SPEAKER-DEPENDENT SPEECH CODING

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND
ELECTRONICS ENGINEERING
AND THE INSTITUTE OF ENGINEERING AND SCIENCES
OF BILKENT UNIVERSITY
IN PARTIAL FULLFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Mete Kemal Kart
August 2006

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. A.Enis Çetin (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Özgür Ulusoy

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Uğur Güdükbay

Approved for the Institute of Engineering and Sciences:

Prof. Dr. Mehmet B. Baray
Director of Institute of Engineering and Sciences

ABSTRACT

SPEAKER DEPENDENT SPEECH CODING

Mete Kemal Kart
M.S. in Electrical and Electronics Engineering
Supervisor: Prof. Dr. Enis Çetin
August 2006

With the rapid expansion of Internet, it became feasible to have low-cost and secure telephone calls via internet. New digital speech compression standards were developed. Digital speech codecs can be used both in regular telephone networks and Internet based systems. Thereby, for a secure call, speech data are firstly compressed by digital speech codecs, and then, these compressed packages are sent in an encoded way through Data Encryption Standard (3DES) [1], Advanced Encryption Standard (AES) [2] encoding methods. Compressing and encoding processes require high processor performance and they may even require the use of high frequency processors and DSP's to encode the binary speech data. Current speech coders are speaker independent, i.e., they don't perform any speaker specific operations. They do not even distinguish between male and female speakers. An interesting way to solve this problem is to send speech after encoding it with a system that is based on a specific user. This system, which can be also called as speaker dependent speech encoding, provides a computationally efficient and relatively secure VoIP call, with high quality and without any encoding compared to the same bit rate standard speech codecs. Despite the disadvantages of requirement of acquiring all the speech characteristics of users and the need for extra data space, it has advantages such as providing secure communication because speech characteristics of the speaker is unknown to other users and the synthesized speech has higher quality compared to a same bit rate LPC compressed speech.

Keywords: SDSC, LPC, CELP, VoIP, codebook, speaker-dependent speech coding, speech encoding, voice compression.

ÖZET

KİŞİYE BAĞIMLI SES KODLAMASI

Mete Kemal Kart
Elektrik ve Elektronik Mühendisliği Bölümü Yüksek Lisans
Tez Yöneticisi: Prof. Dr. Enis Çetin
Ağustos 2006

İnternetin yaygın bir hal almasıyla birlikte, insanlar internet üzerinden ucuz ve güvenilir telefon görüşmesi yapmanın yollarını aramaya başladılar. Sesin sayısal ortamda iletimi ile sesin sıkıştırılarak gönderilmesi sağlandı. Çalışmalar sonucunda yeni ses sıkıştırma standartları geliştirildi. Bununla beraber güvenli bir görüşme için ses verileri önce bu geliştirilen kodekler tarafından sıkıştırıldı ve ardından bu sıkıştırılan paketler 3DES, AES gibi şifreleme yöntemleriyle şifrelenerek gönderilmeye başlandı. Sıkıştırma ve şifreleme işlemlerinin yüksek işlem gücü istemesi, frekansı yüksek işlemci ve DSP'lere rağmen, sınırlı sayıda şifrelenmiş RTP bağlantısı (görüşme) yapma olanağı tanımaktadır. Bu sınırı artırmanın yollarından biri konuşulan kişinin ses bilgilerinin bilinmesine dayanan bir sistemle sesi sıkıştırıp yollamaktır. Kişiye bağımlı ses kodlaması olarak tanımlayabileceğimiz bu sistem aynı bit oranındaki kodeklere nazaran hem daha kaliteli hem de şifrelemeye gerek duymadan nisbeten daha güvenli bir VoIP görüşmesi sağlamaktadır. Görüşülen her kişinin ses karakteristiğini bilme mecburiyeti ve ekstra veri alanı işgal etmesi gibi dezavantajları yanında, kişiye ait ses karakteristiğinin başkaları tarafından bilinmemesi sayesinde güvenli iletişim sağlaması ve aynı bit oranında sıkıştırma yapan LPC'den daha kaliteli ses sentezlenebilmesi nedeniyle incelenmesi gereken bir metottur. Askeri ve günlük alanlarda kullanılması mümkün olan kişiye bağımlı ses kodlamasının nasıl çalıştığı bu makale içinde detaylı bir şekilde anlatılacaktır.

Anahtar Kelimeler: SDSC, LPC, CELP, VoIP, kod defteri, kişiye bağımlı ses kodlaması, ses şifreleme, ses şıkıştırma.

Acknowledgements

I would like to express my deep gratitude to my supervisor Prof. Dr. Ahmet Enis Çetin for his instructive comments and constant support throughout this study.

I would like to express my special thanks to Prof. Dr. Özgür Ulusoy and Assoc. Prof. Dr. Uğur Güdükbay for showing keen interest to the subject matter and accepting to read and review the thesis.

I would also like to thank Uğur Kart, Mustafa Durukal, Burak Güldoğan, İbrahim Onaran and Mehmet Türkan for many helpful suggestions and discussions.

It is a great pleasure to express my special thanks to my family, Metin, Nigar, Mustafa and Uğur who brought me to this stage with their endless patient and dedication.

Table of Contents

1. INTRODUCTION 1

1.1 LINEAR PREDICTIVE CODING 2
1.1.1 Encoder 4
1.1.2 Decoder 5
1.2 CODE EXCITED LINEAR PREDICTION 6
1.2.1 Encoder 6
1.2.2 Decoder 7
1.3 ORGANIZATION OF THE THESIS 8

2. SPEAKER-DEPENDENT SPEECH CODING 9

2.1 SPEECH DATA PROCESSING PROCEDURE 9
2.1.1 Encoder 10
2.1.2 Decoder 12
2.2 GENERATING SDSC CODEBOOK TABLE 13
2.2.1 Raw Codebook Table 13
2.2.2 Huffman Coded Codebook Table 14
2.3 CODEBOOK TABLE SEARCH 14
2.4 FINDING THE PHASE INDEX 15
2.5 PITCH PERIOD ESTIMATION 16
2.6 SDSC WITH HIGHER BIT-RATES 16

3. EXPERIMENTAL STUDIES 19

3.1 COMPARISON OF THE SIGNALS 19
3.2 CODEBOOK TABLE 24
3.3 BIT ALLOCATION 26
3.4 SDSC VERSUS LPC-10 AND MELP 27

4. APPLICATION AREAS OF THE SDSC ENCODER 29

4.1 PRACTICAL USE OF SDSC ENCODER 29
4.1.1 Basic Initiation Scenario 29
4.1.2 Basic Initiation Scenario by Sending Encrypted Codebook Tables 30
4.1.3 SDSC with SIP Integration 33
4.2 SECURITY ANALYSIS OF THE SDSC ENCODER 35

5. CONCLUSIONS 37

APPENDIX A 39

SPEECH SIGNAL FIGURES OF MALE AND FEMALE TEST SOUNDS IN ENGLISH 39

BIBLIOGRAPHY 48

List of Figures

Figure 1.1 LPC model of speech production [14]. 3
Figure 1.2 Block Diagram of LPC encoder [14]. 4
Figure 1.3 Block Diagram of LPC decoder [14]. 5
Figure 1.4 Block Diagram of CELP encoder [14]. 6
Figure 1.5 Block Diagram of CELP decoder [14]. 7
Figure 2.1 Block Diagram of SDSC encoder. 10
Figure 2.2 Block Diagram of SDSC with phase encoder. 11
Figure 2.3 Block Diagram of SDSC decoder. 12
Figure 2.4 Block Diagram of SDSC with phase decoder. 13
Figure 2.5 Codebook table search process. 14
Figure 2.6 Finding phase difference between original excitation and excitation chosen from the codebook table. 15
Figure 2.7 Quantizer approach used in SDSC with high bit-rates. 18
Figure 3.1 Excitation of a voiced speech frame from original signal (above), the best match excitation chosen from codebook table (below). 20
Figure 3.2 A voiced speech frame from original signal (above), the synthesized speech frame using best match excitation chosen from codebook table (below). 21
Figure 3.3 Excitation of an unvoiced speech frame from original signal (above), the best match excitation chosen from codebook table (below). 22
Figure 3.4 An unvoiced speech frame from original signal (above), the synthesized speech frame using best match excitation chosen from codebook table (below). 23
Figure 3.5 Original speech signal (above), the synthesized speech signal using SDSC (below). 24
Figure 4.1 Basic Initiation Scenario of SDSC. 30
Figure 4.2 Basic Initiation Scenario of SDSC with encrypted codebook tables. 31
Figure 4.3 A small but different codebook generation from the big codebook table. 32
Figure 4.4 Codebook generation using permutation from the original codebook table. 32
Figure 4.5 SIP Initiation Scenario. 33
Figure 4.6 SIP INFO packet creation. 34
Figure 4.7 SDSC integration into the SIP Scenario. 35

Figure A.1 Female-1 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 40
Figure A.2 Female-2 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 41
Figure A.3 Female-3 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 42
Figure A.4 Female-4 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 43
Figure A.5 Male-1 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 44
Figure A.6 Male-2 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 45
Figure A.7 Male-3 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 46
Figure A.8 Male-4 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 47

List of Tables

Table 3.1 Huffman coding table of pitch periods. 25
Table 3.2 LPC bit allocation [12]. 26
Table 3.3 SDSC bit allocation. 27
Table 3.4 SDSC with phase bit allocation. 27
Table 3.5 MOS rates of different coders. 28

Chapter 1

Introduction

Speech Coding is one of the most important research areas in signal processing. After the wide use of the Internet, people are looking for ways of reducing the cost of the telephone calls. Voice over IP (VoIP) is one of the methods used to achieve this goal. Because of the limited bandwidth of the networks, 16-bit PCM speech data should be compressed, therefore many speech coding standards and coders are developed (G711 [3], G722 [4], G722.1 [5], G722.2 [6], G723.1 [7], G726 [8], G727 [9], G728 [10], and G729 [11] named vocoders). These coders are speaker independent and based on Linear Predictive Coding (LPC).

In LP-based coders, speech parameters, by which speech can be reproduced, are packed and sent to the receiving end. The receiving end uses these parameters and synthesizes the sound. This procedure is independent of the speaker. We propose a new method that depends on the sound characteristics of the speaker and it has the same bit-rate as Linear Predictive Coding - 10 with a higher speech quality than LPC.

In this thesis, we investigate speaker dependent speech coding (SDSC). In this chapter, first, we introduce Linear Predictive Coding (LPC) [12] and Code

Excited Linear Prediction (CELP) [13] which are the pioneers of the speaker dependent speech coding and then the SDSC algorithm is explained.

In Section 1.1, LPC and, in Section 1.2, CELP are explained. Section 1.3 outlines the work that has been carried out in this research.

1.1 Linear Predictive Coding

Based on a highly simplified model for speech production, the linear prediction coding (LPC) algorithm is one of the earliest standardized coders that works at low bit-rate 2.4 kbps and is a breakthrough in speech coding development, because it is the first computationally efficient digital speech vocoder. Even though the quality of the decoded speech is low and it is quite intelligible [12], [14].

LPC is originally developed for military applications where secure communication is more important than the quality of the sound. It can be said that most of the modern LP-based speech coders which have similar bit-rates with LPC and high sound quality, derived from LPC.

Linear prediction coding based on a more basic model of speech production is shown in Figure 1.1. The model comes from studies on basic properties of speech signals and claims to ape the human speech production mechanism. The combined spectral contributions of the glottal flow, the vocal tract, and radiation of the lips are synthesized by a filter. The driving input of the filter or excitation signal represents either an impulse train (voiced speech) or random noise (unvoiced speech). To select appropriate input, the switch is placed to the certain location. Gain parameter controls thereby energy level of the output. [14]

Linear prediction coding model supposes that the speech signal is divided into the non-overlapping and reasonable small frames (speech samples), so that the

properties of the signal essentially remain constant. For each of the frames of the speech the following parameters are calculated:

- V/UV : The frame is voiced or unvoiced.
- A : Filter coefficients of the synthesis filter.
- T_p: Pitch period of the voiced frames.
- G : Gain, the energy level of the speech sample.

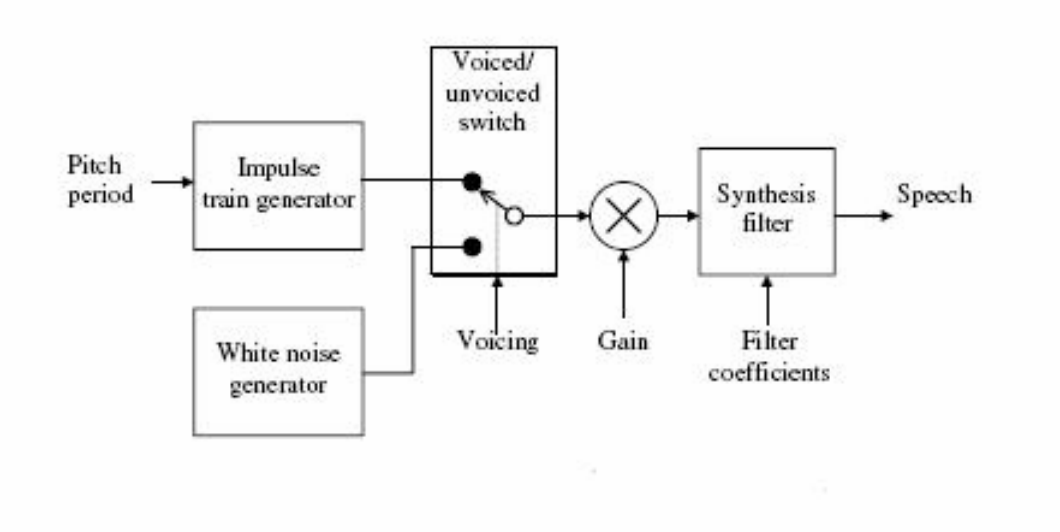


Figure 1.1 LPC model of speech production [14].

For each frame of the speech signal, the parameters V/UV, A, Tp and G are calculated and sent instead of the PCM samples, so that the frame parameters of the model allocates only 2.4 kbps which is nearly 53 times smaller than the corresponding 16-bit PCM. The quality of the model sound is low and irreversible, but in most applications, compression is more important than quality, depending on the sound, which is understandable enough. The structure of the LPC is divided into the two parts, the encoder and the decoder as shown in Figure 1.2 and 1.3 respectively.

1.1.1 Encoder

Figure 1.2 shows the block diagram of the encoder. The input speech is first segmented into the nonoverlapping frames [14]. To adjust the spectrum of the input signal, a pre-emphasis filter is put into action. The voicing detector categorizes the current frame as voiced or unvoiced and outputs one bit showing the voicing state. Ten LPCs derived LP analysis is done with the help of the pre-emphasized signal. Coefficients of LPCs are quantized with the indices that are transmitted as information frames. Thereafter, the quantized LPCs are used to form the prediction-error filter. It has the function to filter the pre-emphasized speech to get prediction error-signal as an output. Here, if and only if the frame is voiced, pitch period can be estimated by use of prediction-error signal. So, with the use of prediction-error signal as an input to pitch period estimation algorithm, the estimation will be more accurate because the formant structure due to the vocal tract is vanished. [14]

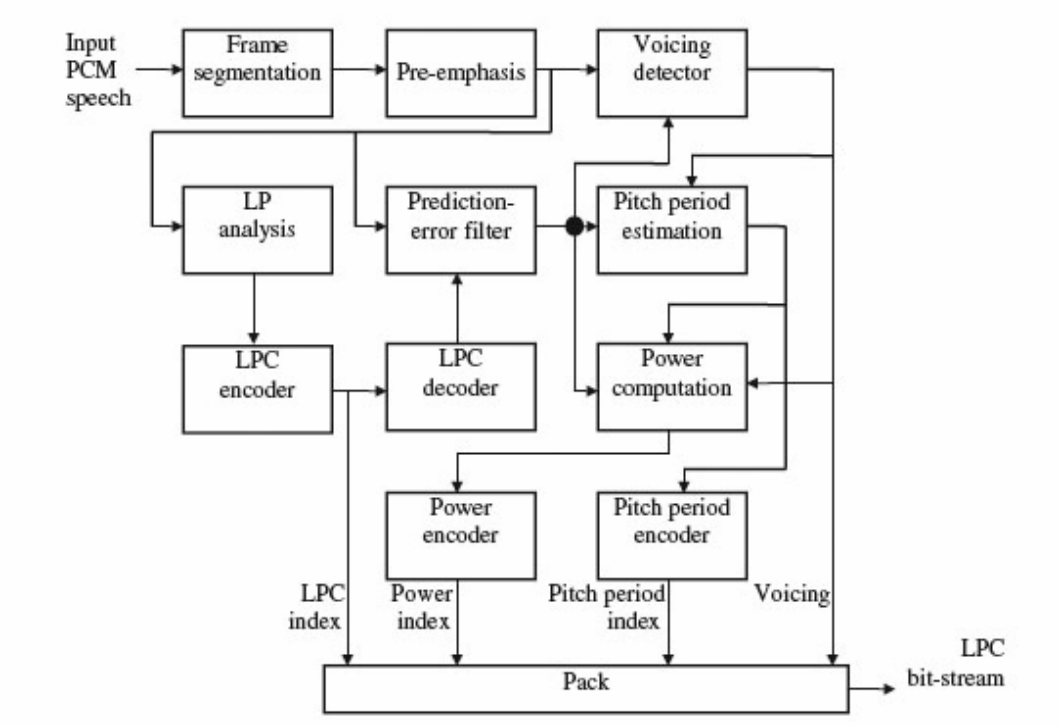


Figure 1.2 Block Diagram of LPC encoder [14].

