1

of/de

1

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur SRC 1970, c. C-30, et ses amendements subséquents.

Acknowledgments

Firstly, I am grateful to my supervisor Dr. Tadeusz Kwasniewski for all that he has done and been for me; as a friend, advisor, professor and fellow researcher.

Secondly, I would like to thank Dr. John Knight and Dr. Patrick van der Puije for their help and support. Also, the staff at the Department of Electronics, namely Mrs. Barbara Lynn.

Thirdly, I would like to acknowledge two of my colleagues who were an important part of this research effort, Timothy Rahrer and Arthur Chee.

Finally, I would like to thank my family who without their support this degree would not have been possible.

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ISBN 0-315-60466-2

Canada

The undersigned recommend to the Faculty of Graduate Studies and
Research the acceptance of the thesis:

“CEPSTRAL ANALYSIS:
A SPEECH PROCESSING STRATEGY FOR THE COCHLEAR IMPLANT”

submitted by Ibrahim Joseph Gedeon in partial fulfillment of the
requirements for the degree of Master of Engineering.

Prof. T.A. Kwasniewski
Thesis Supervisor

Prof. J. S. Wight
Chairman,
Department of Electronics

List of Abbreviations

approx.	approximately
DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
FFT	Fast Fourier Transform
ICASSP	Int'l Conference on Acoustics, Speech and Signal Processing
ILS	Interactive Laboratory System
Int'l	International
LPC	Linear Predictive Coding
MAC	Minimal Auditory Capabilities
M.I.T.	Massachusetts Institute of Technology
PARCOR	Partial Correlation
R-n	Radix size n
RAM	Random Access Memory
ROM	Read Only Memory
SIFT	Simplified Inverse Transform
VLSI	Very Large Scale Integration

Acknowledgments

Firstly, I am grateful to my supervisor Dr. Tadeusz Kwasniewski for all that he has done and been for me; as a friend, advisor, professor and fellow researcher.

Secondly, I would like to thank Dr. John Knight and Dr. Patrick van der Puije for their help and support. Also, the staff at the Department of Electronics, namely Mrs. Barbara Lynn.

Thirdly, I would like to acknowledge two of my colleagues who were an important part of this research effort, Timothy Rahrer and Arthur Chee.

Finally, I would like to thank my family who without their support this degree would not have been possible.

Abstract

This thesis presents the implementation of a novel speech processing strategy for the cochlear implant. Cepstral speech analysis is proposed for speech feature extraction use in the cochlear implant. The selection of cepstral analysis is discussed in light of past, current and undergoing cochlear implant research. An evaluation strategy of the cepstral system is proposed and implemented. The results and the strategy are presented.

Cepstral analysis is implemented in real time using the Texas Instruments TMS320c25 digital signal processor. The implementational techniques and specifications of the cepstral analysis system are also discussed in this thesis.

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Background Information	1
1.3	Thesis Organization	5
2	Speech Processing for the Cochlear Implant	6
2.1	Speech	6
2.1.1	<i>Speech Model</i>	6
2.1.2	<i>Speech Features</i>	6
2.1.2.1	<i>Spectrogram</i>	7
2.1.2.2	<i>Pitch</i>	8
2.1.2.3	<i>Formants</i>	9
2.2	Signal Processing Tools for Feature Extraction	10
2.2.1	<i>Filter Bank</i>	10
2.2.2	<i>The Fourier Transform</i>	10
2.2.3	<i>Convolution and Autocorrelation</i>	12
2.2.4	<i>Predictive Coding</i>	13
2.2.5	<i>The Cepstrum</i>	17
2.3	The Current Spread Problem	19
2.3.1	<i>Theory</i>	20
2.3.2	<i>Algorithm: Formulation and Implementation [Town 87]</i>	22
2.4	Signal Processing in Cochlear Implant Research Projects	23
2.4.1	<i>A General View</i>	23
2.4.2	<i>London</i>	23
2.4.3	<i>Los Angeles</i>	24
2.4.4	<i>Vienna</i>	25
2.4.5	<i>San Francisco</i>	25
2.4.6	<i>Paris</i>	25
2.4.7	<i>Stanford</i>	25
2.4.8	<i>Melbourne</i>	26
2.4.9	<i>Boston</i>	26
2.4.10	<i>Tokyo</i>	26
2.4.11	<i>Nimes</i>	27

