instead.

Table 5.6. Two pulses symbol pattern without and with applying spectral shaping filter.

Codec	Without	With
GSM FR	4.3%	7.8%
GSM HR	36%	29%
AMR 10.2 kbps	5.5%	2.6%
AMR 7.95 kbps	23%	13%
AMR 4.75 kbps	33%	33%
AMR-WB 12.65 kbps	8.4%	10%
AMR-WB 6.6 kbps	13%	20%
TETRA Codec	23%	18%

The wideband codec modes (AMR-WB) don't perform very well with the spectral shaping filter. A reason for this can be that they are wideband codecs and the filter is suited for narrowband codecs. The filter concentrates the frequencies between 0 and 4000 Hz for narrowband codecs which scales to 0 and 8000 Hz for wideband signals. Wideband codecs encode the frequency band 50 to 7000 Hz and the AMR-WB codec only encodes 50 to 6400 Hz frequency band. The higher frequency band, 6400-7000 Hz, is only reconstructed using parameters from the lower band. Another spectral shaping filter which concentrates the frequencies to a more proper frequency range could improve the performance for this codec. Due to lack of time no spectral shaping filter for wideband codecs were implemented.

5.3.5 Wide Pulses

Symbol patterns with "unused" samples (always zero) between possible pulse positions can be modified to have pulses which are more than one sample wide. The wider pulses give more "weight" during the analysis-by-synthesis stage of the encoding process where the pulses should be represented by the speech codec parameters. More similarity between the original signal and the signal transmitted through the speech coded voice channel results in higher correlation for the correct symbol at the demodulator.

Wider pulses also result in reference symbols in the demodulator with fewer non-zero samples. If the symbols contain a low number of pulses and each pulse is one sample wide, the correlation result will depend on only a few samples in the signal which may be reflected in a more uncertain outcome.

Making the pulses wider is an approach which not lowers the bit rate of the data transmission. The number of pulses and the number of pulse positions can

be maintained under the condition that there are unused samples between pulse positions. Since the data bit rate is not affected in a negative direction by making the pulses wider, this approach can always be used if it is possible and improves the robustness of the transmission.

The pulses can not be too wide. All ACELP codecs starts after a certain width to get higher bit error rate again. Different codecs have different optimal widths as will be shown below.

Most codecs perform better with wide pulses if the spectral shaping filter is applied. The wider pulses lead to higher bit error rates for some of the codecs if the filter is not used and some codec, as with 1 sample wide pulses, perform worse with the filter applied.

One symbol pattern, which improves from wider pulses, is the two pulses symbol pattern described in previous section, Section 5.3.4. The symbol pattern has four (nine for the wideband) non-used samples between every pulse positions which can be utilized for wider pulses. All pulses can be changed to be one to five samples wide.

The results from the simulations are shown in Figure 5.5. The spectral shaping filter has been used in all simulations except the simulations of the GSM FR, HR and AMR-WB 6.6 kbps codec. These codecs perform better without the filter.

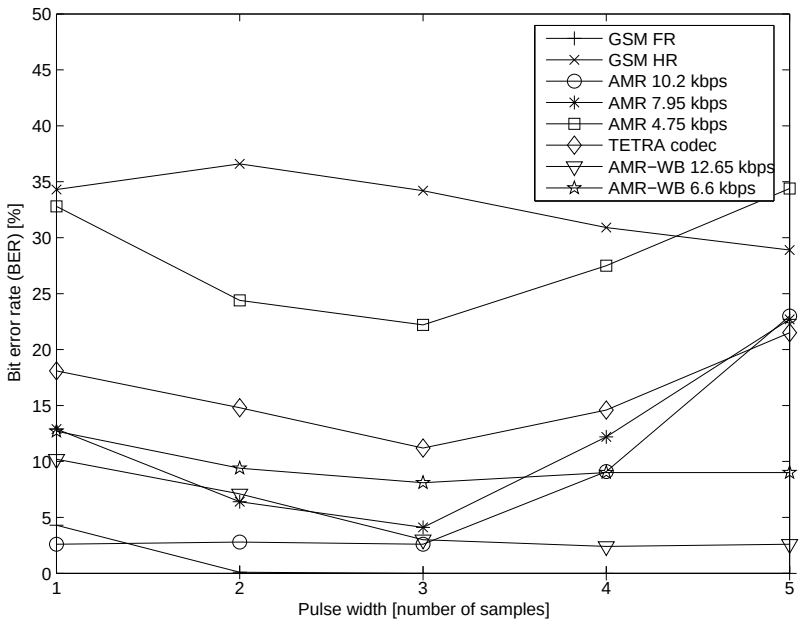

Figure 5.5. Bit error rate as a function of the width of pulses in the symbol.