1.1.2 Decoder

The block diagram of the decoder is shown in Figure 1.3. This is essentially the LPC model of speech production with parameters controlled by the bit-stream [14]. According to the voiced or unvoiced case, impulse train or white noise is generated. For voiced frame, pitch period index is decoded and pitch period is determined. The impulse train is generated according to this pitch period. This signal is then multiplied with the gain factor and is filtered with synthesis filter. As a last step a de-emphasis filter is used.

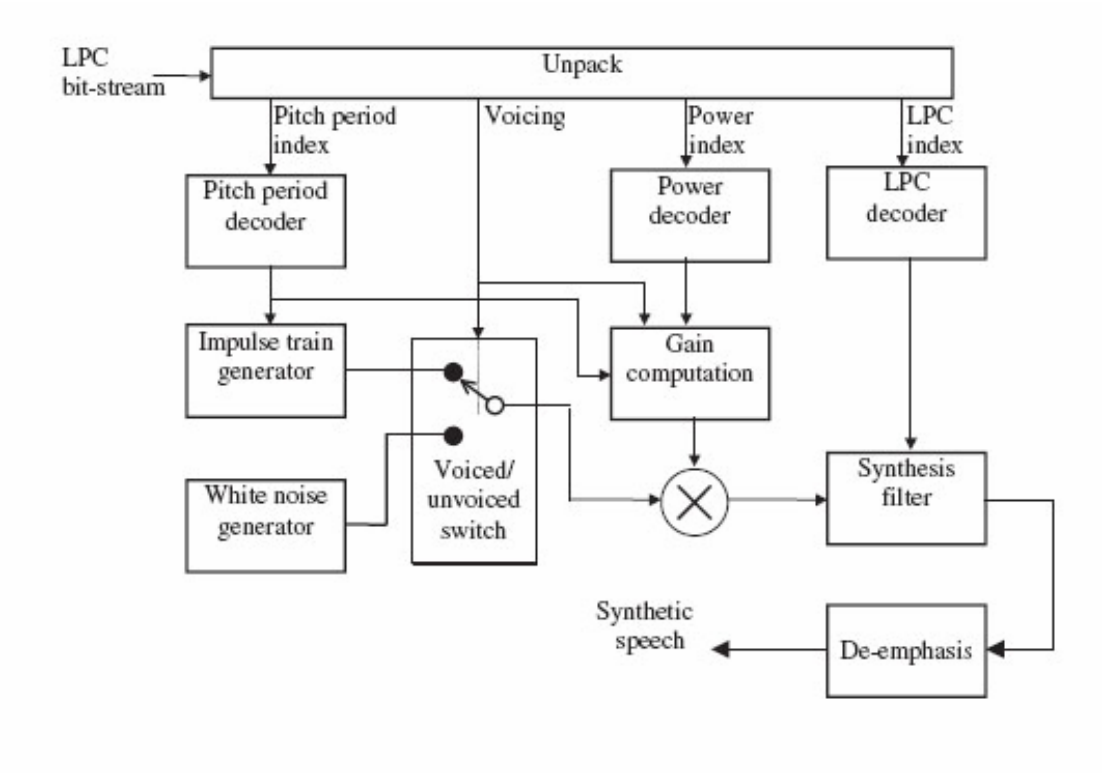


Figure 1.3 Block Diagram of LPC decoder [14].

1.2 Code Excited Linear Prediction

The model of Code Excited Linear Prediction (CELP) is the same as LPC, but CELP adds more information about the excitation. There is a codebook available on the both receiver and sender side. The sender finds the best match to the input speech from the codebook and sends the index of the best match. This procedure is known as Vector Quantization. The bit-rate of CELP is 4800-9600 bps [15]. CELP can be explored in two parts: encoder and decoder.

1.2.1 Encoder

The block diagram of CELP encoder is shown in Figure 1.4. First the input speech is divided into frames and subframes. On each frame short-term LP parameters are calculated and the input for this operation is taken the pre-emphasized speech.

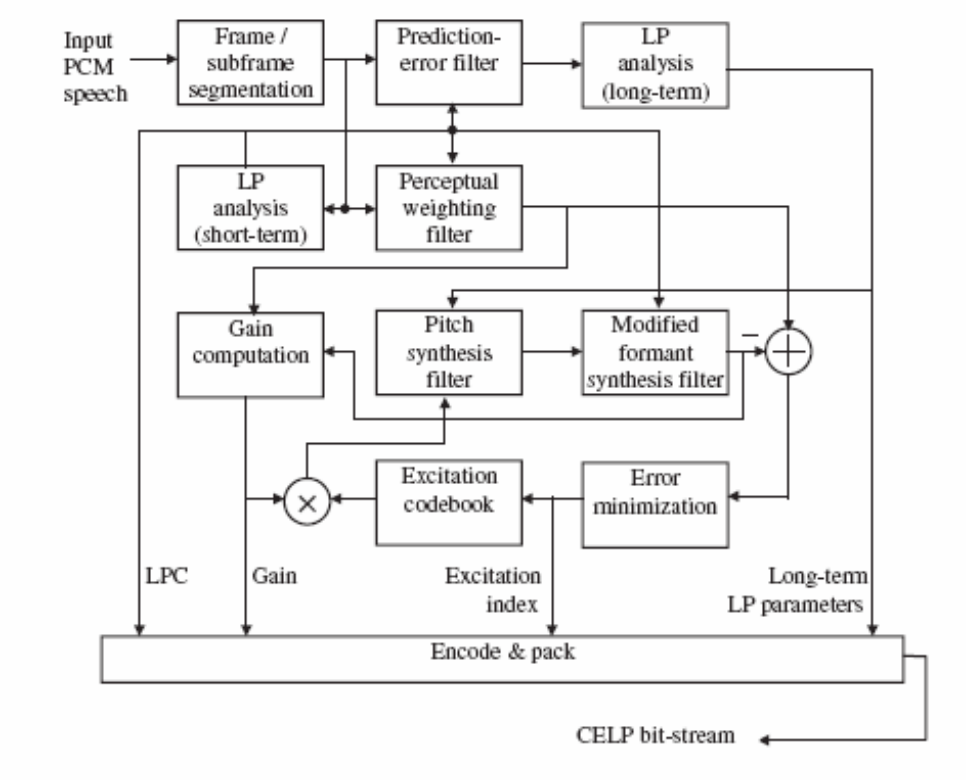


Figure 1.4 Block Diagram of CELP encoder [14].

Then long-term LP analysis is performed on each subframe by using prediction error signal. Coefficient of the perceptual weighting filter, pitch synthesis filter, and modified formant filter are known after this step [14]. Afterwards, the codebook search is applied to all subframes and the best match excitation signal is chosen respect to the lowest error. Gain is also computed in this turn. Last of all, short-term LP parameters, long-term LP parameters, gain and best match excitation index are encoded and packed [13].

1.2.2 Decoder

The block diagram of CELP decoder is shown in Figure 1.5. As shown in the figure the bit-stream is unpacked and decoded and the parameters are directed to the corresponding blocks. Then the speech is synthesized. The purpose of the post filter is to increase the quality of speech.

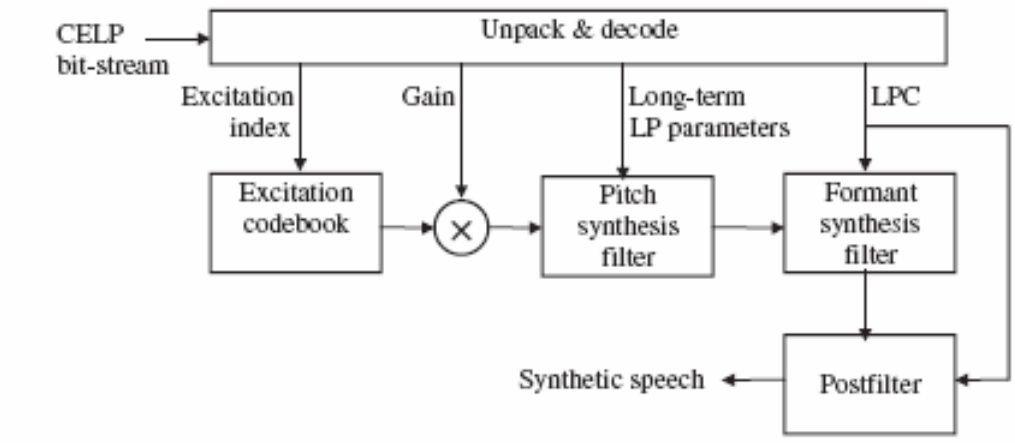


Figure 1.5 Block Diagram of CELP decoder [14].

1.3 Organization of the Thesis

Chapter 2 explains Speaker Dependent Speech Coding (SDSC) in detail. First of all, the speech data processing structure of SDSC is described by explaining SDSC encoder and the decoder. In addition, an improved version of SDSC by preserving phase information of the original signal is explained. Afterwards, element indexing of the codebook table is explored, and then search process in the codebook table is summarized. In this chapter, lastly the process of the phase difference between the original excitation and the excitation chosen from the codebook table is explained. Though this may lead to higher bit-rates.

In Chapter 3, simulation experiment results with SDSC are presented. The figures show the difference between the original signals and the synthesized signals using SDSC. Besides, bit allocation of SDSC is described and compared with LPC' bit allocation. In this chapter, Huffman code representation of the codebook table is also presented. Lastly, the speech quality of SDSC is compared with LPC-10 and MELP [16].

Chapter 4 is aimed to explore the application areas of SDSC in real life. Some scenarios are proposed and advantages and disadvantages are discussed. As a last scenario, SDSC is integrated with the Session Initiation Protocol (SIP) [17]. Lastly, a security analysis of SDSC is made.

The last chapter concludes the thesis with a summary of speaker dependent speech coding and possible future applications.

Chapter 2

Speaker-Dependent Speech Coding

In Linear predictive coding (LPC), a general method of speech compression for any person was developed. So it does not depend on the human being who generates the sound. We propose a new speech data compression method that is specific to a person.

Unlike LPC, we do not send the pitch period of the voiced part of the speech signal and whether the sound is voiced or unvoiced. Instead we have a speaker dependent normalized residue signal table (codebook) on the receiving side and we send the index of the best matching residue signal. At the receiving end, we synthesize the speech signal by using the LPC coefficients, normalization index and the residue signal, whose index was transmitted. As an improvement to SDSC, phase information is also sent to the receiver. This structure is called "SDSC with phase" (SDSCP) in this thesis. The general procedure is explained in the next part.

2.1 Speech Data Processing Procedure

The speaker-dependent speech coding procedure can be explained in two parts. They are the analysis part which is at the encoder and the synthesis part which is at the decoder.

2.1.1 Encoder

The block diagram of the SDSC encoder is shown in Figure 2.1. The PCM Input speech is first segmented into non-overlapping frames. A pre-emphasized filter is used to adjust the spectrum of the input signal. The voicing detector classifies the current frame as voiced or unvoiced. Unvoiced frames are thought as signals with zero pitch periods. The pre-emphasized signal is used for LP analysis, where ten LPC coefficients are derived. These coefficients are quantized with the indices transmitted as information of the frame. The quantized LPCs are used to build the prediction-error filter, which filters the pre-emphasized speech to obtain prediction error-signal at its output. Pitch period is estimated from the prediction-error signal if and only if the frame is voiced. [14] The prediction-error signal is normalized and the normalization (gain) factor is encoded as normalization (gain) index.

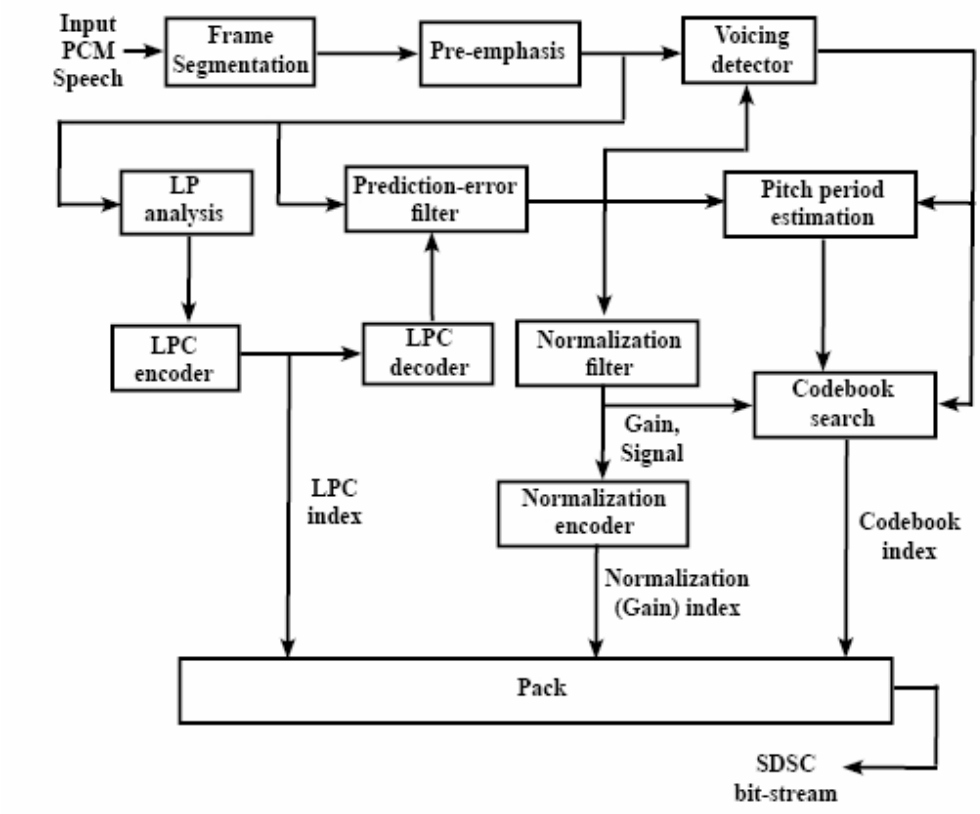


Figure 2.1 Block Diagram of SDSC encoder.

The normalized prediction-error signal and pitch period are used in the codebook search and the best match of the codebook signal is determined and the index of this signal is sent. In the search process the best match is chosen from the signals which have the same pitch period of the normalized prediction-error signal.

The block diagram of SDSCP encoder is shown in Figure 2.2. The only difference from SDSC without phase is, a phase index is generated by using gain, prediction-error and excitation chosen from the codebook table. This phase index is also packed.

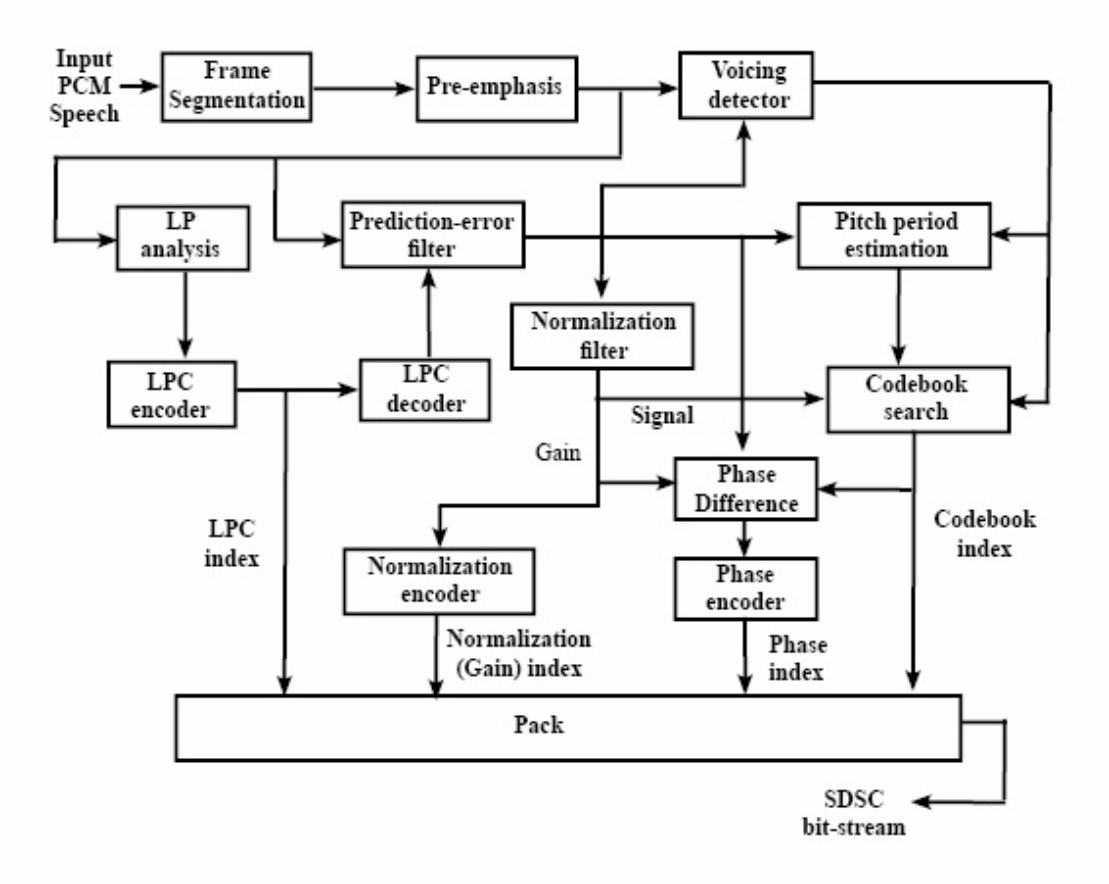


Figure 2.2 Block Diagram of SDSCP encoder.