2.4.12	<i>Ottawa-Sherbrooke</i>	28
2.4.13	<i>Summary</i>	28
2.5	Suggested Signal Processing Strategy	29
3	Cepstral Analysis: Software-Hardware	32
3.1	Theory of Cepstral Analysis	32
3.1.1	<i>Signal Cepstral Analysis</i>	32
3.1.2	<i>Speech Cepstral Analysis</i>	33
3.2	Software	35
3.2.1	<i>The Fast Fourier Transform</i>	38
3.2.2	<i>The Magnitude Routine</i>	38
3.2.3	<i>The Logarithm Routine</i>	39
3.2.4	<i>The Feature Extraction Routines</i>	40
3.2.5	<i>The Windowing (Liftering) Routine</i>	40
3.2.6	<i>Interrupt Service Routine</i>	42
3.3	Hardware	42
3.3.1	<i>Processor Selection</i>	42
3.3.2	<i>The TMS320C25</i>	44
3.3.3	<i>The Board Level Implementation</i>	44
3.3.4	<i>Hardware Summary</i>	45
3.3.5	<i>The Cochlear Implant System Hardware</i>	45
4	System Evaluation and Performance	47
4.1	Integrity of Cepstral Analysis Development	47
4.2	Validity of The Cepstral Implementation	48
4.2.1	<i>Non-Speech Data</i>	49
4.2.2	<i>Cepstrum Comparison</i>	50
4.3	The Cochlear Implant System Hardware	51
4.4	Discriminatory Capabilities of Cepstral Analysis	51
5	Conclusions	54
5.1	Achievements and Contributions	54
5.2	Future Work	54
	References	60
	Bibliography	65

List of Figures

1-1	An Analog Cochlear Implant System	3
1-2	A Digital Cochlear Implant System	3
1-3	Frequency Mapping along the Basilar Membrane	4
2-1	Cepstral Voiced Speech Model	7
2-2	The All-Pole Filter Model (LPC)[OpSc 75]	8
2-3	The Spectrogram of a Speech Sample [OpSc 89]	9
2-4	LPC Basic Speech Synthesis Model	15
2-5	Block Diagram of SIFT[RaSc 78]	17
2-6	Cepstral Analysis for Pitch Detection[ScRa 70]	18
2-7	The Resistor Model of The Cochlea[GeCh 87]	21
2-8	Error vs n_s [Town 87]	23
2-9	Threshold Theory Current Correcting Method	24
2-10	Hokkaido's speech processor, an eight channel implant[IfWh 86]	27
2-11	Genin-Charachon: Analog circuitry for 1 channel [GeCh 86]	28
2-12	The rolling process of the electrode made at Carleton	29
3-1	The Cepstrum of a Speech Frame	36
3-2	Cepstral Analysis, A Block Diagram	37
3-3	The Board	45
4-1	The Test Speech Sample "We Were Away A Year Ago"	48
4-2	The Spectrogram of The Test Sample	49
4-3	Tracking of The First Three Formant Frequencies	50
4-4	Target System Hardware	52
4-5	Discriminatory Test Using "atta" and "adda"	53

List of Tables

2-1	The Various Cochlear Implant Groups	30
3-1	Timing and Memory Information for The Routines	43
4-1	Non-Speech Data Test	51

List of Abbreviations

approx.	approximately
DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
FFT	Fast Fourier Transform
ICASSP	Int'l Conference on Acoustics, Speech and Signal Processing
ILS	Interactive Laboratory System
Int'l	International
LPC	Linear Predictive Coding
MAC	Minimal Auditory Capabilities
M.I.T.	Massachusetts Institute of Technology
PARCOR	Partial Correlation
R-n	Radix size n
RAM	Random Access Memory
ROM	Read Only Memory
SIFT	Simplified Inverse Transform
VLSI	Very Large Scale Integration

Chapter 1

Introduction

1.1 Motivation

The recent advances in digital signal processing methods and hardware have created new challenges in the implementation of innovative real-time speech processing algorithms for a number of applications, such as telecommunications, cochlear implants, radar etc. The need for an accurate, real-time speech processing strategy for the cochlear implant was the main motivation behind this research project. A new speech processing strategy, based on cepstral analysis, is proposed.