A pulse width of three gives the lowest bit error rate for most of the codecs. Best width for the wideband codec modes are also three or a few samples more, reason for this can be that there are more non-used samples between and/or that wideband codecs can encode wider frequency spectrum. The AMR and TETRA speech codecs should not be used with wider pulses than three samples since the performance rapidly decreases and soon gets worse than using pulses of width one.

GSM FR codec gets very low bit error rate if the pulse is three samples wide or wider. The codec regularly samples every third sample to use as excitation vector and if the pulses are three samples or wider at least one sample of each pulses will be part of the excitation vector. The somewhat strange behavior for the GSM HR codec is probably explained by the different codebook structure used for the codec.

5.3.6 Pulse Redundancy

One rather obvious way to improve the bit error rate is to introduce redundancy in the transmitted signal. If two pulses are used to encode every group of bits instead of one pulse the performance will increase. During the demodulation the extra information transmitted with redundant pulses can be used to make a better decision regarding which symbols that have been transmitted.

The negative side of introducing pulse redundancy is a lower data bit rate. Using twice as many pulses to encode the data will half the possible data bit rate for the system. Pulse redundancy is only recommended at severe conditions or at transmission of very small amounts of data.

Two different symbol patterns and a total of six different variants of these patterns were used to simulate the effect of redundancy. The first symbol pattern encodes 2 bit of data in one pulse located in one of four different pulse positions in the first half of symbol. To introduce redundancy a second pulse is placed in the corresponding pulse position in the second half of the symbol with a negative sign. The symbol is repeating itself but with a negative pulse instead of a positive. For simulations two different lengths, 40 and 64 samples, of the symbol have been used which give a distance between the center of the pulses of 5 and 8 samples and data bit rates of 400 bps respective 250 bps. In both cases the pulses have a width of 3 samples. Pulse positions for the 40 sample long symbol pattern are shown in Table 5.7.

Table 5.7. Double pulse symbol pattern.

Pulse	Pulse positions				Sign
Pulse 1	0-2	5-7	10-12	15-17	positive
Pulse 2	20-22	25-27	30-32	35-37	negative
Data bits	00	01	10	11	-

Second pattern also uses two pulses to encode one group of bits. But instead of placing the pulses after each other, the pulses are placed in opposite pulse positions

and have opposite sign. If the first pulse is placed in the second pulse position with positive sign, the second pulse is placed in the second last pulse position with negative sign. Both pulses have a width of three samples. This symbol pattern has been simulated with encoding 2 bit (4 pulse positions) and 3 bit (8 pulse positions) into the pulses and symbols with length 40 and 64 samples.

Example 5.2

If the 3 bit `110` of data that should be transmitted with the two opposite pulses symbol pattern with symbols of length 64 samples and 8 pulse positions, there should be a positive pulse on sample 48-50 and the negative pulse on sample 8-10. The example symbol is shown in Figure 5.6.

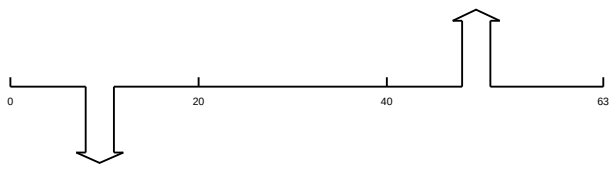

Figure 5.6. Example of the opposite pulses symbol.

Table 5.8 gives a summary over some properties for the simulated pulse redundant symbols and Table 5.9 gives the bit error rates from the simulations. All codecs have been simulated with the spectral shaping filter. Only codecs with low bit rate have been simulated since redundancy is not necessary for most other codecs.

Table 5.8. Summary of simulated pulse redundant symbol patterns.

Pattern	Symbol length	Pulse positions	Bits per symbol	Number of symbols	Data bit rate
Double pulse 1	40	4	2	4	400 bps
Double pulse 2	64	4	2	4	250 bps
Opposite pulses 1	40	8	3	8	600 bps
Opposite pulses 2	40	4	2	4	400 bps
Opposite pulses 3	64	8	3	8	375 bps
Opposite pulses 4	64	4	2	4	250 bps

With these patterns there are possibilities to achieve rather low bit error rates even for the codecs with the lowest bit rates. The opposite pulses pattern gives slightly better performance than the double pulse pattern. Longer distance between pulse positions is probably one reason which gives the opposite pulses pat-

Table 5.9. Bit error rate for simulated pulse redundant symbol patterns.