2.1.2 Decoder

The block diagram of the SDSC decoder is shown in Figure 2.3. The codebook index is first used to determine the normalized best match error signal. The normalization index is decoded and the normalization (gain) factor is calculated. The codebook error signal is then multiplied with the gain factor and filtered with a synthesis filter. As a last step, a de-emphasis filter is used.

As a difference, SDSCP decoder first decodes the phase index and generates the phase difference. This phase difference is used to make phase adjustment to the excitation signal chosen from the codebook table. After the phase adjustment, the signal is filtered with a synthesis filter and then a de-emphasis filter is used. The block diagram of the SDSCP decoder is shown in Figure 2.4.

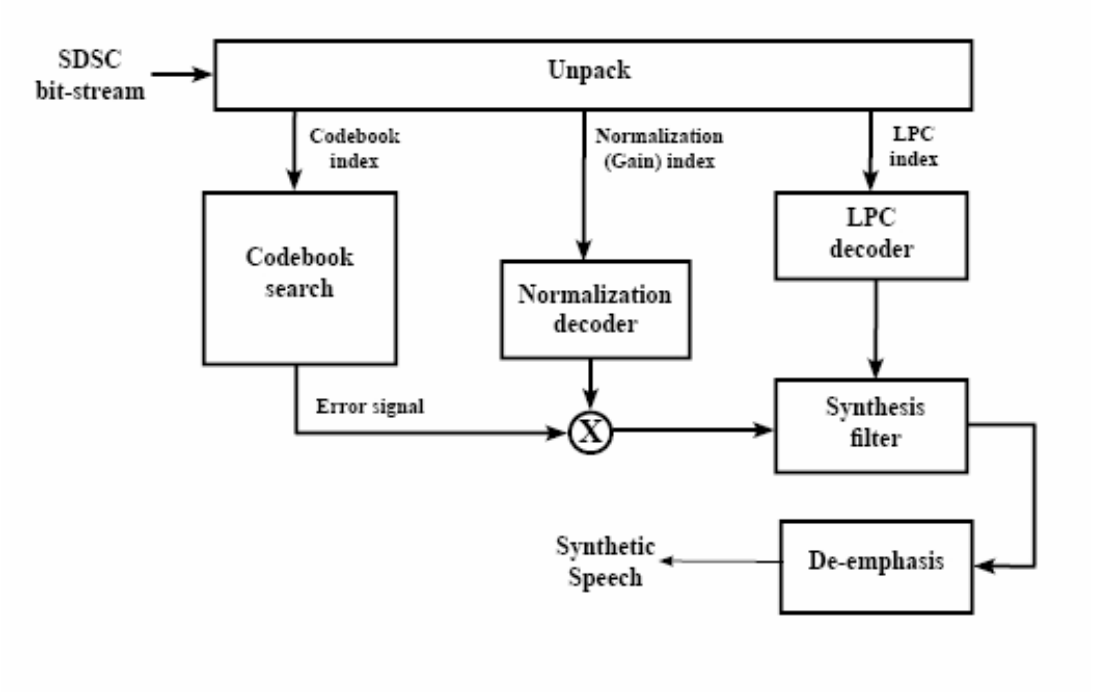


Figure 2.3 Block Diagram of SDSC decoder.

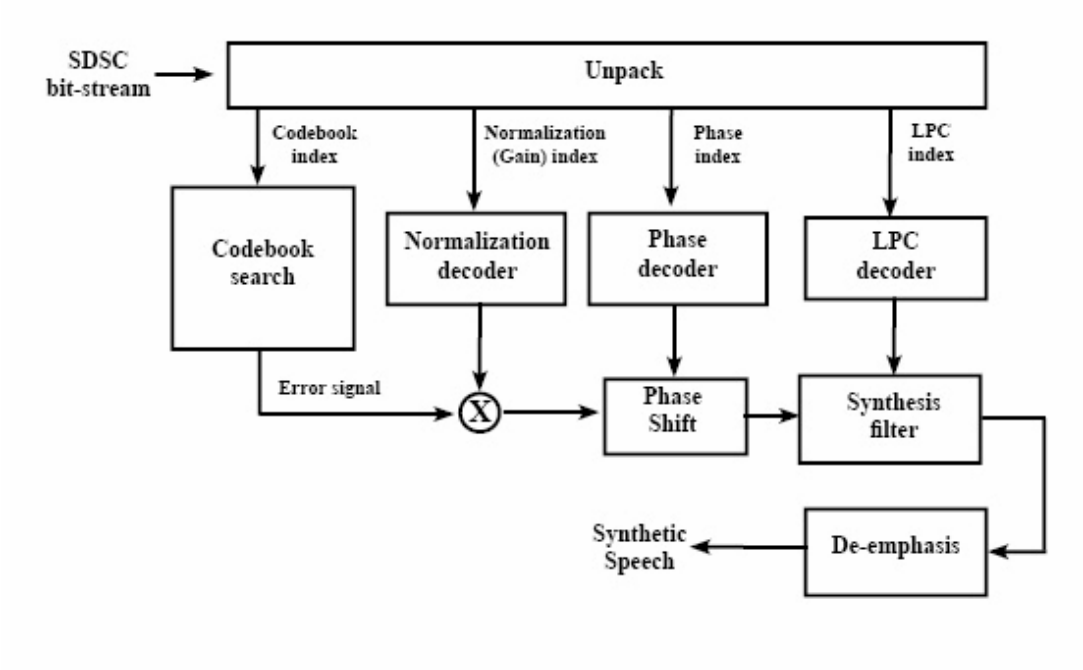


Figure 2.4 Block Diagram of SDSC with phase decoder.

2.2 Generating SDSC Codebook Table

The codebook table of SDSC should have an optimized number of elements so that the search cost and error reduction are both balanced. Two methods of table element indexing are explained below.

2.2.1 Raw Codebook Table

In this scenario, we think that the codebook table is divided according to the pitch periods of the speech frames. The number of occurrence of the pitch periods in the experiment are shown in Table 3.1. Each excitation has an index value and this index value is sent to the remote end. The encoding cost of a codebook table of size 256 is 8 bits.

2.2.2 Huffman Coded Codebook Table

We can also optimize the table by using Huffman Coding. The frequency of the occurrence of each pitch period is known, so, we can send less data bits for pitch periods with high occurrences and more data bits for pitch periods with low occurrences. The Huffman encoded codebook table bit representations are shown in Table 3.1. The average data bit rate for an index of codebook table of size 256 is 7 bits.

2.3 Codebook Table Search

Codebook table search process can be summarized in Figure 2.5. First, the excitation which comes from the original signal is normalized and the normalization factor (G) is encoded and packed. The normalized excitation is windowed. Then, the FFT of this windowed excitation is calculated and compared with the 128 values of FFT's of the excitations in the codebook table having the same pitch period with the original excitation signal.

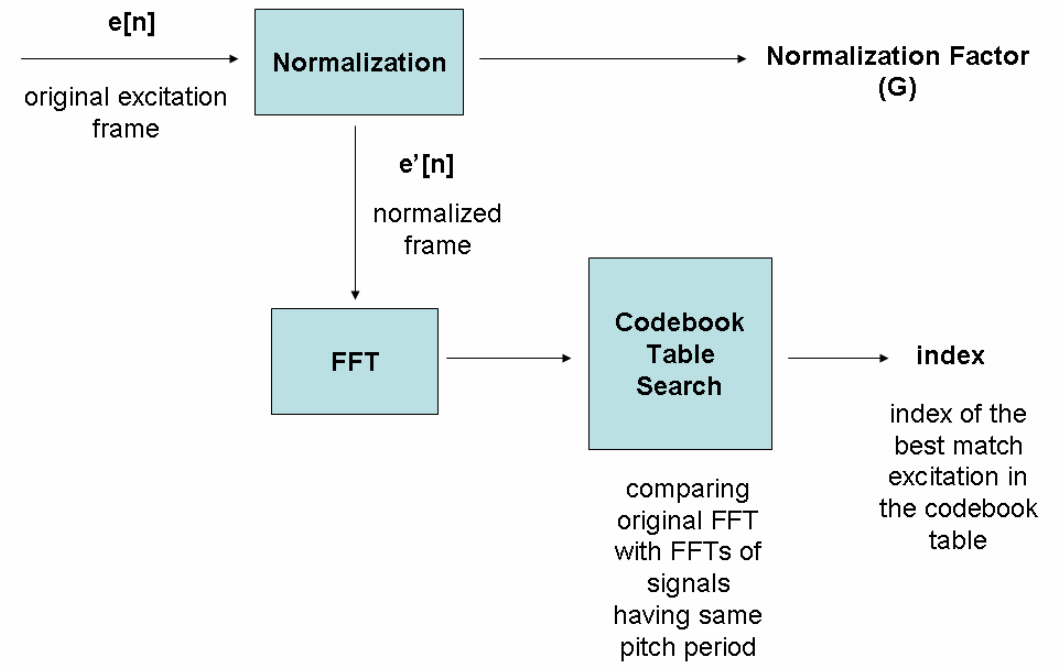


Figure 2.5 Codebook table search process.

The excitation signal with the smallest mean square error in the FFT domain is chosen and the corresponding index is packed and sent to the receiving side together with other parameters.

2.4 Finding The Phase Index

Improved version of the SDSC sends also the phase information between the original excitation and excitation chosen from the codebook table. After calculating the codebook index and normalization factor (G) shown in Figure 2.5, the signal chosen from the codebook table is multiplied with normalization factor and a denormalized signal is generated. However there may be a phase difference between the original excitation signal and the reconstructed signal. Let the phase difference be: *k*. The range of *k* should be:

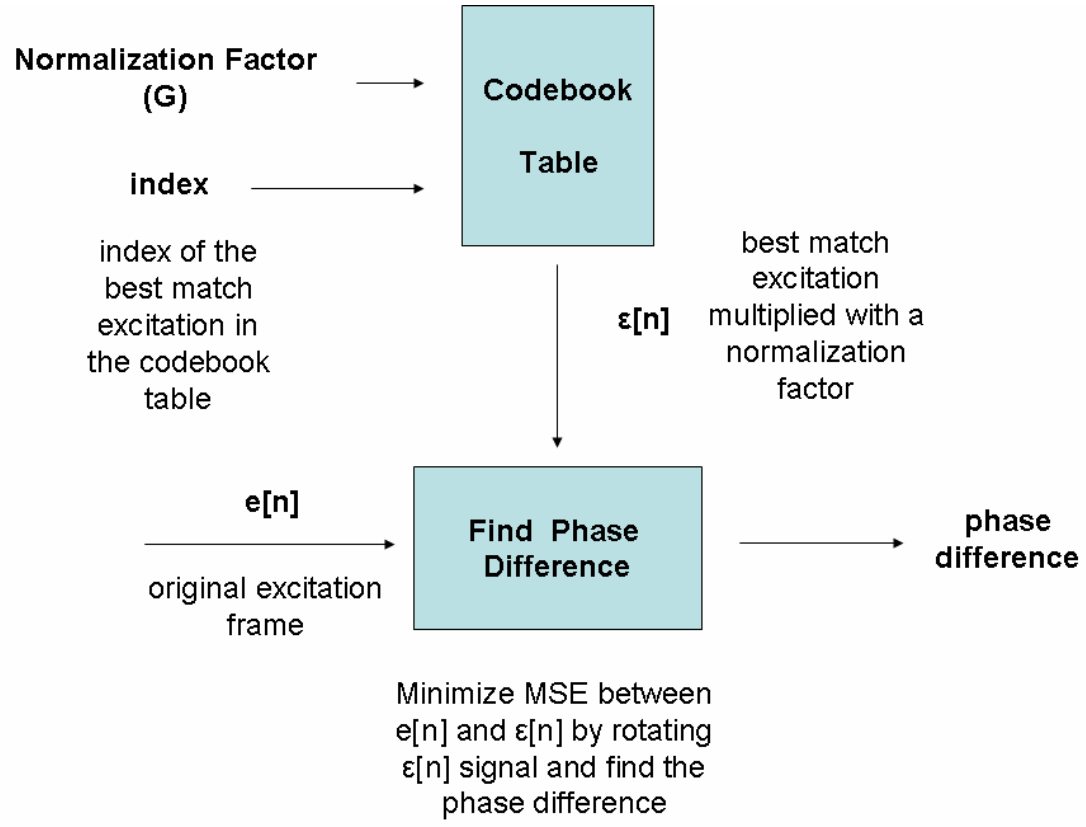


Figure 2.6 Finding the phase difference between the original excitation and the excitation chosen from the codebook table.

$$0 \leq k < N$$

where N is the number of elements in the signal. The parameter k is selected to minimize the mean square error of the original excitation *e[14]* and the reconstructed excitation signal *ε[14]* rotated by k. This procedure is shown in Figure 2.6.

2.5 Pitch Period Estimation

By calculating the autocorrelation values for the entire time-shifted version of the original signal (lag), it is possible to find the value of lag associated with the highest autocorrelation representing the pitch period estimate, since in theory, autocorrelation is maximizes when the lag is equal to the pitch period. Lag value l is a positive integer representing a time lag. The range of lag is selected so that it covers a wide range of pitch period values. For instance, for l = 20 to 147 (2.5 to 18.3 ms), the possible pitch frequency values range from 54.4 to 400 Hz at 8 kHz sampling rate. [14]

2.6 SDSC with Higher Bit-rates

By using more data or information about each SDSC speech frame we transmit, we synthesize higher quality sound. Although the excitation signal chosen from the codebook table is constructed from the original excitations, there is always a difference between a given speech frame and its corresponding codebook entry, and in some cases this difference is big enough to produce a speech frame with a high error with respect to the original speech frame. We experimentally observed that the difference between the original excitation and the excitation chosen from the codebook table on the voiced frames is high, the generated speech quality decreases. We don't encounter this problem for unvoiced frames. Unvoiced frames generally have low amplitudes and can be synthesized by

using random noise and the speech quality is not affected drastically from the excitation error.

The following procedure is implemented to increase the quality of the SDSC encoder, but it costs more bits to encode a given speech frame. To improve the speech quality, the excitation difference of a voiced original frame and the frame chosen from the codebook table, denormalized and shifted with phase index explained in Section 2.4, is first quantized and then packed. In the synthesis part this difference is unpacked and added to the phase shifted and denormalized excitation chosen from the codebook table. Figure 2.7 summarizes this scenario. In this case, the excitation differences are in the rage of ***[-0.2 0.2].*** Almost ***90%*** of these differences are between ***-0.1*** and ***0.1*** and ***5-10%*** of them are between ***-0.01*** and ***0.01***. If the excitation difference is in the interval ***[-0.01 0.01],*** the produced speech frame is similar to the original frame. In this experiment, a uniform quantizer step size of ***0.02*** taken. In the first case, the quantizer minimum and maximum values are chosen as ***-0.6*** and ***0.8***. In this case, ***3 bits*** are used to represent a value in the difference excitation signal. In the second case, the minimum and the maximum values of the quantizer are taken as ***-1.4*** and ***1.6*** respectively. So in the second case, the excitation difference signal is represented with ***4 bits/sample***. To calculate the expected bit-rates of these two versions of SDSC, the following assumptions are done:

- 2/3 of the total frames are unvociced and 1/3 of them are voiced.
- Whole data of 10% of the excitation differences of voiced frames are in the range of ***[-0.01 0.01]*** and called perfect match excitation.
- A frame contains 180 samples. With a sampling rate 8000 Hz, there are 44.44 frames with 180 samples.
- Unvoiced frame uses random noise to synthesize the signal and therefore 54 bits/frame is used. See Table 3.2.
- Voiced frames allocates 8 bits for codebook table with 256 different excitation signals, 5 bits for gain, 41 bits for LPC coefficients, 6 bits for

phase, 1 bit if the excitation is a perfect match or not, if not frame size multiply with number of bit required to represent the quantizer output.

a) 3-bits/sample for the quantizer output:

$$E[X] = \frac{2}{3} \cdot 54 + \frac{1}{3} \cdot (\frac{1}{10} \cdot 61 + \frac{9}{10} \cdot (61 + 180 \cdot 3))$$

which gives 218.3 bits/frames and multiple it with 44.44 frames/sec. The average bit-rate is ***9.7 kbps***.

b) 4-bits/sample for the quantizer output:

$$E[X] = \frac{2}{3} \cdot 54 + \frac{1}{3} \cdot (\frac{1}{10} \cdot 61 + \frac{9}{10} \cdot (61 + 180 \cdot 4))$$

which gives 272.3 bits/frames and multiple it with 44.44 frames/sec. The average bit-rate is ***12.1 kbps***.