1.2 Background Information

The cochlear implant is an auditory prosthesis that bypasses the main functions of the ear and electrically stimulates the auditory nerve using an electrode. The auditory nerve interfaces to the inner ear through hair cells in the cochlea. The cochlea is a cone shaped organ that constitutes the coupling between the eardrum and the auditory nerve. It contains fluids that aid the auditory process, and has approximately 40000 hair cells. A problem in electrically stimulating the cochlea using an electrode is the resulting current spread that occurs due to the highly conductive fluid (perilymph) which fills the cochlea.

In order to understand the system design of the cochlear implant, it is necessary to have an appreciation of the sequence of events that take place in the healthy ear from the incidence of an acoustic wave on the eardrum to the excitation of the acoustic nerve which transmits the signal to the brain.

The incidence of the acoustic wave causes the eardrum to vibrate. The three bones in the middle ear (hammer, anvil and stirrup) conduct the vibration of the eardrum through the oval window to the fluid (perilymph) which fills the scala tympani. The wave travels along the length of the cochlea through the helicotrema and backwards towards the round window. The two waves travelling in opposite directions cause a standing wave on

the basilar membrane whose peak occurs at a point along the length of the cochlea, depending on the instantaneous short-term amplitude spectrum carried by the incident wave. High frequencies produce peaks near the basal end of the cochlea while low frequencies produce peaks closer to the apex. The hair cells closest to the position of the standing wave are triggered by the mechanical vibration and produce an electrical discharge which is communicated by the nerve fibers to the brain where they are perceived as sound. This description, while necessarily an oversimplification of the hearing process is sufficient to develop an understanding of the main features of the cochlear implant.

The most common cause of hearing loss is due to the inability of the hear cells to transduce the mechanical vibration of the basilar membrane to electrical discharge. The cochlear implant is designed to by-pass the process up to the operation of the hair cells and to provide an imitation of the electrical discharge to the nerve endings in the cochlea.

The observation that electricity might provide auditory sensations was first made in 1800 by Volta [Volt 1800]. The first recorded implant was by Djourno and Eyries in 1957 during an operation to remove a tumor from the ear of a patient. A temporary electrode, was placed in the patient's cochlea and stimulated electrically, the patient reported hearing sounds during stimulation [Feig 87]. It is important to mention the work of Simmons, which provided some of the earliest recorded results on cochlear implants [Simm 66].

Cochlear implant systems consist of three basic components, a speech processor, a neural stimulator and an electrode. There are analog and digital designs. A typical analog cochlear implant system is made up of a microphone, a level limiter, a filter bank and an electrode as shown in figure 1-1. A digital system would consist of a microphone, a digital speech processing unit, a transmitter, a receiver/stimulator and an electrode. A typical digital system is shown in figure 1-2.

The implanted human cochlea has the following characteristics [MiTC 84]:

- 1- Electrical stimulation of the basal end of the cochlea produces a sensation similar to high frequencies, while an electrical stimulation of the apical end of the cochlea produces a low frequency sensation. Figure 1-3 shows the frequency distribution along the scala tympani [Simm 77].
- 2- An electrical threshold exists in the amount of stimulation current needed to cause an auditory sensation [Town 87].

Figure 1-1: An Analog Cochlear Implant System

Figure 1-2: A Digital Cochlear Implant System

- 3- Loudness is perceived to be a function of the energy of the transmitted signal [WDDE 87].
- 4- In general, implantees are insensitive to a stimulation rate greater than 300Hz in the case of pulsatile stimulation.

Figure 1-3: Frequency Mapping along the Basilar Membrane

The speech processing strategy, used in auditory prostheses aims to maximize the patient's speech perception. A logical stimulation strategy would take advantage of the frequency selective nature of the cochlea and use a stimulation rate less than 300Hz. The strategy would also apply a comfortable level of loudness to the patient to overcome the

stimulation threshold [StHo 87, Geis 87]. This thesis deals mainly with the speech processing aspects of the cochlear implant system.

1.3 Thesis Organization

This thesis presents a number of cochlear implant systems and their speech processing strategies. Signal processing techniques are introduced outlining their capabilities and the problems involved with their implementation. A real time speech processing method is implemented and proposed for use in the cochlear implant. An evaluation scheme for the proposed strategy is presented along with the results of the evaluation scheme.