	GSM HR	AMR 5.15	AMR 4.75	TETRA Codec
Double pulse 1	24%	8.2%	12%	3.6%
Double pulse 2	12%	4.3%	5.4%	1.1%
Opposite pulses 1	26%	10%	13%	3.0%
Opposite pulses 2	15%	8.0%	9.8%	3.3%
Opposite pulses 3	16%	4.5%	6.5%	0.9%
Opposite pulses 4	11%	3.4%	4.8%	0.7%

tern the advantage. The GSM HR performs better with the opposite pulses 4 pattern if no spectral shaping filter is used. In that case a bit error rate of 8.7% is achieved.

The codec causing most problems is the GSM HR codec which have the highest bit error rate of all the codecs. GSM FR and GSM HR have different excitation vectors techniques compared to the other codecs and the one used for GSM HR is not as suitable for coding information in pulse positions as the for example ACELP is.

If even lower bit error rates are needed, further redundancies can be introduced to the cost of even lower data bit rates. If only a few bytes of data should be transferred every time, this may be the way to ensure that the data arrives.

5.3.7 GSM FR Specialized Symbol Pattern

GSM Full Rate (FR) speech codec has a quite different coding technique compared to the other investigated speech codecs. The codec is a Regular Pulse Excitation (RPE) and sample the input signal regularly. Only every third sample of the input signal is transmitted which makes it difficult to use all samples in a symbol to encode the data, which is the case with the Surrey pattern.

The special structure of the FR codec can be exploited to achieve very low bit error rate for the FR codec by using only every third sample to place pulses on. Synchronization between the symbol sequence and the FR codec is not necessary, in every subframe the codec has four 13 samples sequences which are sampled every third sample and select the sequence which is the one with the maximum energy. Since only every third sample is used, the probability that the pulse positions will be transferred correctly is very high.

One possible symbol pattern is to place eight 1 bit pulses in 48 samples long symbols. One pulse position is located every third sample and each pulse is placed in one of two positions depending on the bit. This symbol pattern will give a data bit rate of 1.33 kbps for the narrow band speech codecs. Table 5.10 shows the simulation result for some codecs.

Although this symbol pattern gives a very low bit error rate for the GSM FR codec it will be hard to utilize in a system with an unknown speech codec since the result for the other codecs are moderately good. If there are some way to enforce

Table 5.10. BER for RPE specialized symbol pattern.

Codec	BER
GSM FR	0.0%
GSM HR	38%
AMR 7.95 kbps	21%
AMR 4.75 kbps	38%
TETRA Codec	28%

or ensure that the FR speech codec is used in the whole channel this pattern can be a good choice.

5.4 Other Data Encodings Approaches

During the work some other simple approaches which didn't encode the data in the pulse positions were also simulated. None of the approaches have been more deeply investigated and most of them were rejected after a first simulation due to bad performance compared to the pulse positions data encoding. None of the approaches could achieve anything close in performance to the pulse position coding for higher data bit rates. However, for very low data rates the pulse sign pattern gave relatively low bit error rate.

5.4.1 Pulse Sign Encoding

The pulse sign pattern encoded the data in the sign of pulses equally distanced from each other in the symbols. Each pulse encoded one bit of data by making the pulse positive if the bit was `1` and negative if the data bit was `0`.

The GSM Half Rate (HR) speech codec doesn't perform very well with the pulse positions encoding but can achieve a relative low bit error rate with pulse sign encoding, under the circumstance that the pulses are not too close to each other. Since only one bit is encoded in each pulse and the pulses is relatively far apart the data bit rate is low.

Figure 5.7 shows simulation results with the HR codec with signed pulses with 10, 20, 30 and 40 samples apart. These distances give data bit rates of 800 bps, 400 bps, 267 bps and 200 bps, respectively, for narrowband codecs. For comparison, two AMR modes are also in the figure. HR has been simulated without the spectral shaping filter while the simulations with AMR have applied the filter. However, the AMR codec can achieve high data bit rates with pulse position encoded data while still remaining relatively low bit error rates. With shorter distance between the pulses, the bit error rate increase rapidly and pattern become unusable for all three of the codecs.