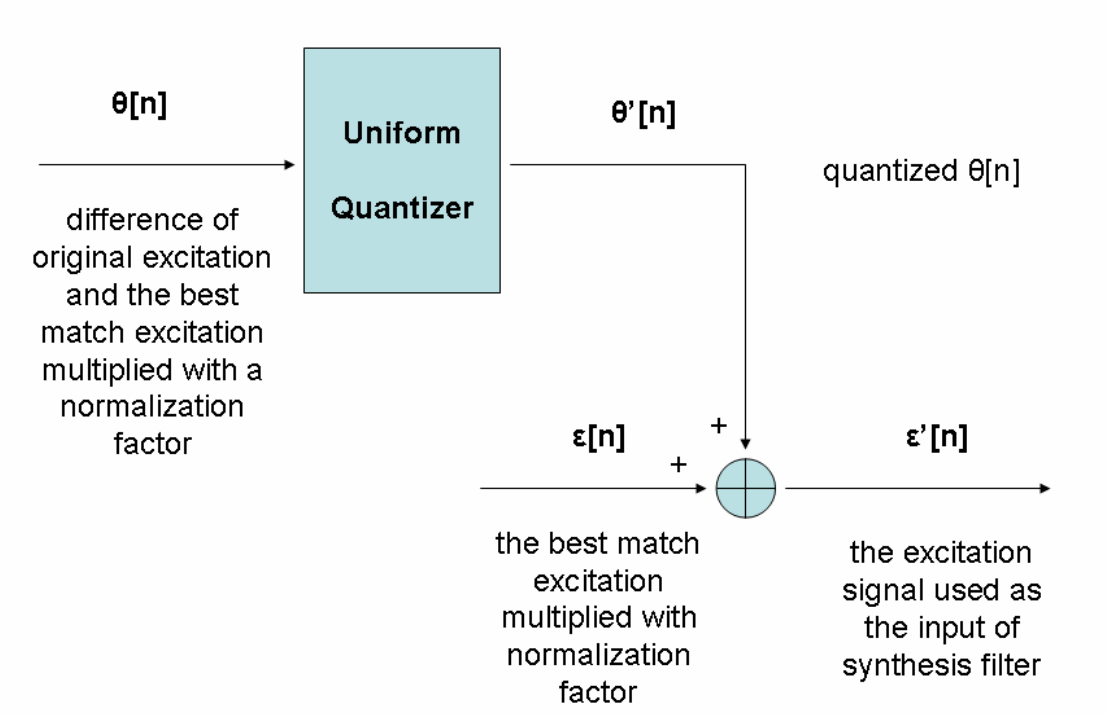


Figure 2.7 Quantizer approach used in SDSC with high bit-rates.

Chapter 3

Experimental Studies

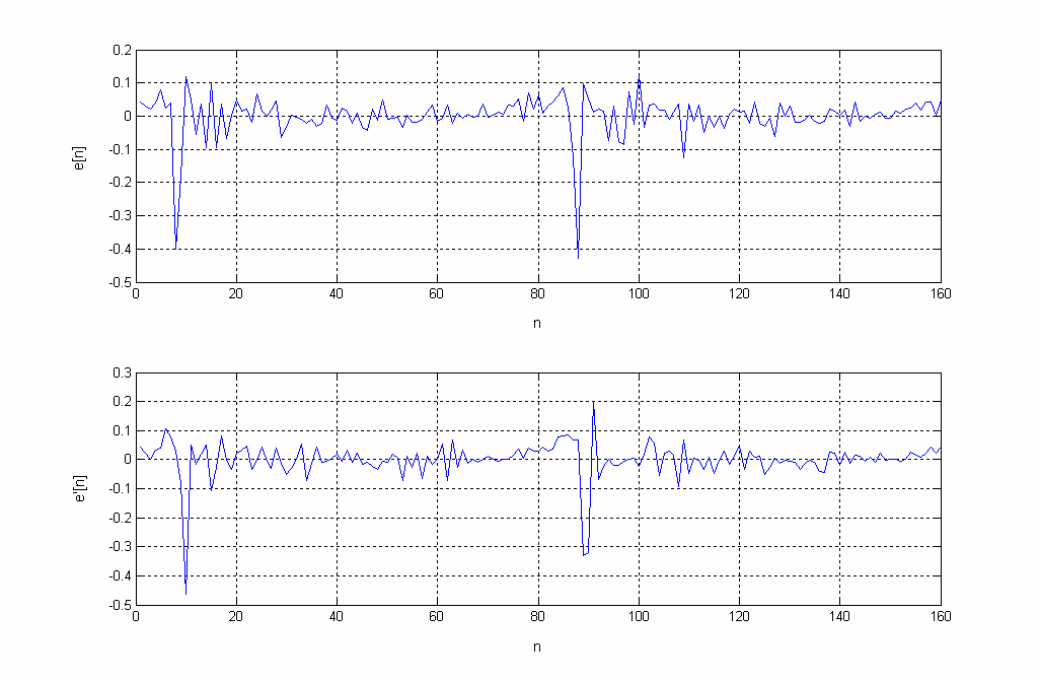
The tests are made with four different utterances of a male person in French, four different utterances of a female person in English, four different utterances of a male person in English and nine different utterances of a male person in Turkish. Each utterance is between 6 and 20 seconds long. Three of the utterances of each person are used for codebook table generation and one of the different utterances, which is not used in the codebook table generation, is used for analysis and then synthesis using the SDSC encoder.

3.1 Comparison of the signals

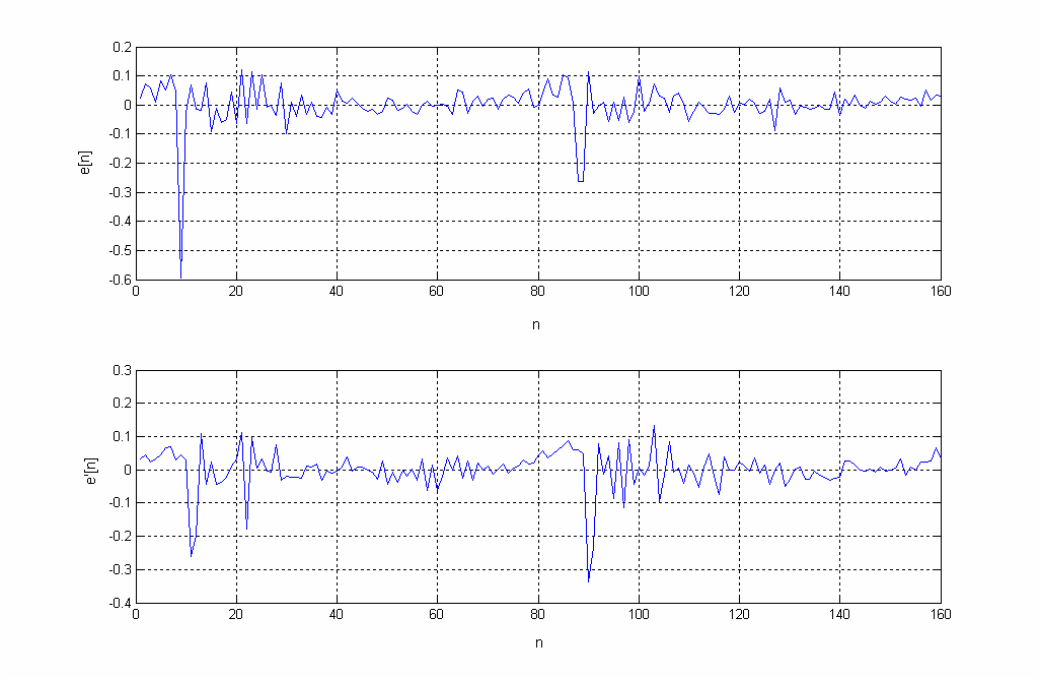
The following figures are generated by using the test utterances of the French male person. In Figure 3.1 (a) and (b) excitations of two different voiced signals are shown (top plots). The signals, below, are the best match excitations chosen from the codebook table. In Figure 3.2 (a) and (b), signals on top are the original voiced frame signals, where the excitation signals of these frame are shown in Figure 3.1 (a) and (b) in the top parts. The bottom signals in Figure 3.2 (a) and (b) are the synthesized speech frames from the best match excitations chosen from the codebook table.

In Figure 3.3 (a) and (b), top signals are excitations of two different unvoiced signals. The bottom signals are the best match excitations chosen from the codebook table. In Figure 3.4 (a) and (b), top signals are the original unvoiced frame where the excitation signals of these frame are shown in Figure 3.3 (a) and (b) in the above parts. The bottom signals in Figure 3.4 (a) and (b) are the

synthesized speech frames from the best match excitations chosen from the codebook table.

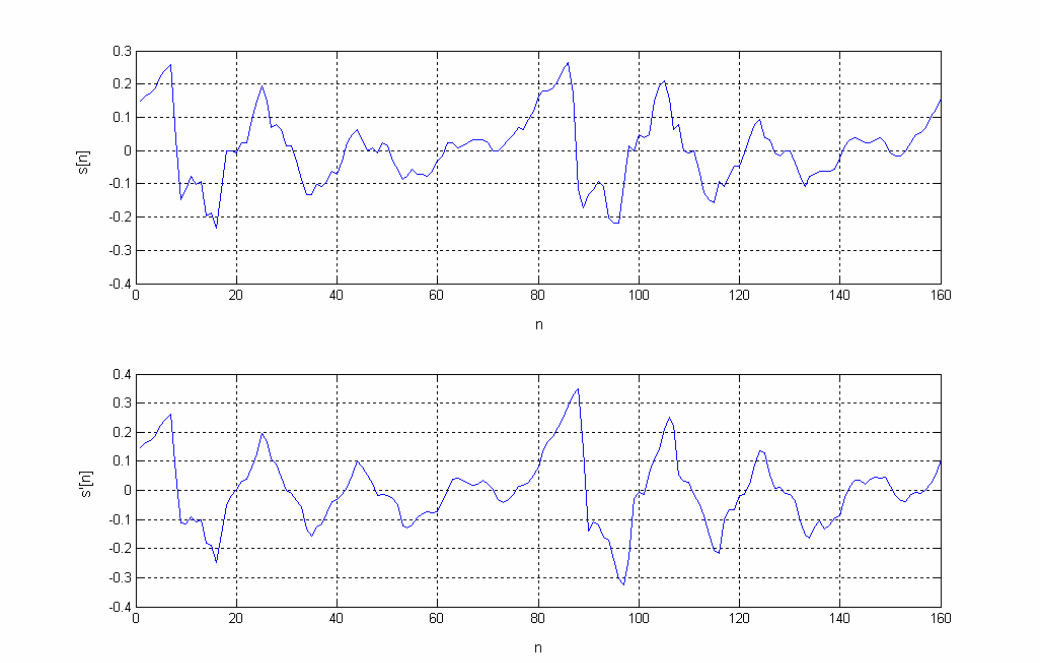


(a)

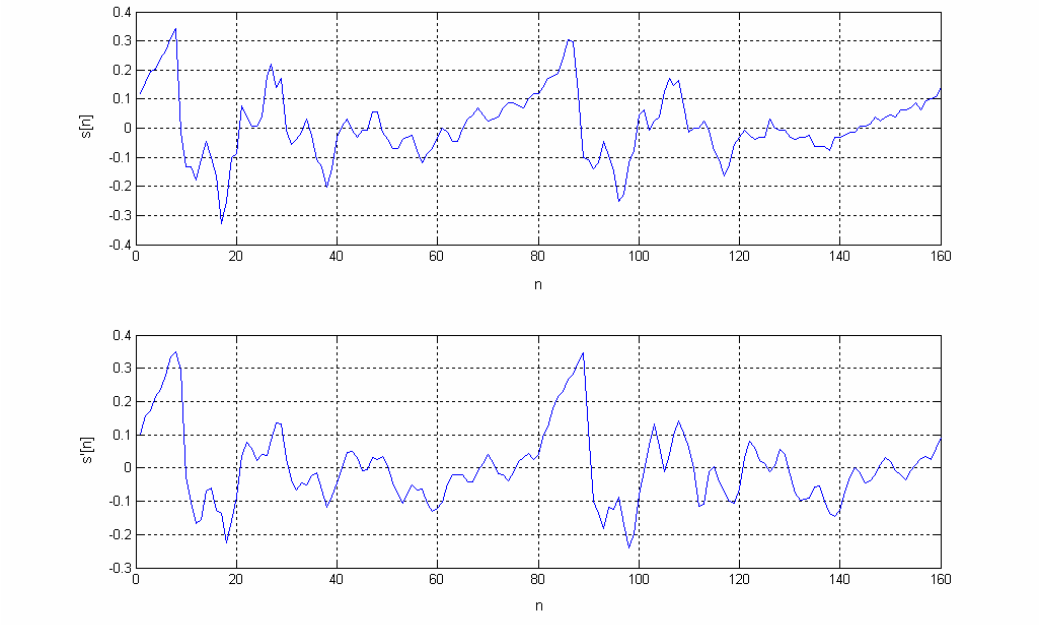


(b)

Figure 3.1 Excitation of a voiced speech frame from original signal (above), the best match excitation chosen from codebook table (below).

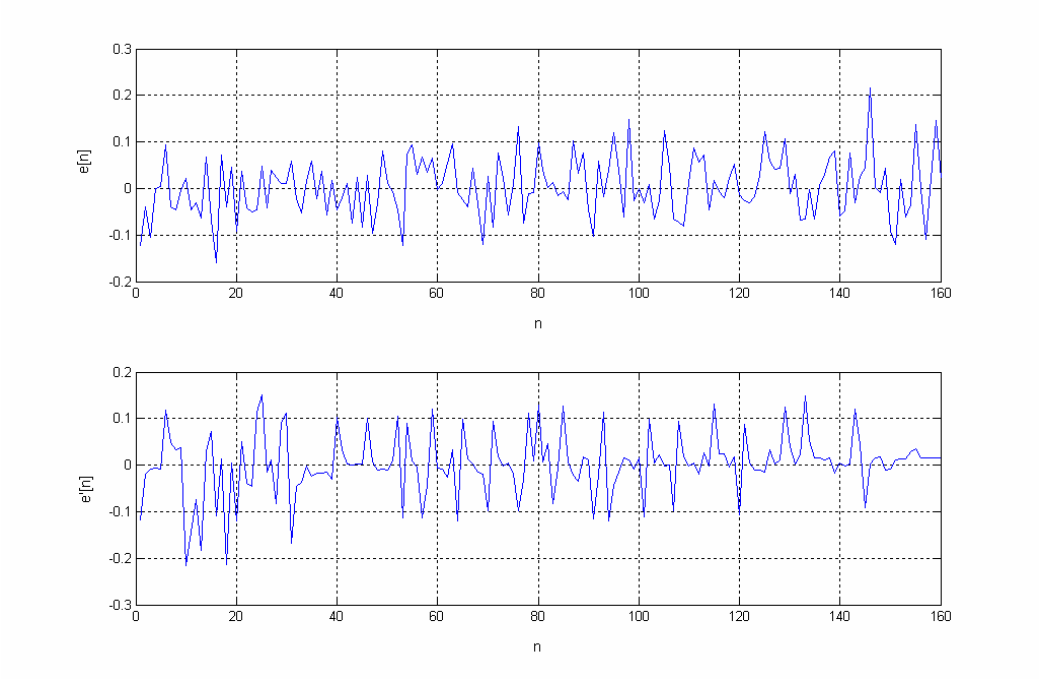


(a)

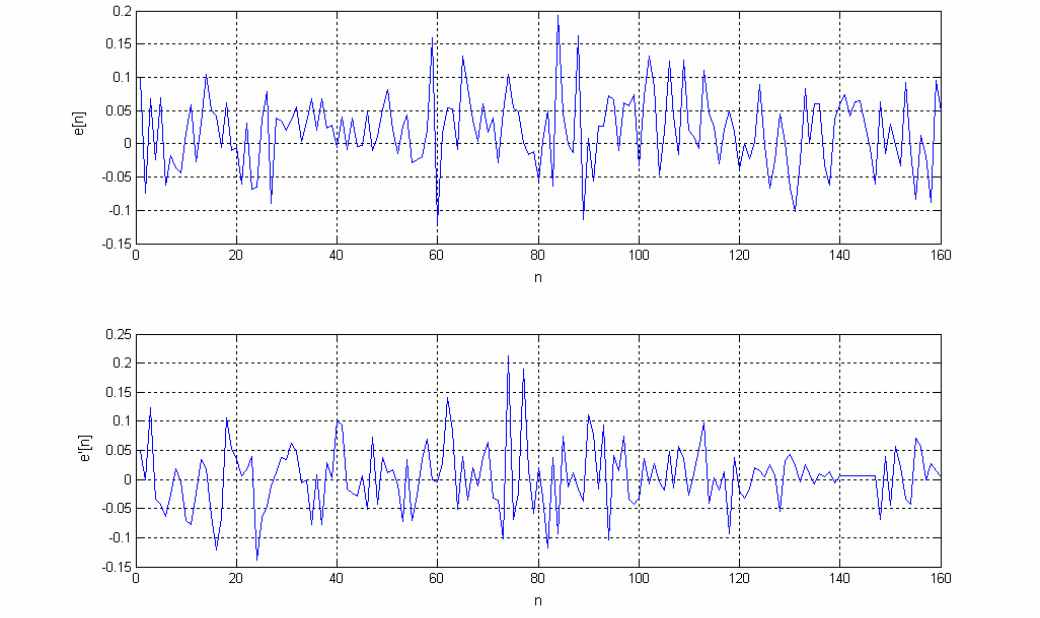


(b)

Figure 3.2 A voiced speech frame from original signal (above), the synthesized speech frame using the best match excitation chosen from codebook table (below).