Chapter 2

Speech Processing for the Cochlear Implant

Successful Implementation of a speech processing strategy for the cochlear implant requires a thorough understanding of all the variables involved. This chapter discusses the following subjects: speech, signal processing tools, the current spread problem and signal processing strategies currently used in cochlear implant systems. Some speech processing characteristics, such as the temporal nature of speech, will not be addressed in this thesis. due to the scope of this research project.

2.1 Speech

2.1.1 Speech Model

Two speech production models are presented in this section, the cepstral model and the all-pole filter model. The cepstral model considers speech as the time convolution of the vocal tract characteristics and the pitch pulse excitations [RaGo 75]. The all-pole model, or Linear Predictive model, considers speech as the output of an all-pole filter whose transfer function models the vocal tract, and whose excitation produces the pitch pulses for voiced speech [RaSc 78]. The vocal tract transfer function is the direct effect of a combination of three physical entities: the length of the vocal tract, the coupling of the nasal cavities and the glottal flow function [Flan 72]. Figures 2-1 and 2-2 display schematically the speech production process of the cepstral and the all pole models, respectively.

2.1.2 Speech Features

In this section some of the theories and methods used in speech feature extraction are presented. The speech features considered are the pitch, the spectrogram, the formants and the cepstrum.

Speech features are usually defined in both the frequency and time domains. Therefore, it is important at this time to define an operating frequency limit for feature extraction considerations. The human ear perceives sounds up to 20kHz in frequency. Human speech

Cepstral Model for Voiced Speech

(* is the convolution function)

Figure 2-1: Cepstral Voiced Speech Model

however, does not usually exceed 8kHz in frequency content. A conventional telephone system uses an upper cutoff frequency of 3.6kHz which is known to give an intelligible degree of speech perception. Adopting the same upper cutoff frequency and conforming to sampling theory requirements, a minimum of 7.2kHz sampling frequency is required [RaGo 75].

2.1.2.1 Spectrogram

The spectrogram is a graphical representation of the spectrum as it evolves in time, it is the energy of speech displayed with respect to frequency and time. The spectrogram may be extracted using the filter bank method, the Fourier transform method and others. Using the filter bank method, a speech signal is passed through a filter bank and the outputs are plotted as a power variation in time (the outputs are proportional to filter in-band energy). Applying the Fourier transformation method, a set of speech samples is displayed in the frequency domain. The filter bank method provides continuously varying energy information while the Fourier Transform operates on a frame basis. When using the Fourier Transform the output is available at the end of each analysis period. Figure 2-3

Figure 2-2: The All-Pole Filter Model (LPC)[OpSc 75]

shows a spectrogram of a speech sample.

2.1.2.2 Pitch

Pitch, by definition, is the fundamental frequency of speech, it is denoted by F_0 and is a speech feature unique to the speaker. Pitch is proportional to the inverse of the period of the impulse excitations introduced earlier in the speech production models. The pitch impulse excitations are speech independent, they depend on the physical characteristics of the speaker.

It is important to introduce the concept of voiced and unvoiced speech in pitch definitions and calculations. Voiced speech is speech produced in the vocal tract excited by pitch pulses, while unvoiced speech is speech produced by modeling of one's tongue and mouth without pitch excitation. In other words, unvoiced speech is speech whose excitation pulses are undetectable.

Figure 2-3: The Spectrogram of a Speech Sample [OpSc 89]

Voiced speech has a distinct pitch value, while unvoiced speech does not contain pitch information. Most pitch extraction techniques apply to voiced speech. A large number of techniques calculate the pitch by estimating the frequency of maximum energy.

2.1.2.3 Formants

Formants are an important feature for speech recognition applications. Formants are defined as the major resonances in the speech spectrum. Formants are usually tracked and extracted over a segment (frame) of speech. Formant frequencies and their amplitudes depend largely on the speech sample itself rather than the speaker. There are many methods

for formant extraction: LPC and cepstral analysis will be discussed in the next section.

2.2 Signal Processing Tools for Feature Extraction

Speech research has been developing rapidly since the early 1940s, encouraged by interest in military communications during the Second World War. In this section some basic speech processing tools and methods are introduced: filter banks, convolution and auto-correlation, the Fourier transform, predictive coding and the cepstrum.