The simulation results vary very much from simulation to simulation for a GSM HR codec channel. This big variations make it hard to really know how good this

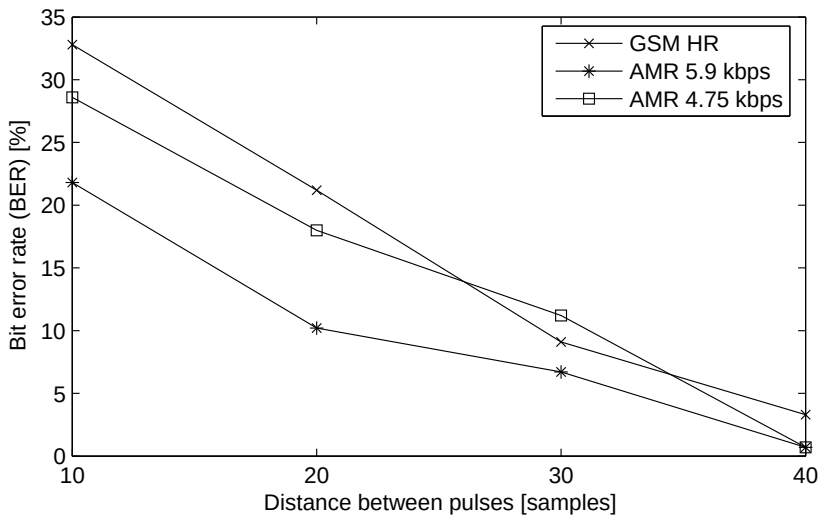

Figure 5.7. Bit error rate as a function of the distance between two pulses in the pulse sign symbol pattern.

symbol pattern is. Some simulations with a GSM HR channel and signed pulses with 40 samples between the pulses give almost no errors while others simulations have a bit error rate over 10%. The standard deviation for the simulations is 4.2%.

This symbol pattern may benefit from some other demodulation than correlate all symbols with reference symbols. The performance may increase if the demodulation just looked on the signs for the sample where the pulses should be located and a few samples around instead of correlating the whole symbol since only the signs are important, not the positions.

5.4.2 Sinusoid Waves

A further symbol pattern simulated encoded the data into sinusoid waves in the symbols. The sinusoid waves had frequencies between 600 Hz and 2600 Hz. Each of the frequencies corresponded to one bit pattern of data to be transmitted. This approach performed very bad and was almost immediately rejected. In likeness with the pulse sign pattern this pattern will probably benefit from another demodulation technique. If the demodulator looks on the frequency component of the received signal instead of correlate the signal the performance would most likely increase. A more deeper investigation should probably be done before the pattern is totally rejected.

Chapter 6

Robustness against Channel Distortions

During the transportation of the speech from one cellular phone to another cellular phone there can be other circumstances than just the speech codecs which introduce distortion to the signal. In this chapter there are a few simulations of how these distortions affect the data transmission. The simulations are performed in the same way as for the improvements for low rate speech coded channel described in Section 5.1 and the referred symbol patterns are the same as found in Chapter 5.

6.1 Voice Activity Detector

The Voice Activity Detector (VAD) does not really introduce any distortion and it is a part of the speech codecs. However, the VAD functionality of the codecs will degrade the performance of the data transmission if the speech frames are marked inactive and not transmitted. Inactive frames are replaced with comfort noise and almost all information is lost. Some codecs update parameters for the comfort noise during the silent periods, these parameters contain rather little information and will be hard to utilize for the data transmission.

Periods marked as inactive by the VAD must be avoided by modulation schemes which can "fool" the codec to believe that it is speech transmitted and not background noise. Table 6.1 shows some of the symbol patterns from previous chapter and the bit error rate with and without the VAD activated. Both VAD options, AMR1 and AMR2, have been simulated for the AMR codec.

The result shows the VAD activated only degrade the performance slightly for all codecs except for the GSM HR codec and when the AMR2 option is used for the AMR codec. If the spectral shaping filter and the inverse spectral shaping filter are applied when simulating the Surrey pattern with AMR 12.2 kbps and the AMR2 VAD option, the bit error rate goes down to 32%. The bit error rate is much lower but it is still too high to be useful for any application. For the GSM HR codec applying the spectral shaping filter doesn't improve the performance.

Table 6.1. BER degradation caused by VAD.

Symbol pattern	Spectral shaping	Codec [(VAD opt.)]	BER no VAD	BER with VAD
Surrey pattern	no	GSM EFR	1.1%	2.7%
Surrey pattern	no	AMR 12.2 (AMR1)	1.2%	1.5%
Surrey pattern	no	AMR 12.2 (AMR2)	1.3%	48%
Surrey pattern (stretched)	no	AMR-WB 15.85	1.6%	2.1%
Two pulses 3 sample wide	yes	AMR 7.95 (AMR1)	4.1%	4.7%
Two pulses 3 sample wide	yes	AMR 7.95 (AMR2)	4.1%	50%
Opposite pulse 4	yes	AMR 4.75 (AMR1)	4.8%	4.8%
Opposite pulse 4	yes	AMR 4.75 (AMR2)	4.8%	50%
Signed pulses (200 bps)	no	GSM HR	3.0%	50%

Due to lack of time no further investigation to improve the performance for the GSM HR and AMR2 have been conducted. A similar approach as Surrey used, which apply the spectral shaping filter only 20 ms of 80 ms or maybe the opposite, applying the filter 60 ms of 80 ms, may improve the performance for these codecs and VADs.