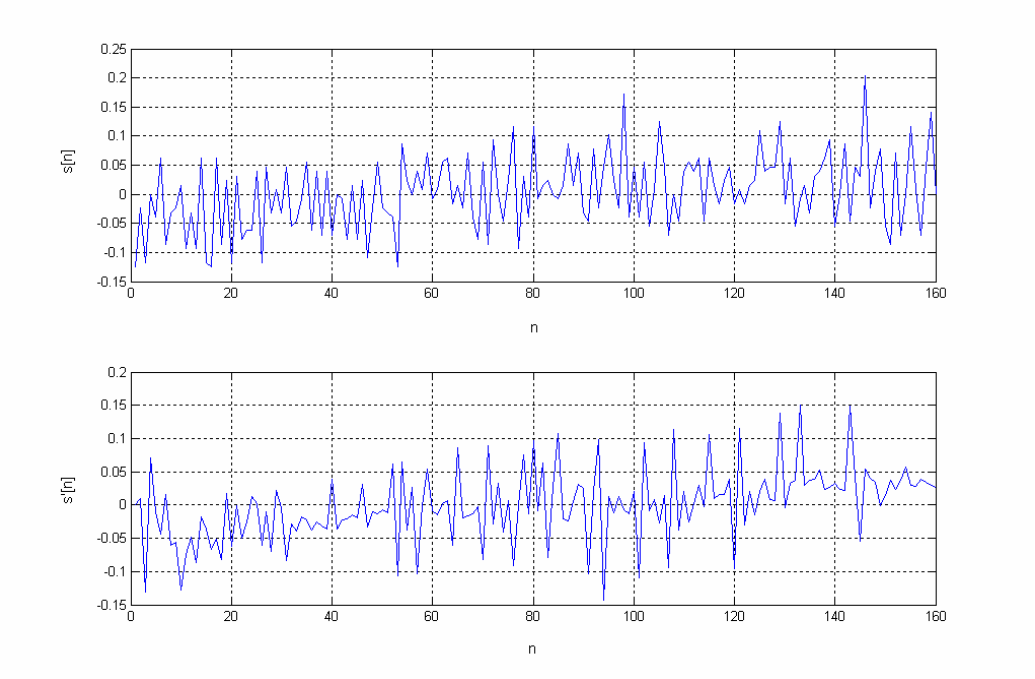


(a)

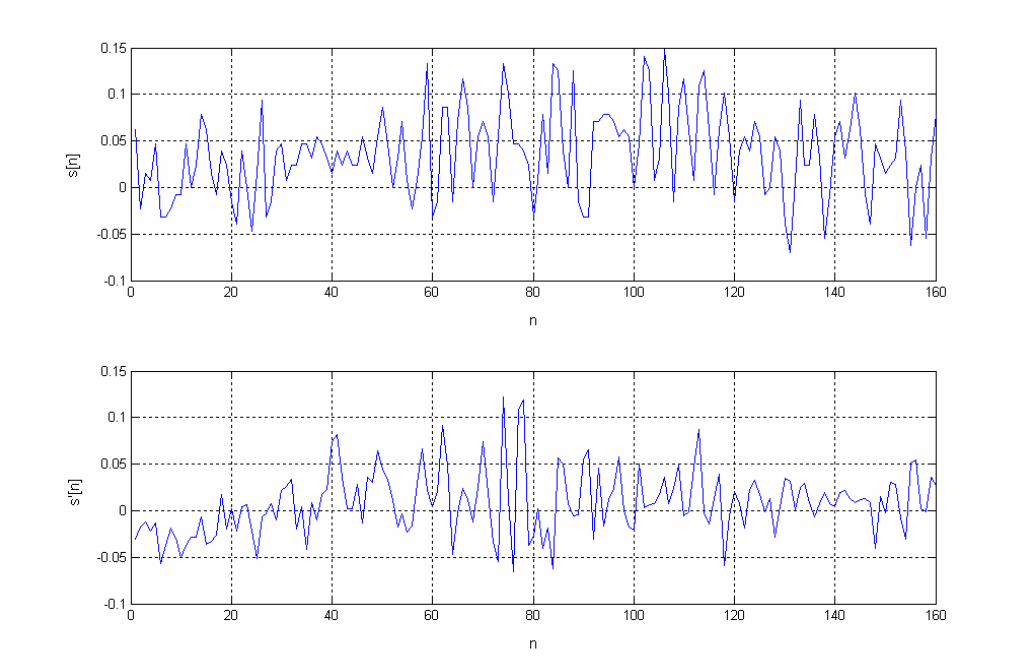


(b)

Figure 3.3 Excitation of an unvoiced speech frame from original signal (above), the best match excitation chosen from codebook table (below).



(a)



(b)

Figure 3.4 An unvoiced speech frame from original signal (above), the synthesized speech frame using best match excitation chosen from codebook table (below).

In Figure 3.5 at the above part, a whole original speech signal is shown; in the same figure at the below part, the synthesized speech signal using SDSC is illustrated.

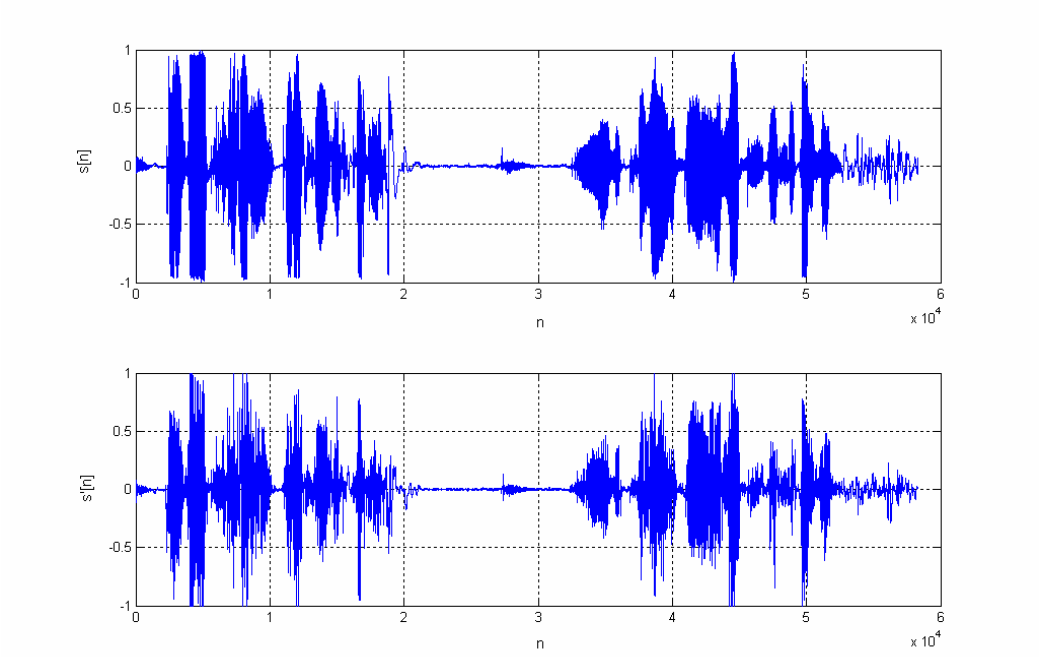


Figure 3.5 Original speech signal (above), the synthesized speech signal using SDSC (below).

3.2 Codebook Table

Three sample speech signals are used for the codebook table generation. There are a total of 748 frames and different excitation signals, therefore there are 748 codebook table entries. Unvoiced sound is supposed to have zero pitch periods. The number of occurrence of each pitch period and the corresponding Huffman code representation is shown in Table 3.1

Pitch period	Number of occurrences	Huffman bit representation
0	518	1
4	2	00100010
11	3	01101100
13	2	00100001
24	1	000101101
45	2	00100000
53	3	01101111
54	3	01101110
55	7	0111101
56	8	001010
57	10	010110
58	7	000100
59	6	0110100
60	19	01001
61	15	00011
62	26	01110
63	20	01010
64	15	00000
65	15	00001
66	11	010111
67	6	0110101
68	9	001110
69	12	011000
70	17	00110
71	14	011111
72	2	00100011
73	5	0100010
74	9	010000
75	5	0010111
76	8	001001
77	9	001111
78	12	011001
79	5	0100011
80	6	0111100
81	4	0010110
82	2	00010101
86	1	000101100
119	2	00010100
125	1	000101111
127	3	01101101
136	1	000101110

Table 3.1 Huffman coding table of pitch periods.

3.3 Bit Allocation

In LPC for each frame total number of bits required is 54 bits [18]. 1 bit is for determining whether the frame is voiced or unvoiced, 6 bits are for pitch period index where 60 pitch period values are considered, 41 bits are for LPC parcor coefficients, 5 bits are for gain index and 1 bit is for synchronization. Bit allocation for LPC is illustrated in Table 3.2. A speech with sampling rate 8000 samples/second and frame rate 180 samples/frame has a bit rate of 2400 bits/second.

Parameter	Voiced	Unvoiced
Pitch period / voicing	7	7
Power	5	5
LPC	41	20
Synchronization	1	1
Error protection	-	21
Total	54	54

Table 3.2 LPC bit allocation [12].

SDSC requires total number of 54 bits for a frame. 8 bits of them are for codebook index, 5 bits are for the gain index calculated when excitation signal is normalized. The remaining 41 bits are for LPC parcor coefficients. Bit allocation for SDSC is illustrated in table 3.3. A speech with sampling rate 8000 samples/second and frame rate 180 samples/frame has a bit rate of 2400 bits/second (2.4 Kbps). SDSC with phase also adds phase information to the packet. 6 bits extra cost makes the bit-rate 2666 bits/second (2.7 kbps).

Parameter	Bit Allocation
Codebook (256 entries)	8
Normalization (Gain)	5
LPC	41
Total	54

Table 3.3 SDSC bit allocation.

Parameter	Bit Allocation
Codebook (256 entries)	8
Normalization (Gain)	5
LPC	41
Phase	6
Total	60

Table 3.4 SDSC with phase bit allocation.

3.4 SDSC versus LPC-10 and MELP

Experimental speech graphs of a female and male tester are shown in Appendix. In each graph, plot (a) is the original signal, (b) is the speech, generated using SDSC with phase information, (c) and (d) are SDSC with bit-rates 9.7 kbps and 12.1 kbps explained in Chapter 2.6 and lastly, (e) and (f) are LPC-10 and MELP (Mixed Excited Linear Prediction) coded signals. These graphs give us an idea about the synthesized coders.

SDSC with bit-rates 9.7 kbps and 12.1 kbps, SDSCP and MELP have a higher speech quality than LPC-10. Besides, SDSCP and MELP compete with each other. The Mean Opinion Score (MOS) [19] rates of each coder are shown in Table 3.5.

Coder	MOS	Bit-rate (kbps)
LPC-10	2.3	2.4
MELP	3.2	2.4
SDSC (phase)	3.5	2.7
SDSC (3-bits)	3.7	9.7
SDSC (4-bits)	4.0	12.1

Table 3.5 MOS rates of different coders.

Chapter 4

Application Areas of the SDSC Encoder

SDSC speech quality is better than LPC with the same bit rates. Because of the compression rate and quality, SDSC can be used in real life scenarios. In addition, SDSC is a computationally efficient secure encoder, if the codebook table of SDSC is kept a secret between the two users who want to communicate with each other. In this chapter, first real life usage of SDSC is explained and then the security issues of SDSC is analyzed.

4.1 Practical use of SDSC Encoder

4.1.1 Basic Initiation Scenario

Basic call initiation scenario of two sides (side A and side B) is shown in Figure 4.1. First the caller side A sends its own codebook table to side B. Side B saves the codebook table of side A and sends an acknowledgement message (ACK) with its own codebook table to side A. After side A takes the codebook table of side B, side A saves the codebook table of side B and sends an ACK, then the

conservation begins. In this scenario, the codebook tables can be sniffed in the network by an attacker; therefore, the codebook table would not be a secret between two sides A and B. Our solution to the sniff problem is explained in the next section.

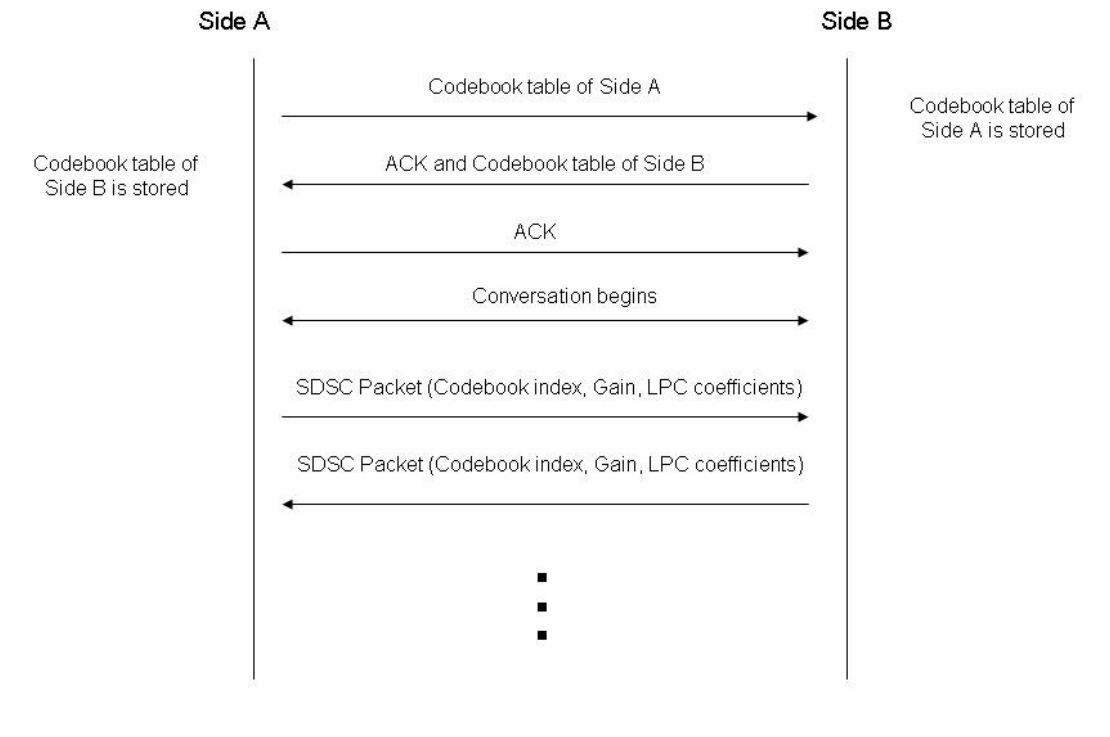


Figure 4.1 Basic Initiation Scenario of SDSC.

4.1.2 Basic Initiation Scenario by Sending Encrypted Codebook Tables

To achieve a secure communication with SDSC encoder the codebook tables should be secret between the two sides. The scenario is summarized in Figure 4.2. In this scenario side A encrypts its own codebook table with the public key of side B and sends it to side B. Side B decrypts the encrypted codebook table of side A with its private key and saves the codebook table of side A. Then side B

encrypts its own codebook table with the public key of side A and sends an ACK with this encrypted table. Side A takes the encrypted codebook table of side B and decrypts it with its private key, then saves the table. Afterwards, side A sends an ACK and the conservation begins. In this scenario, the codebook table is protected from sniffing attack, but what happens if the attacker sets up a conversation with side A and side B before, because in this case the attacker has the codebook tables of both side A and side B. To prevent from this type of an attack, different tables should be sent by the callers for different conversations.

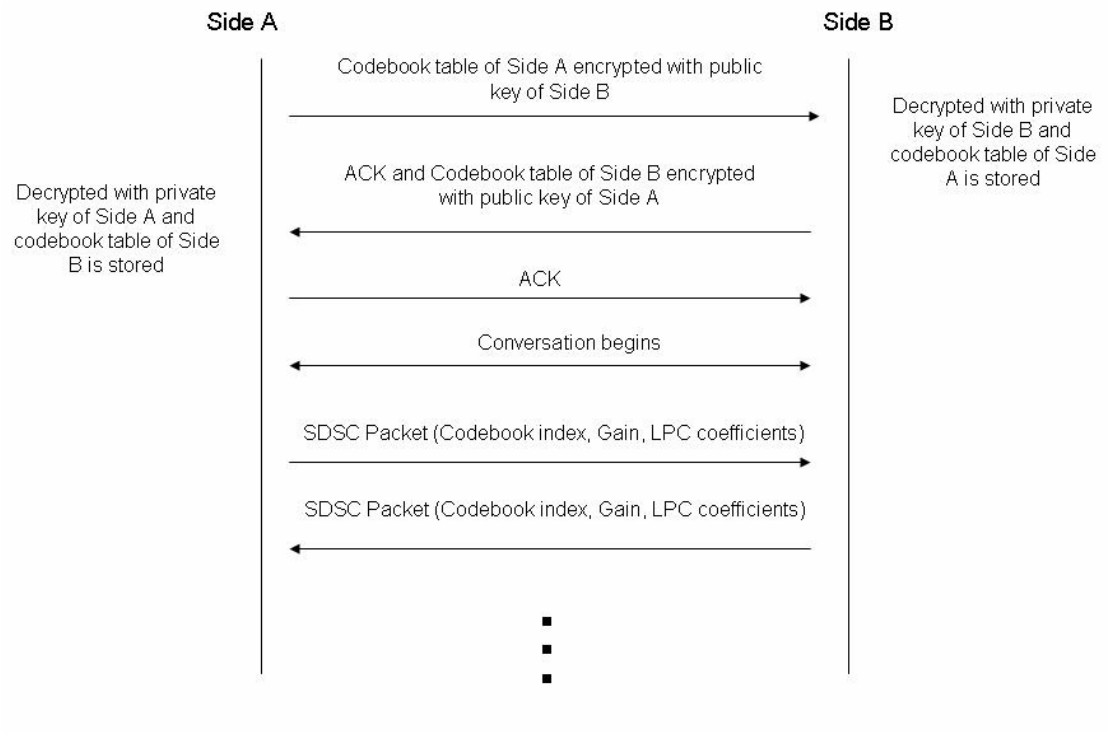


Figure 4.2 Basic Initiation Scenario of SDSC with encrypted codebook tables.