2.2.1 Filter Bank

A typical filter bank consists of a chosen number of bandpass filters. Speech is passed through each bandpass filter and the output of each filter shows the content of speech at that band. The Fourier transform and filter banks perform similar operations, yielding a common result of energy per cycle. The resolution of the filter bank depends on the number of filters in the bank and the bandwidth of each individual filter. Filter banks are easily realizable, but are limited in their accuracy.

2.2.2 The Fourier Transform

Due to the important usage of the Fourier Transform, this section starts with a theoretical overview of the transform. Consider a signal $x(t)$ of period T (i.e. $x(t)=x(t+T)$) this signal can be transformed into a Fourier series of harmonic terms of the following form [WoFa 85]:

$$x_a(t) = a_0 + \sum_{m=1}^{\infty} (a_m \cos(m\omega t) + b_m \sin(m\omega t)) \quad (2.1)$$

where $\omega = \frac{2\pi}{T}$, a_m and b_m are defined as follows:

$$a_m = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \cos(m\omega t) dt \quad (2.2)$$

and

$$b_m = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \sin(m\omega t) dt \quad (2.3)$$

This transformation may be applied to discrete signals, by assuming the period T mentioned in the above equations to be the duration of the signal.

In digital signal processing applications signals are usually sampled, in this case the sampled Fourier Transform or discrete Fourier transform (DFT) is applied. A finite sampled sequence $x(n)$ of length N is considered periodic and every N samples and may be transformed into the frequency domain using a DFT of the form

$$X(k) = \sum_{n=0}^{N-1} \left(x(n) e^{-j \frac{2\pi kn}{N}} \right) \quad (2.4)$$

where k is an integer such that $0 \leq k \leq N - 1$ and $f_s k/N$ is the frequency content of the signal at the k th point of the transform.

The computational version of the DFT is called the Fast Fourier Transform (FFT), the FFT is more efficient in terms of execution time and memory space consumption than the DFT. The FFT breaks the Fourier computation into several smaller Fourier Transform computations (called passes) instead of one. A Fourier Transformation may be decomposed into any number of passes (>1), the smallest decomposition factor is called the radix.

It is important to note the reduction in the number of required operations between the FFT and DFT. The multiplication operation in an N point DFT will be considered in this discussion. A direct DFT calculation requires N^2 multiplications. Decomposing the N point DFT into 2 ($N/2$) point DFTs requires $2(N/2)^2$ multiplications, a two fold saving [WoFa 85]. Further decomposition gives $4(N/4)^2$ multiplications. Decomposing till the smallest FFT is of size 2, one gets $N(\log_2 N)$ multiplication operations. It can be seen that if the smallest decomposition was 4, the number of multiplications would have been $N(\log_4 N)$, where the radix of the FFT is the base of the Log. Increasing the value of the radix, and hence the minimum FFT decomposition size, the number of multiplications is reduced. This results in an added reduction of the execution time. For DSP FFT applications, the value of the radix is bound by the the number of temporary registers available for intermediate calculations and storage [Morr 85].

The first FFT was implemented by Cooley and Tukey [Berg 69]. The Cooley and Tukey FFT set a basic algorithm that other programmers improved upon by trying to optimize certain computational features of the FFT such as speed or memory. The Prime Factor Algorithm is one of the latest methods used in FFT calculations, developed by Burrus and Eschenbacher, it is considered fast and efficient by many professionals in the field [BuEs 81]. The Winograd algorithm is also considered as the optimal FFT implementation in terms of execution speed and memory requirements [Morr 78]. With the current digital signal

processors the time to perform a multiplication is comparable with the time to perform an addition so there is a potential saving in minimizing both the number of additions and multiplications. This led to the development of the split radix FFT algorithm. The split-radix FFT uses a combination of R-2 and R-4 FFTs aiming to reach an optimum in minimizing both the number additions and multiplications. The split radix algorithm was first implemented by Duhamel and Hollman, it resembles a basic 3 loop Cooley-Tukey FFT [SoHB 86]. Sorensen and Heidman and Burrus have developed a split radix FFT whose efficiency exceeds that of a radix-8 FFT, with a code size comparable to that of a radix-4 FFT, and the flexibility of a radix-2 FFT [SoHB 86].