6.2 Bit Errors

Between two base stations, bit errors can be introduced to the PCM signal. The errors change the PCM waveform, how much depends on the importance of the erroneous bits. These bit errors will decrease the performance of the data transmission as the signal will be less similar to the original signal. An increase of bit errors in the PCM channel will result in an increase of the bit error rate of the data transmission.

The expected increase in bit error rate is confirmed with the simulation results in Table 6.2. These simulations have been performed with a bit error probability of 0.1% in the PCM channel between the base stations.

A bit error probability of 0.1% gives the probability of $1-(1-0.001)^{16} = 1.59\%$ that at least one bit in each 16 bit PCM sample is toggled. All simulations except GSM HR increase less than 1.59% in bit error rate which shows that a bit error in a pulse encoding data not necessary lead to an incorrect decoded symbol.

Table 6.2. BER for a PCM channel with a probability of 0.1% bit error.

Symbol pattern	Spectral shaping	Codec	BER error free channel	BER, channel with errors
Surrey pattern	no	AMR 12.2	1.2%	1.8%
Surrey pattern (stretched)	no	AMR-WB 15.85	1.6%	2.7%
Two pulses 3 sample wide	yes	AMR 7.95	4.1%	5.0%
Opposite pulse 4	yes	AMR 4.75	4.8%	5.7%
Signed pulses (200 bps)	no	GSM HR	3.0%	4.3%

6.3 Lost Speech Frames

Lost speech frames will result in high bit error rate since the data transmitted also will be lost. Lost frames are substituted with the previous good frame or an extrapolated frame which help improve the quality for a regular speech conversation. However, this error concealment techniques used by the codecs don't help for data transmission.

One lost frame is 20-30 ms and corresponds to around 2-4 lost symbols. There is not much that can be done with the modulation or the symbols to overcome this problem except make very long symbols, which anyway only work if the subsequent number of lost frames is not too long. Some of the effects by lost speech frames can probably be reduced by interleaving of data and channel coding. The loss of speech frames have not been simulated as there is no obvious approach to improve the modulation or/and demodulation for better performance. The increase in bit error rate should be proportional to the number of lost frames.

6.4 Analog PCM Errors

Most traffic in mobile communication networks of today is digital. However, a voice call can be transmitted over an analog link between two base stations and that analog link can introduce extra distortion. An analog connection also exist in the channel if the speech coded voice channel modem is connected to the cellular phone with an analog connection, as is the case at University of Surrey. Among other things, random noise and a DC offset can be introduced to the signal and these two distortions have been simulated to see the modulation robustness against these kind of errors.

Both errors have been simulated to be a maximum of 10% of the maximum value of the PCM signal. The distortions have been applied between the two speech codecs in the channel. The random noise is applied as a random value

between +10% and -10% of the maximum value to each sample. This is maybe not the way random noise behaves like in real systems, but gives some indication of the robustness against random noise. For the DC offset, 10% of the maximum value is added to each sample. Table 6.3 and Table 6.4 contain the simulation results.

Table 6.3. BER for a noisy analog PCM channel.

Symbol pattern	Spectral shaping	Codec	BER noise free channel	BER, noisy channel
Surrey pattern	no	AMR 12.2	1.2%	2.0%
Surrey pattern (stretched)	no	AMR-WB 15.85	1.6%	3.6%
Two pulses 3 sample wide	yes	AMR 7.95	4.1%	4.9%
Opposite pulse 4	yes	AMR 4.75	4.8%	6.8%
Signed pulses (200 bps)	no	GSM HR	3.0%	4.2%

The noisy channel simulated increase the bit error rates more than the bit error of a probability of 0.1% do. This also shows the advantage of having a digital interface between the speech coded voice channel modem and the cellular phone.

Table 6.4. BER for an analog PCM channel with a DC offset.