The following cases show how different codebook tables can be sent for different conversations. In Figure 4.3, there is a big codebook table at each side, and in conversation, a subtable is generated by random selection from the big codebook table and sent.

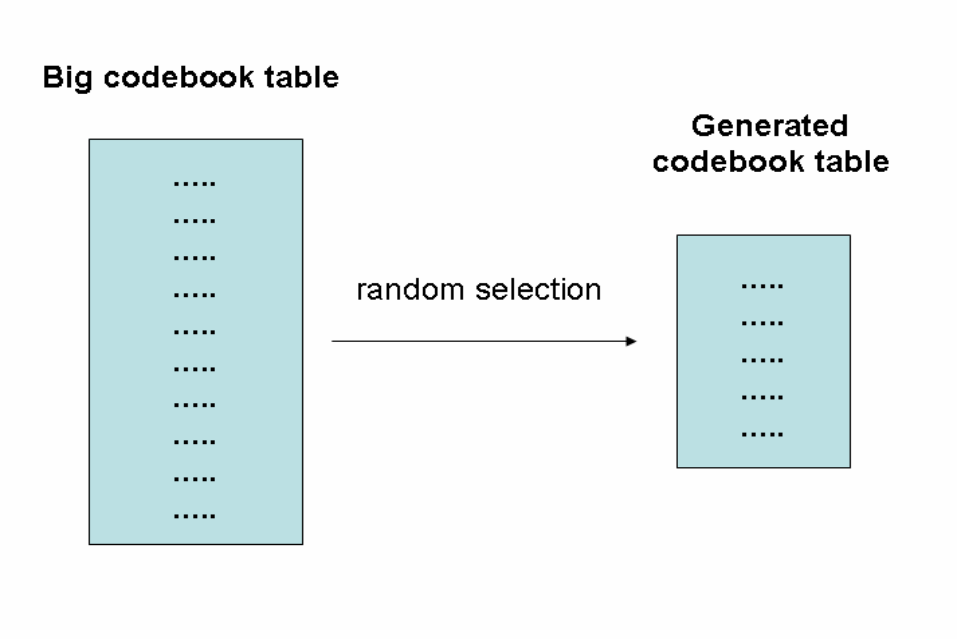


Figure 4.3 A small but different codebook generation from the big codebook table.

Another case is shown in Figure 4.4. The codebook table elements are permuted and a new codebook table is generated with the same signals with different indices.

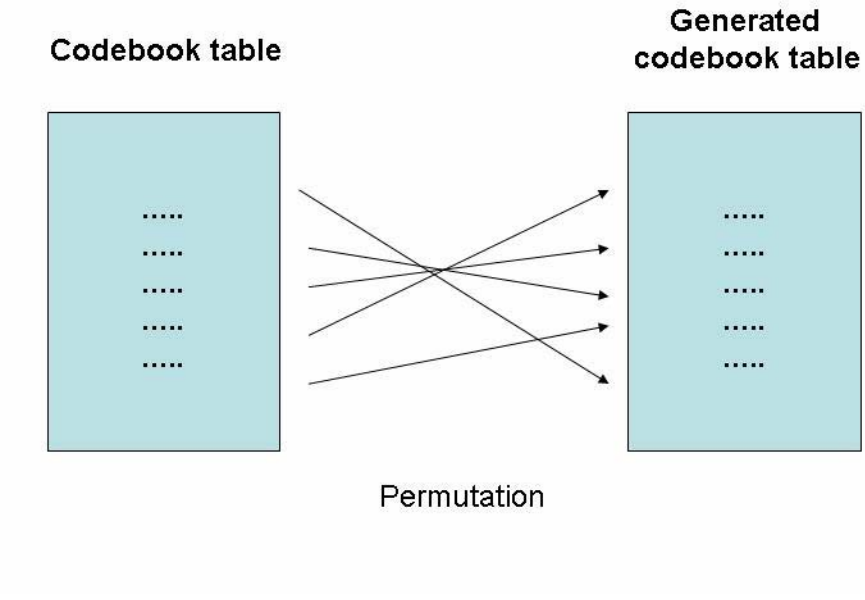


Figure 4.4 Codebook generation using permutation from the original codebook table.

4.1.3 SDSC with SIP Integration

Speaker dependent speech coding can be integrated with most common VoIP initiation protocols like Session Initiation Protocol (SIP). The session initiation of two sides is figured out in Figure 4.5. Side A sends an INVITE packet to side B. In this INVITE packet, side A tells its IP, Real-time Transport Protocol (RTP) [20] port and the speech codec which it supported. Afterwards, side B takes the INVITE packet and, because side B is ready for a call, sends RINGING packet to side A. When side B off hooks the phone, it sends an OK packet to side A. In this OK packet, side B sends its own IP, RTP port and the codec which it selects from side A choices. After side A takes the OK packet, side A sends an ACK and RTP session is established.

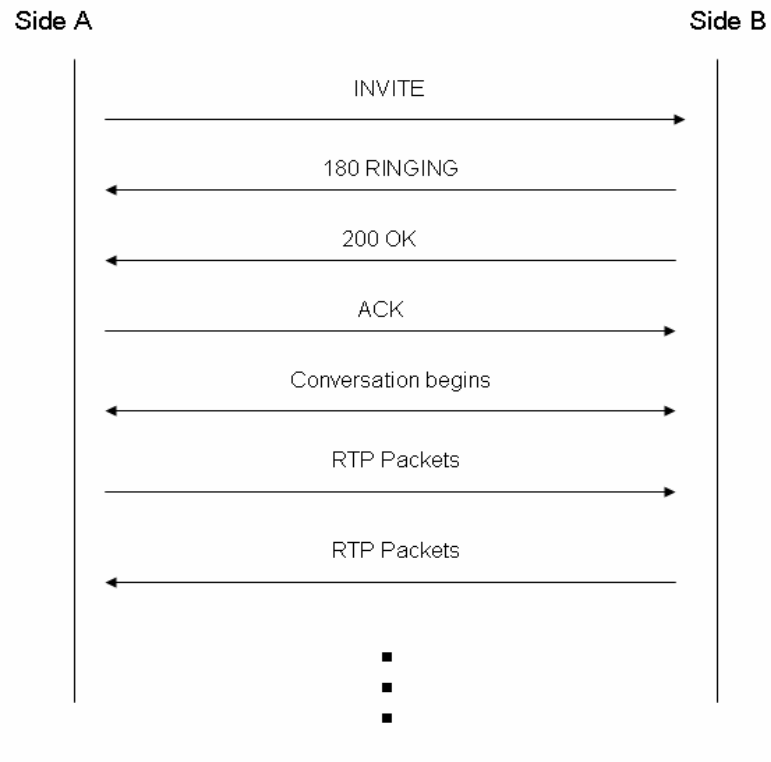


Figure 4.5 SIP Initiation Scenario.

This is the basic SIP scenario and we will integrate the SDSC in this scenario. In SDSC initiation scenario, the codebook tables should be sent to the both sides. Thus, after side B selects the SDSC codec and sends an OK packet, side A sends an ACK packet. The RTP connection is not established in this case. Before the RTP session, both sides send the SIP INFO packets that contain the codebook table of each side to each other. The codebook tables are first encrypted explained in section 4.1.2 and then converted using Base64, because SIP is a text based protocol. This procedure is summarized in Figure 4.6.

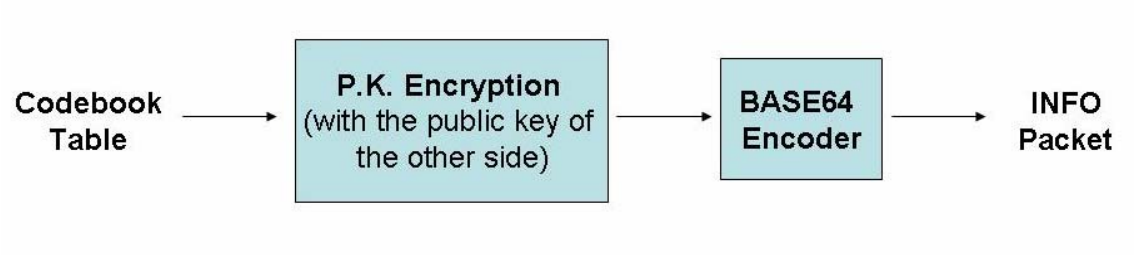


Figure 4.6 SIP INFO packet creation.

After all the codebook table INFO packets are sent, both sides send INFO packet that says sending codebook table is completed. After the end of codebook table packet, RTP session is established. Figure 4.7 illustrates this scenario.

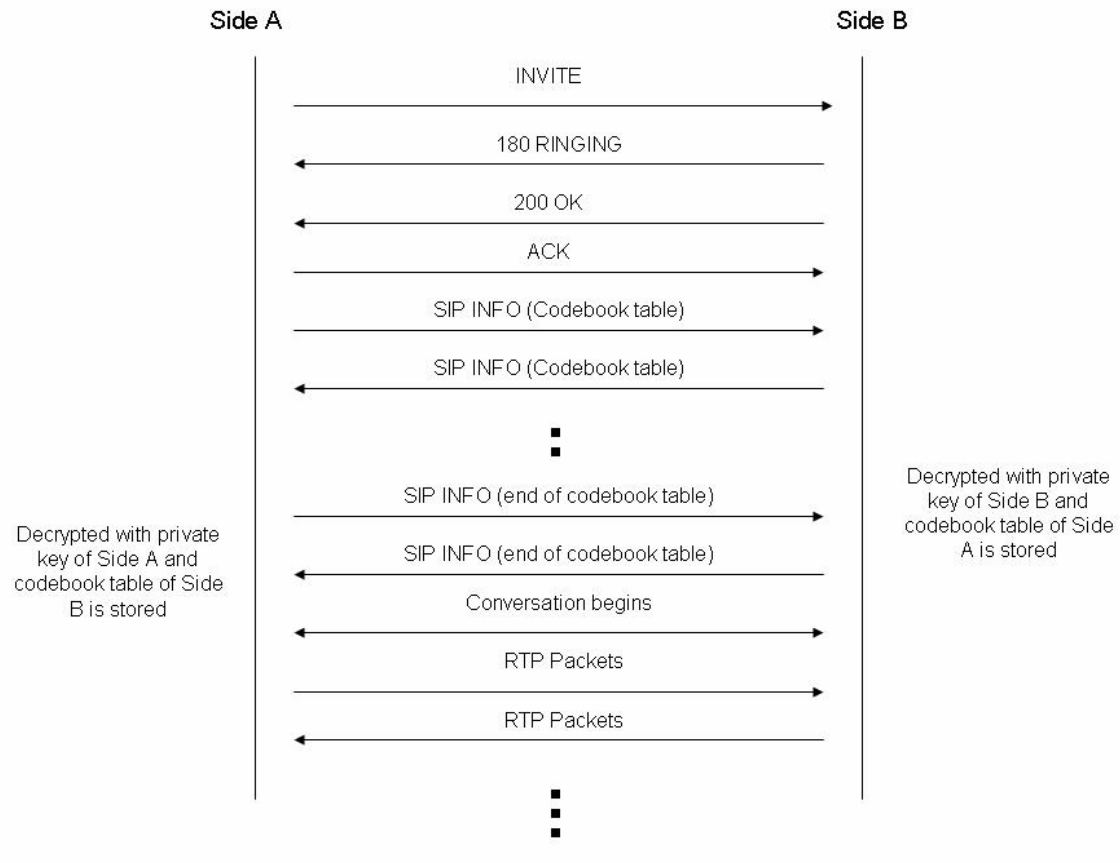


Figure 4.7 SDSC integration into the SIP Scenario.

4.2 Security analysis of the SDSC encoder

The LPC is designed for military purposes. SDSC also can be used in the military communications. Because the security is important in the military communications, SDSC provides a way of secure communication with no extra encryption during communication, data payload and processing time. The codebook table can be considered to be the private key of the communication.

Anyone except sender and receiver must not have any knowledge about the private key (codebook table). On the other hand, there is a risk of predicting the voice signal using cryptanalysis methods, although the cryptanalyst does not have

any knowledge about the codebook table. This cryptanalysis method is the following: If a cryptanalyst can predict the pitch periods of the frames (unvoiced frames are supposed to be zero pitch period), he/she can generate an understandable sound by using LPC speech synthesis method. This is possible because gain and LPC filter coefficients can be known using sniffing attacks in the network. To prevent from this cryptanalysis method, the gain and LPC coefficients should be also secret. To achieve this, the following procedure can be used: We know that gain and parcor coefficients are encoded. We can generate new encode tables for gain and parcor coefficients using the permutation applied to the codebook table shown in Figure 4.4. These new generated encode tables are first encrypted and then sent to the receiving end as in the codebook table. In the communication, these tables are used to encode gain and LPC coefficients and no extra encryption is needed during communication. Gain and LPC coefficients are now secret and it is impossible to synthesize an understandable sound without knowing gain, LPC coefficients and pitch period information.

To sum up, SDSC provides a high level security with no extra encryption during communication, data payload and processing time, when

- The codebook table is secret between the sender and receiver.
- The encode tables for the gain and LPC coefficients are secret between the sender and receiver.
- Attacker doesn't have any statistical information about the sound of speakers.

Chapter 5

Conclusions

Voice over IP (or Internet telephony which is almost the same thing) is one of the several technologies that allow us to make phone calls over the Internet instead of the regular telephone network. Some more advanced and secure systems use a private data network instead of the Internet. This technology has been around since the 1970s but hasn't been practical until recently because we needed a broadband/high-speed connection for it to be effective. Specifically, we need bit rates of more than 100kbps per connection using modern VoIP transmission technologies. This has only recently become common among residential broadband subscribers. That kind of bandwidth has been available in businesses for longer and the technology is already well established in the business market – but even there the necessary broadband has only been commonly available for the last four years. [21] Bandwidth is available for a VoIP transmission, but companies want to make as many VoIP calls as possible without losing any speech quality. Therefore speech compression is very important in this case. Although many speech codecs have very high compression rates and high speech quality, speaker dependent speech coding provide us a new aspect of speech

compression, which is to using human specific speech characteristics and also provide a good quality speech in spite of its low bit rate.

The advantages of SDSC compared to LPC are as follows. Firstly, in SDSC system, more qualitative speech can be obtained in the same bit rate compared to LPC system. Besides, LPC has a limitation so that LPC uses strictly random noise or strictly periodic impulse train as excitation which does not match practical observations using real speech signals [14]. On the other hand, SDSC uses a codebook table which is generated from the speaker dependent excitation signals and excitations are extracted from real speech signals and match practical observations. As an improved model of SDSC, the phase information is preserved. Besides, SDSC provides a way of secure communication with no extra encryption during communication, data payload and extra processing time. There are also some disadvantages of SDSC. Speech frames cannot be classified as strictly voiced or unvoiced. This is a limitation of both LPC and SDSC systems. In addition, codebook table in SDSC system adds an extra burden of storage. The caller has to have the codebook table of the called user, which requires an extra storage and prior transmission of the codebook table.

Future work is needed to incorporate the SDSC to a practical speech coding system. This is possible because SDSC can be easily integrated into Session Initiation Protocol (SIP) as mentioned in Chapter 4.1.3. We also need to improve the speech quality at low bit rates. The speech quality of SDSC at higher bit-rates explained in Chapter 2.6 can be improved by using non-uniform quantizers or using transform coding to pack the excitation difference signal with lower bit-rates. Speech quality can be improved by using bigger codebook tables. In addition, codebook tables can be generated using the best excitation signals representing the speaker characteristics. On the other hand, the bigger the codebook is, the slower the search process. Therefore, an optimal solution has to be determined between these two constraints, the codebook size and time required for the search process.