Symbol pattern	Spectral shaping	Codec	BER DC offset free channel	BER channel with DC offset
Surrey pattern	no	AMR 12.2	1.2%	1.8%
Surrey pattern (stretched)	no	AMR-WB 15.85	1.6%	1.6%
Two pulses 3 sample wide	yes	AMR 7.95	4.1%	4.2%
Opposite pulse 4	yes	AMR 4.75	4.8%	4.5%
Signed pulses (200 bps)	no	GSM HR	3.0%	2.3%

Adding a DC offset to the signal decreases the performance slightly from some of the simulations while the bit error rate is unchanged for some. The DC offset has in most cases very little effect of the overall performance and in most cases

not a really big problem as the symbol patterns are robust against this kind of error. Some codecs and patterns perform slightly better with the offset, but the differences are within the standard deviations from the different simulations. A reason why a DC offset doesn't affect the performance is probably that correlation with all reference symbols is used to demodulation which is relatively insensitive to DC levels in the input.

Chapter 7

Conclusions and Further Studies

This chapter sums up the results from the conducted work. A section is also dedicated to suggestions of further possible studies which may improve data transmission over a speech coded voice channel.

7.1 Conclusions

A number of different possible symbol patterns have been presented and simulated with different results. The simulated voice channel has been built up of different codecs to evaluate the data transmission during different channel conditions which can appear during a regular voice call.

The simulations with the Surrey pattern showed that it only will work with the codecs GSM EFR, AMR 12.2 kbps and most modes of AMR-WB codec. To improve the robustness of the data transmission for channel conditions with other speech codecs different methods are needed. The speech codec (mode) with lower bit rates caused the biggest problem and specially the GSM HR codec. For these codecs, symbol patterns which give much lower data transmission rates must be used to achieve a low bit error rate.

To improve the robustness for low rate speech codecs this thesis present several different methods. Encoding the data in pulse positions seems to be the best approach for most codecs. The robustness of the Surrey pattern is in most cases improved if the pattern is changed in some of the following ways:

- reduce the number of pulses
- increase distance between pulse positions
- wider pulses
- applying the spectral shaping filter
- reduce the number of pulse positions

- introduce pulse redundancy

The most important change to the Surrey pattern is to reduce the number of pulses in the symbol.

The choice of symbol pattern must probably be selected depending on the application the data transmission should be used for. If only a small amount of data should be transported maybe a more reliable symbol pattern for all conditions should be selected. In good channel conditions this pattern will give low bit error rate and in more severe conditions there will at least be some possibilities to transport the data. On the other hand in some real time application, maybe a minimum data bit rate is required and therefore a less robust pattern should be selected, and make the service unavailable during severe channel conditions.

To achieve both a system robust against errors and high data transmission rates during good conditions, it is probably best to introduce some adaptation of the symbol pattern. In this case the system must in some way measure the conditions of the channel and select the symbol pattern on the basis of these measurements. With adaptation to the channel conditions the system can provide a system with good data bit rates at good conditions and still guarantee that the data reaches the receiver during more severe conditions. Push-To-Talk (PTT) is an example of where this approach could be useful. The message will always arrive and if the conditions are good, the receiver doesn't need to get annoyed because of extra long delay before the message arrives.

Most likely no special case needs to be done for the AMR-WB codec. All three network GSM, UMTS and TETRA, support narrowband codecs. If a digital interface is used, e.g. the Bluetooth handsfree profile, there must be some indication if wideband speech codecs are used or some way to select between the narrowband and the wideband speech codecs since wideband codecs require twice the sample rate on the input data. However, there can be advantages in using the AMR-WB to achieve high data bit rates and lower bit error rates.

In the thesis there are also some simple evaluations of the effects of distortion on the channel. If the signal is distorted, the performance of the data transmission will degrade. However, most distortions have a moderate effect on the bit error rate for the simulated symbol patterns. The AMR2 VAD option in the AMR codec and the HR VAD may cause big problems and need further investigation to find a combination of modulation technique and symbol pattern which will not mark the speech frames carrying data as inactive to improve the performance.

7.2 Further Studies

AMR2

As mentioned in the conclusion section, the AMR2 voice activity detector can be a problem. A further study could look closer to this problem and try to find a way to modulate the data, filter the signal and/or a symbol pattern which the AMR2 VAD will not classify as inactive speech frames while still maintaining a good data bit rate.

Channel Coding

To be able to use the system in a real application there must most likely also be added some kind of channel coding before the modulation. If the demodulation and the channel decoding are combined by using some kind of matching metric for the correlation between each of the reference symbols and the input signal which is used together with the channel decoding the demodulation should perform better. For example, a Viterbi decoder could be implemented.

Synchronization

The simulation framework has a very simple synchronization mechanism implemented. This synchronization need to be improved/replaced before it can be used in a real system to overcome problems not existing in the simulations. Some time need to be spent on investigating a good approach to synchronize the modulator and demodulator. It may also be interesting to have some kind of continuous synchronization during the data transmission, similarly to the one Surrey use, to resynchronize if the two sides get unsynchronized. This is extra interesting if some analogue links in the transmission channel are expected.

Adapt to Channel Condition

Different codecs in the voice channel permit rather different data bit rates to be achieved with a not too high bit error rate. If, as mentioned in the conclusion, the symbol pattern could be chosen on basis of the current conditions, a better data transmission service could be provided. The adaptation could be either a procedure in the beginning of the transmission or something that is adapted during the whole data transmission.

Higher Data Bit Rates

This thesis has focused on increasing the robustness of the data transmission over a speech coded voice channel. In some applications, higher data bit rates are required and a further study could investigate approaches to achieve that. Higher data bit rates should not be impossible, at least not for the AMR-WB codec.

Real Performance

The findings of this thesis have only been simulated and not tried on a real system. An interesting study would be to implement and test the system on a real network. This could lead to insight about problems not foreseen or encountered during simulations.

VoIP Codec

Telephone calls over Internet get more and more popular which make an investigation of how well it works with data transmission over voice channels going over Internet. This will lead to some other codecs and maybe other conditions to adapt to.

Demodulation efficiency improvements

The demodulation is performing an exhausted correlation with all reference symbols to find the correct symbol. This is very time and computationally consuming if there are many reference symbols. If the demodulation could be improved to perform a not exhausted search, less hardware should be required to run the system in real time.

Bibliography

[1] M. G. Bellanger. *Adaptive Digital Filters*. Marcel Dekker Incorporated, second edition, 2001. ISBN 0-8247-0563-7.

[2] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen. The adaptive multirate wideband speech codec (amr-wb). *IEEE Transactions on Speech and Audio Processing*, 10(8):620–636, November 2002.

[3] I. Boyd. Speech coding for telecommunications. *Electronics & Communication Engineering Journal*, 4:273–283, October 1992.

[4] J. Degener and C. Bormann. Gsm 06.10 lossy speech compression. `ftp://ftp.cs.tu-berlin.de/pub/local/kbs/tubmik/gsm/gsm-1.0.10.tar.gz`.

[5] E. Ekudden, R. Hagen, I. Johansson, and J. Svedberg. The adaptive multirate speech coder. In *IEEE Workshop on Speech Coding Proceedings, 1999*, pages 117–119, 1999.

[6] ETSI. *European digital cellular telecommunications system; Half rate speech Part 3: Substitution and muting of lost frames for half rate speech traffic channels (GSM 06.21), ETS 300 581-3*, November 1995.

[7] ETSI. *European digital cellular telecommunications system; Half rate speech Part 6: Voice Activity Detector (VAD) for half rate speech traffic channels (GSM 06.42 version 4.1.1), ETS 300 581-6*, November 1995.

[8] ETSI. *Digital cellular telecommunications system (Phase 2); Voice Activity Detector (VAD) for Enhanced Full Rate (EFR) speech traffic channels (GSM 06.82 version 4.0.1), EN 301 249*, December 1997.

[9] ETSI. *Digital cellular telecommunications system (Phase 2); Full rate speech; Part 3: Substitution and muting of lost frames for full rate speech channels (GSM 06.11 version 4.0.6), ETSI ETS 300 580-3*, March 1998.

[10] ETSI. *Digital cellular telecommunications system (Phase 2) Full rate speech; Part 6: Voice Activity Detection (VAD) for full rate speech traffic channels (GSM 06.32 version 4.3.1), ETS 300 580-6*, April 1998.

[11] ETSI. *Digital cellular telecommunications system (Phase 2); Half rate speech; Part 2: Half rate speech transcoding (GSM 06.20 version 4.3.1), ETS 300 581-2*, May 1998.

[12] ETSI. *Digital cellular telecommunications system (Phase 2+); Voice Activity Detector (VAD) for Adaptive Multi-Rate (AMR) speech traffic channels; General description (GSM 06.94 version 7.1.1 Release 1998), EN 301 708*, December 1999.

[13] ETSI. *Digital cellular telecommunications system (Phase 2+); Adaptive Multi-Rate (AMR) speech transcoding (GSM 06.90 version 7.2.1 Release 1998), ETSI EN 301 704*, April 2000.

[14] ETSI. *Digital cellular telecommunications system (Phase 2+); Enhanced Full Rate (EFR) speech transcoding (GSM 06.60 version 8.0.1 Release 1999), ETSI EN 300 726*, November 2000.