Appendix A

Speech Signal Figures of Male and Female Test Sounds in English

Figure A.1 Female-1 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 40
Figure A.2 Female-2 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 41
Figure A.3 Female-3 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 42
Figure A.4 Female-4 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 43
Figure A.5 Male-1 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 44
Figure A.6 Male-2 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 45
Figure A.7 Male-3 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 46
Figure A.8 Male-4 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech. 47

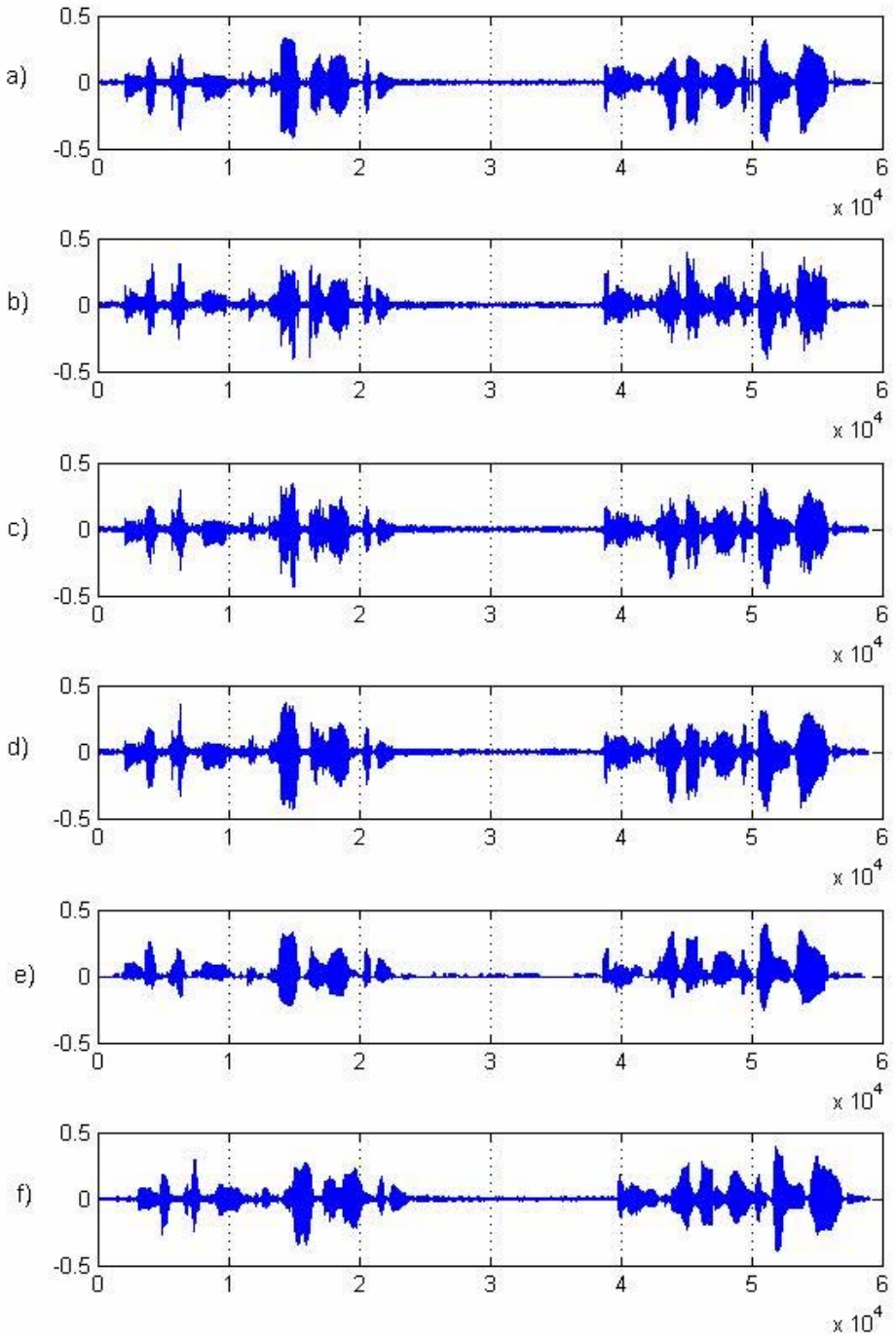


Figure A.1 Female-1 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech.

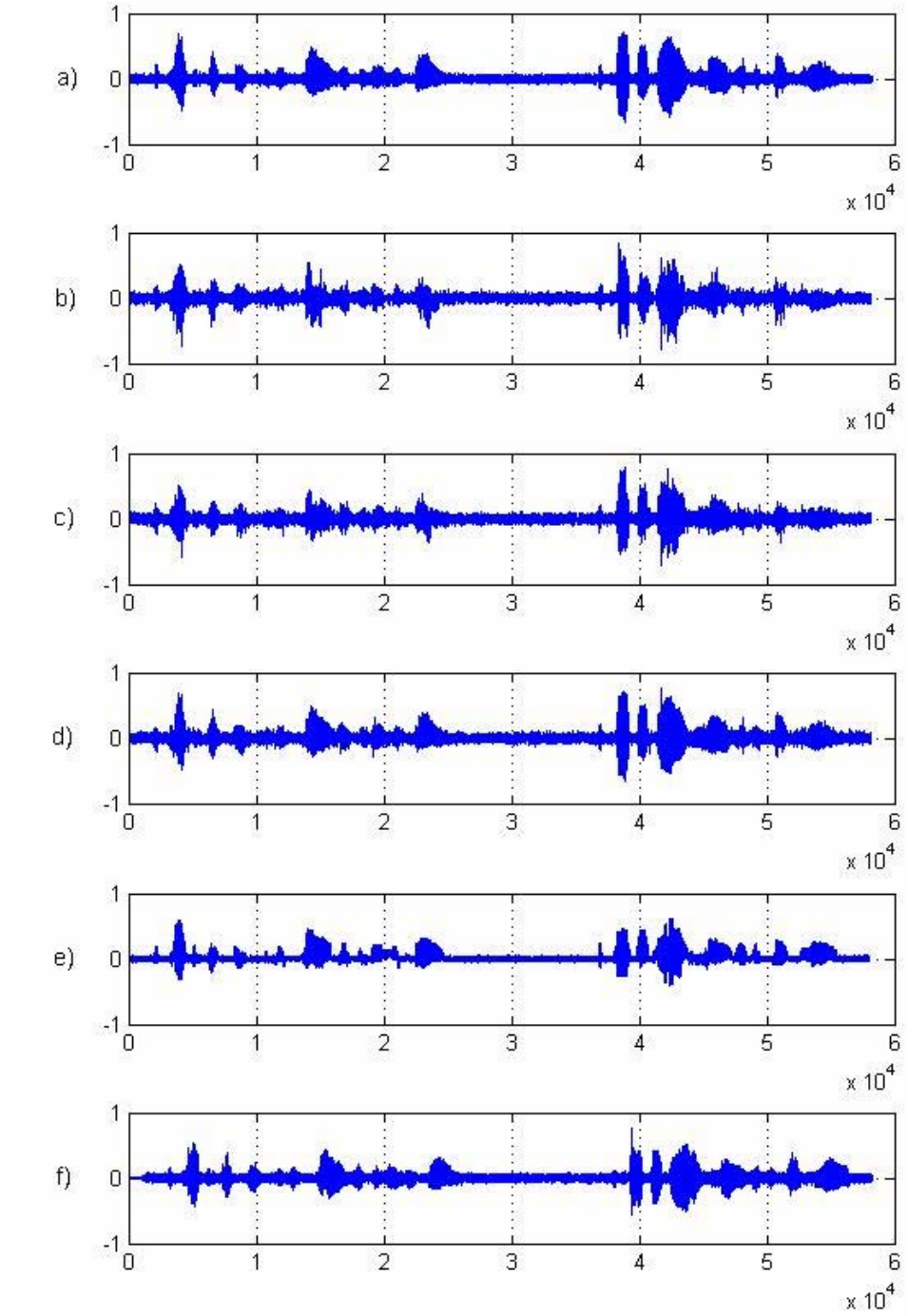


Figure A.2 Female-2 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech.

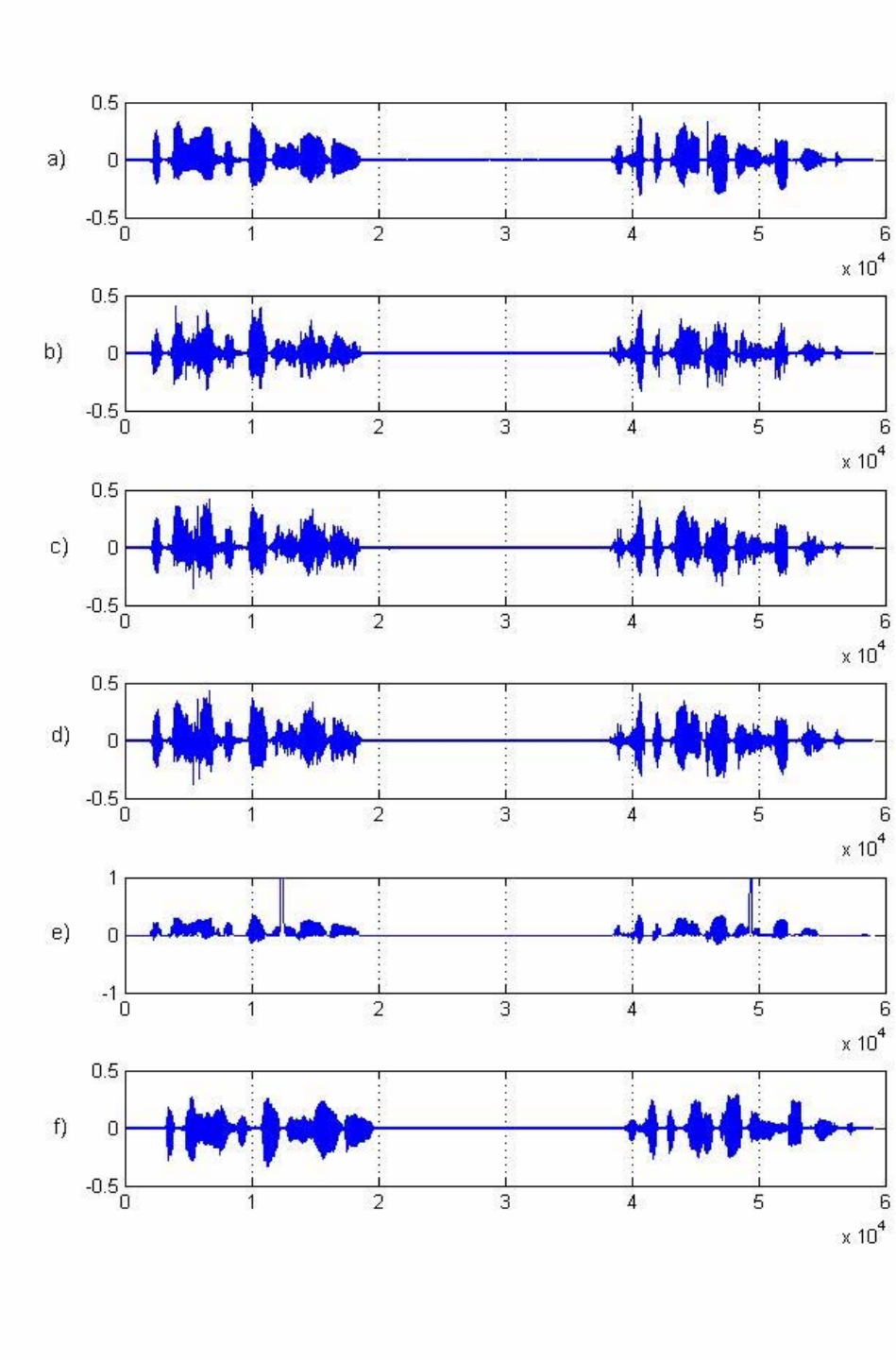


Figure A.3 Female-3 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech.

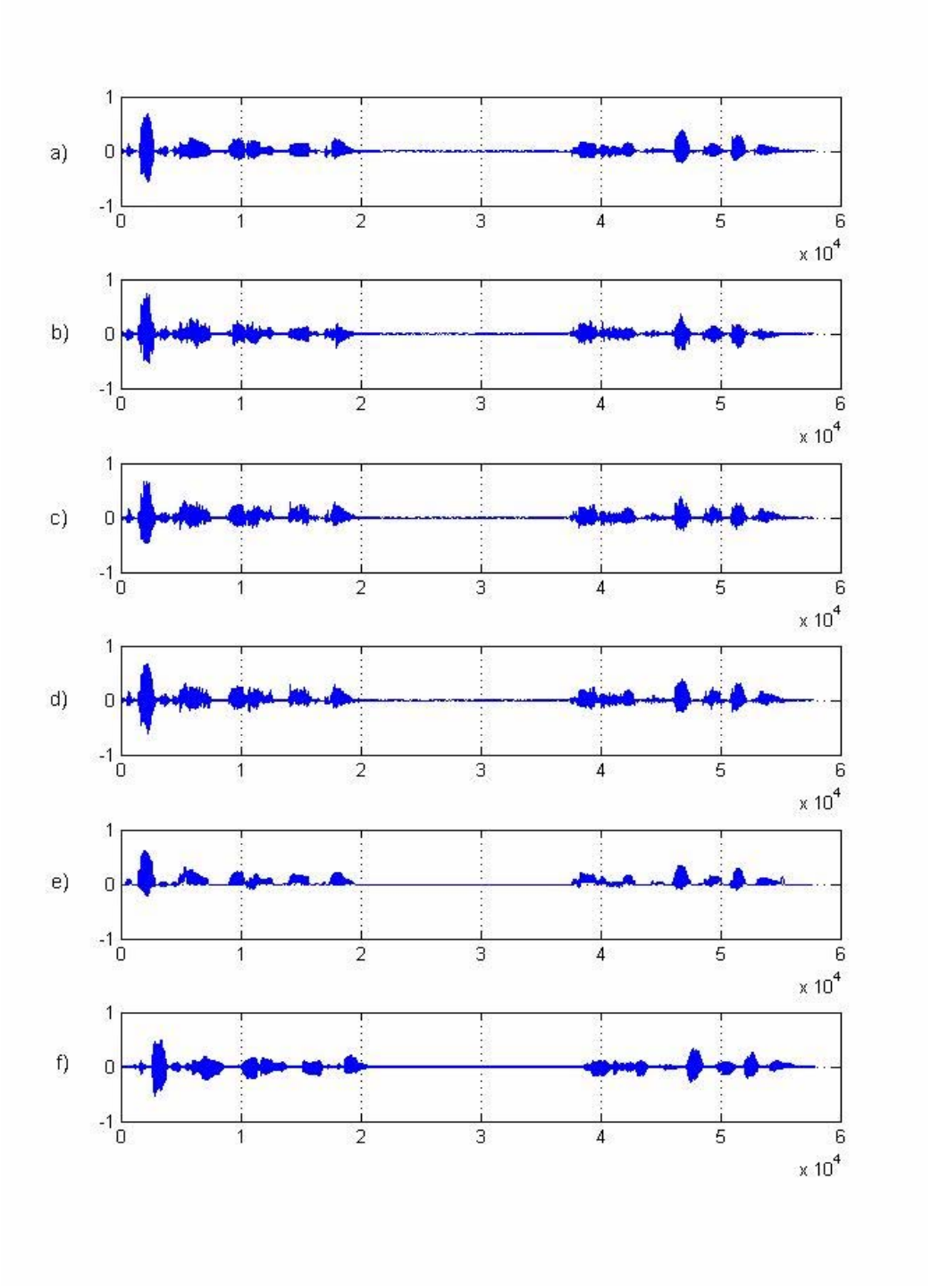


Figure A.4 Female-4 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech.

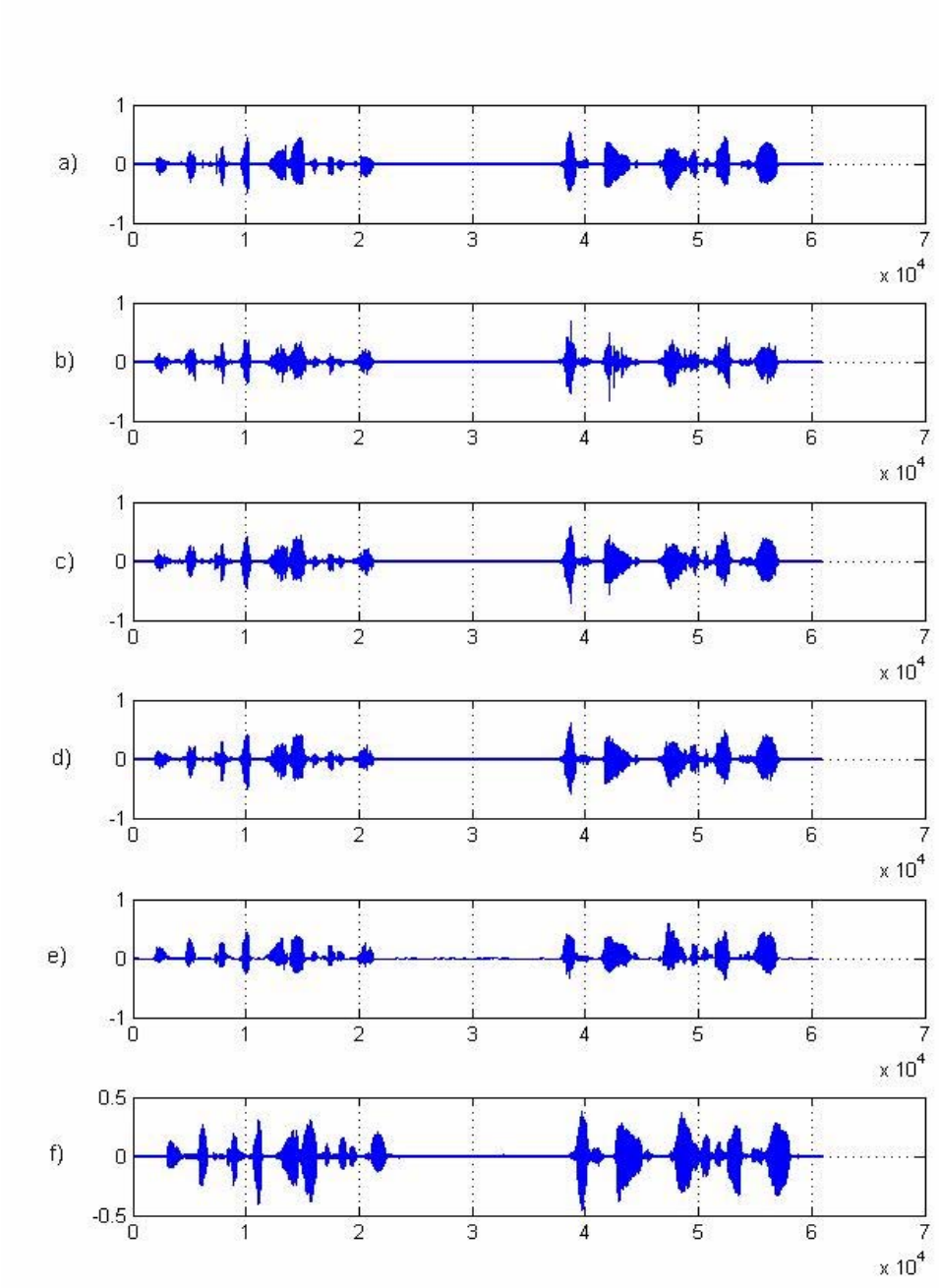


Figure A.5 Male-1 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech.