[15] ETSI. *Digital cellular telecommunications system (Phase 2); Full rate speech; Part 2: Transcoding (GSM 06.10 version 4.2.1), ETSI ETS 300 580-2*, December 2000.

[16] ETSI. *Digital cellular telecommunications system (Phase 2+) (GSM); Adaptive Multi Rate (AMR) speech; ANSI-C code for the AMR speech codec (GSM 06.73 version 7.4.1 Release 1998), ETSI 301 712*, September 2000.

[17] ETSI. *Digital cellular telecommunications system (Phase 2+) (GSM); ANSI-C code for the GSM Enhanced Full Rate (EFR) speech codec (GSM 06.53 version 8.0.1 Release 1999), ETSI EN 300 724*, November 2000.

[18] ETSI. *Digital cellular telecommunications system (Phase 2+) (GSM); Half rate speech; ANSI-C code for the GSM half rate speech codec (GSM 06.06 version 8.0.1 Release 1999), EN 300 967*, November 2000.

[19] ETSI. *Digital cellular telecommunications system (Phase 2+); Substituion and muting of lost frames for Enhanced Full Rate (EFR) speech traffic channels (GSM 06.61 version 8.0.1 Release 1999), ETSI EN 300 727*, November 2000.

[20] ETSI. *Digital cellular telecommunications system (Phase 2+); Substitution and muting of lost frames for Adaptive Multi Rate (AMR) speech traffic channels (GSM 06.91 version 7.1.1 Release 1998), ETSI EN 301 705*, April 2000.

[21] ETSI. *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); AMR speech codec, wideband; Error concealment of lost frames (3GPP TS 26.191 version 6.0.0 Release 6), ETSI TS 126 191*, December 2004.

[22] ETSI. *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions (3GPP TS 26.190 version 6.1.1 Release 6), ETSI TS 126 190*, July 2005.

[23] ETSI. *Terrestrial Trunked Radio (TETRA); Speech codec for full-rate traffic channel; Part 2: TETRA codec, EN 300 395-2 v1.3.1*, January 2005.

[24] ETSI. *Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); ANSI-C code for the Adaptive Multi-Rate - Wideband (AMR-W) speech codec (3GPP TS 26.173 version 6.0.0 Release 6), ETSI TS 126 173*, January 2006.

[25] J. D. Gibson. Speech coding methods, standards, and applications. *Circuits and Systems Magazine, IEEE*, 5(4):30–49, Fourth Quarter 2005.

[26] F. Gustafsson, L. Ljung, and M. Millnert. *Signalbehandling*. Studentlitteratur, 2000. ISBN 91-44-01709.

[27] K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, and J.-P. Adoul. Gsm enhanced full rate speech codec. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 771–774, 1997.

[28] N.N. Katugampala, K.T. Al-Naimi, S. Villette, and A.M. Kondoz. Real time data transmission over gsm voice channel for secure voice and data applications. In *The 2nd IEE Secure Mobile Communications Forum: Exploring the Technical Challenges in Secure GSM and WLAN*, London, September 2004.

[29] N.N. Katugampala, K.T. Al-Naimi, S. Villette, and A.M. Kondoz. Real time end to end secure voice communications over gsm voice channel. In *13th European Signal Processing Conference*, Antalya, Turkey, September 2005.

[30] N.N. Katugampala, S. Villette, and A.M. Kondoz. Secure voice over gsm and other low bit rate systems. In *IEE Secure GSM and beyond: end to end security for mobile communications*, London, February 2003.

[31] A.M. Kondoz. *Digital Speech: Coding for Low Bit Rate Communications Systems*. John Wiley & Sons Ltd, 1994. ISBN 0-471-95064-5.

[32] A.M. Kondoz. *Digital Speech: Coding for Low Bit Rate Communications Systems*. John Wiley & Sons Ltd, second edition, 1996. ISBN 0-471-87008-7.

[33] A.M. Kondoz, N.N. Katugampala, and K.T. Al-Naimi. Data transmission. European Patent Office, patent number WO2005109923, 2005.

[34] M. Mouly and M.-B. Pautet. *The GSM System for Mobile Communications*. published by the authors, France, 1992. ISBN 2-9507190-0-8.

[35] R. Salami, C. Loflamme, B. Bessette, J.-P. Adoul, K. Jarvinen, J. Vainio, P. Kapaenen, T. Honkanen, and P. Haavisto. Description of gsm enhanced full rate speech codec. In *IEEE International Conference on Communications, 1997. ICC 97 Montreal, 'Towards the Knowledge Millennium'. 1997*, volume 2, pages 725–729, June 1997.