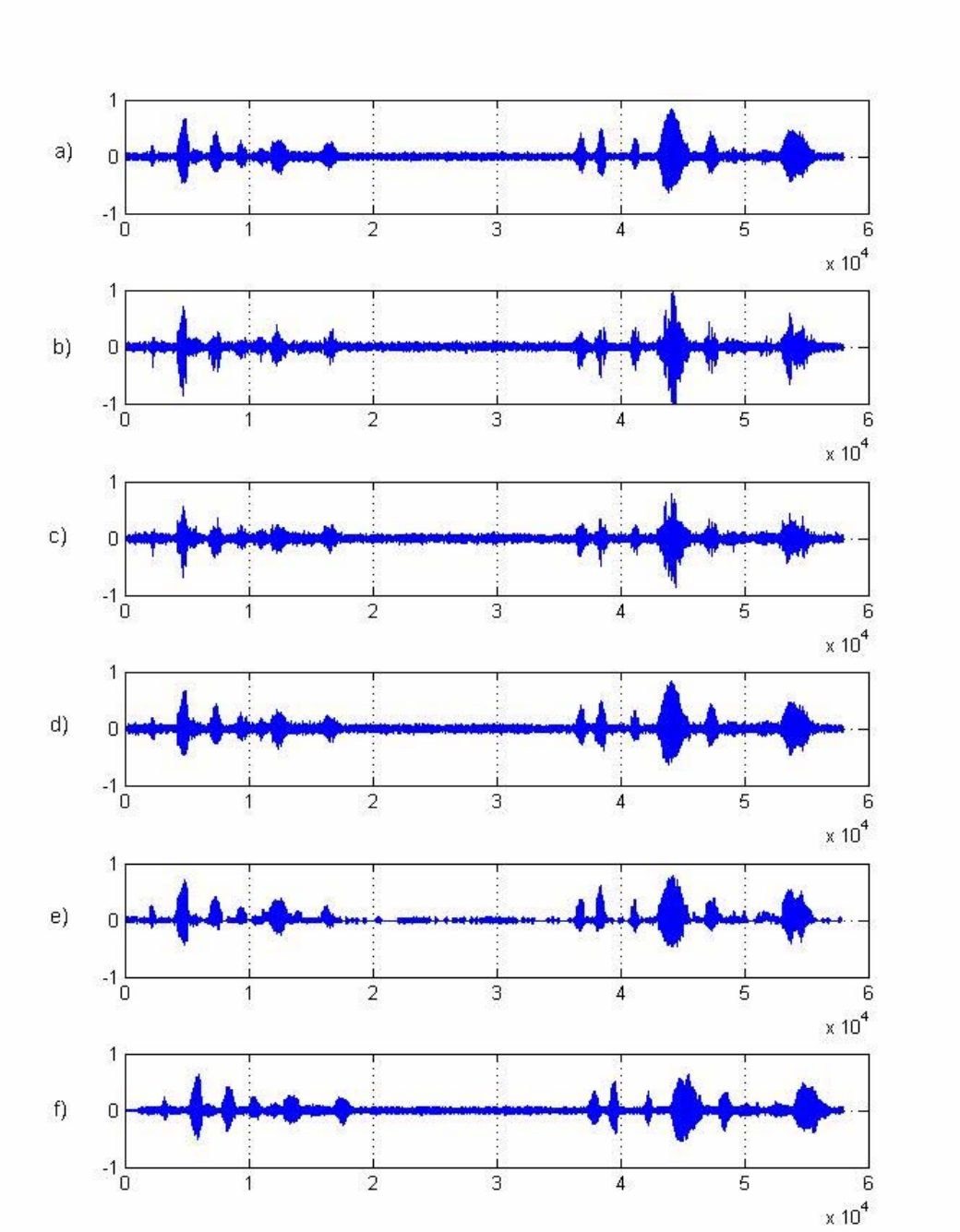


Figure A.6 Male-2 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech.

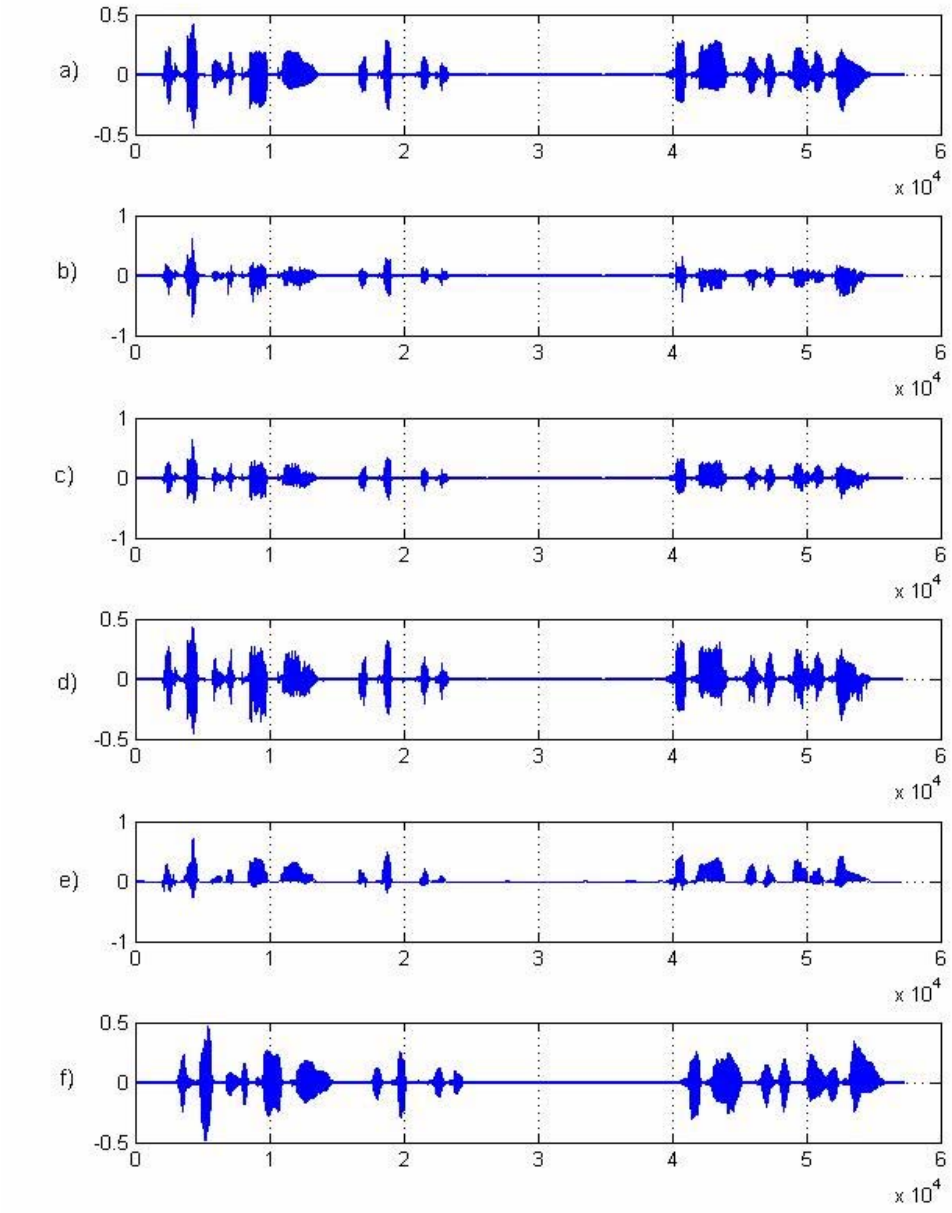


Figure A.7 Male-3 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech.

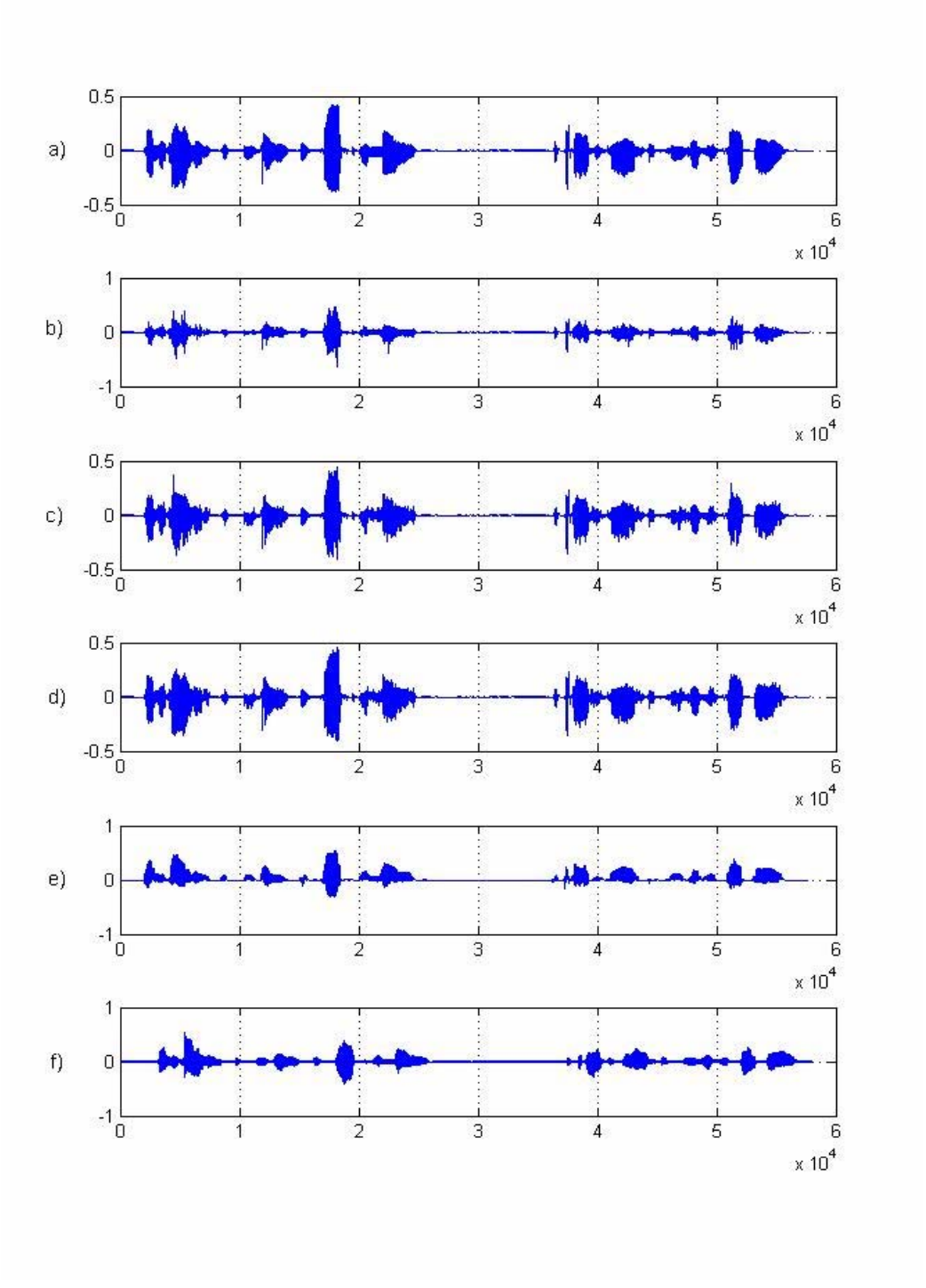


Figure A.8 Male-4 a) original speech, b) SDSC with phase information speech, c) 9.7 kbps SDSC speech, d) 12.1 kbps SDSC speech, e) LPC-10 coded speech, f) MELP coded speech.

BIBLIOGRAPHY

[1] William C. Barker. (May 2004) NIST, *Recommendation for the Triple Data Encryption Algorithm (TDEA) Block Cipher*, Special Publication 800-67. http://csrc.nist.gov/publications/nistpubs/800-67/SP800-67.pdf

[2] J. Daemen and V. Rijmen, *AES Proposal: Rijndael*, AES Algorithm Submission, September 3, 1999, available at AES page available via http://www.nist.gov/CryptoToolkit

[3] Vocal Technologies, Ltd Homepage. (2004). *G711: Pulse Code Modulation (PCM) of Voice Frequencies.* 21 August 2006 http://www.vocal.com/data_sheets/g711.html

[4] Vocal Technologies, Ltd Homepage. (2004). *G722: 7 kHz audio coding within 64 kbit/s using Sub-Band Adaptive Differential Pulse Code Modulation (SB-ADPCM).* 21 August 2006 http://www.vocal.com/data_sheets/g722.html

[5] Vocal Technologies, Ltd Homepage. (2004). *G722.1: Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss.* 21 August 2006. http://www.vocal.com/data_sheets/g722d1.html

[6] Vocal Technologies, Ltd Homepage. (2004). *G722.2: Adaptive Multi-rate Wideband AMR-WB Vocoder Algorithm.* 21 August 2006 http://www.vocal.com/data_sheets/g722d2.html

[7] Vocal Technologies, Ltd Homepage. (2004). *G723.1: Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s.* 21 August 2006. http://www.vocal.com/data_sheets/g723d1.html

[8] Vocal Technologies, Ltd Homepage. (2004). *G726: 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM).* 21 August 2006 http://www.vocal.com/data_sheets/g726.html

[9] Vocal Technologies, Ltd Homepage. (2004). *G727: 5, 4, 3 and 2 – bits sample embedded Adaptive Differential Pulse Code Modulation (ADPCM).* 21 August 2006 http://www.vocal.com/data_sheets/g727.html

[10] Vocal Technologies, Ltd Homepage. (2004). *G728: Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction.* 21 August 2006. http://www.vocal.com/data_sheets/g728d0.html

[11] Vocal Technologies, Ltd Homepage. (2004). *G729: Coding of Speech Signals at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP).* 21 August 2006 http://www.vocal.com/data_sheets/g729d0.html

[12] Thomas E. Tremain, "The Government Standard Linear Predictive Coding Algorithm: LPC-10," *Speech Technology Magazine*, April 1982, p. 40-49.

[13] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 10, pp. 937-940, 1985.

[14] Wai C. Chu., *Speech Coding Algorithms Foundation and Evolution of Standardized Coders*. Wiley, 2003.

[15] I. Wakeman. (31 January 2002). *5.6 Code Excited Linear Predictor (CELP)*. 21 August 2006. http://www.cogs.susx.ac.uk/users/ianw/teach/ms/node71.html

[16] A. V. McCree, K. Truong, E. B. George, T. P. Barnwell, and V.Viswanathan (1996). "A 2.4 kbit/s MELP Coder Candidate for the New U.S. Federal Standard," in *IEEE ICASSP,* pp. 200-203.

[17] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler (June 2002). *SIP: Session Initiation Protocol.* 21 August 2006 http://www.ietf.org/rfc/rfc3261.txt

[18] J. Brudbury (5 December 2000). *Linear Predictive Coding.* 21 August 2006 http://www.ece.ubc.ca/~elec466/lpc_paper.pdf

[19] G. Liu, *Mean Opinion Score.* 21 August 2006. http://umsis.miami.edu/~gliu/

[20] H. Schulzrinne, S. Casner, R. Frederick, V. Jaconson. (July 2003). *RTP: A Transport Protocol for Real-Time.* 21 August 2006 http://www.ietf.org/rfc/rfc3550.txt

[21] O. Linderholm. (20 Juny 2006). *Why VoIP?*. 21 August 2006. http://www.voip-news.com/news/why-voip-2404006/