FFT pitch extraction involves the computation of the harmonic product spectrum which is defined as follows:

$$P_n(e^{j\omega}) = \prod_{r=1}^k \left(X_n(e^{j\omega}) \right)^2 \quad (2.5)$$

where k is a chosen degree for the harmonic spectrum calculation. The product of the harmonics will coincide, but only at the fundamental frequency will they be reinforcing each other. This technique has been found to be especially resistant to additive noise, since the contributions of the noise to $X_n(e^{j\omega})$ does not have a coherent structure when viewed as a function of the frequency. It is observed that unvoiced speech will not exhibit a peak in the log harmonic spectrum P_n' , where $P_n' = \text{LOG} | P_n |$ [RaSc 78]. Also the peak at the fundamental frequency need not be in $X_n(e^{j\omega})$ for there to be a peak in P_n' .

2.2.3 Convolution and Autocorrelation

Some digital signal operations are difficult to perform by applying their defined mathematical format. The ability to transform an operation into either an addition or multiplication in another domain would make the operation manageable. Convolution and autocorrelation are two such techniques.

Discrete convolution is a linear-shift invariant operation applied to an input sequence $x_1(n)$ by another sequence $x_2(n)$ [WoFa 85]. Convolution of $x_1(n)$ by $x_2(n)$ is calculated using the following expression:

$$y(n) = x_1(n) * x_2(n) = \sum_{k=-\infty}^{\infty} [x_1(k)x_2(n-k)] \quad (2.6)$$

where the operation '*' denotes convolution. This computational definition is true for any

two sequences $x_1(n)$ and $x_2(n)$. One of the properties of convolution is that convolution in the time domain corresponds to multiplication in the frequency domain, and vice versa.

The autocorrelation function, assuming a stationary signal, is defined as

$$\Phi(k) = \sum_{m=-\infty}^{\infty} (x(m)x(m+k)) \quad (2.7)$$

For applications relevant to the scope of this thesis, the autocorrelation operation is performed over a finite interval $0 \leq m \leq N-1$. The autocorrelation function has three major properties [RaSc 78]:

- The function is even, $\Phi(k) = \Phi(-k)$
- It has a maximum at $k=0$; $|\Phi(k)| \leq \Phi(0)$ for all k
- The quantity $\Phi(0)$ is equal to the energy for a deterministic signal and the average power for a random or periodic signal.

The autocorrelation function may be used as a basis for pitch extraction. In the autocorrelation pitch extraction method, speech is passed through a low pass filter with approx. 1kHz cutoff frequency and the autocorrelation function is then calculated for time shifts between 0 and 20 ms to cover the expected range of speech pitch periods for both male and female speakers. The time shift at which the function's largest value occurs corresponds to the pitch period of the speech segment under analysis [WoFa 85]. The value of the function when normalized with respect to the speech energy is used to decide whether to classify the speech segment as voiced or unvoiced. The autocorrelation method has many advantages such as accuracy and ease of implementation. Its disadvantage is in the complexity of the computations involved. This rules out its use (when implemented using equation 2.6) for most real-time applications.

2.2.4 Predictive Coding

Linear predictive coding (LPC) will be discussed in this section, other forms of predictive coding such as adaptive predictive coding techniques and applications may be found in [TeIn 1981]. LPC speech analysis is based on the time-varying all-pole filter speech production model. The poles of the transfer function of the all-pole speech production filter, are what is known as the linear prediction coefficients. LPC attempts to model the vocal tract using the prediction coefficients.

Assuming that the time variations of the vocal tract shape can be approximated with sufficient accuracy by a succession of stationary shapes, it is possible to define a transfer function in the z -domain for the filter in figure 2-4. For an all-pole filter, the transfer function $H(z)$ can be written as follows:

$$H(z) = \frac{G}{1 - \sum_{k=1}^p (a_k z^{-k})} \quad (2.8)$$

The filter is fully defined by a gain factor G and p coefficients $a_1, a_2, \dots, a_p$. The linear filter has p poles which are either real or occur in complex conjugate pairs. The number of coefficients p , required to represent any speech segment adequately is determined by many factors such as the length of the vocal tract, the coupling of the nasal cavities, the place of the excitation and the nature of the glottal flow function.

There are many ways to calculate the LPC coefficients. This is usually achieved by calculating the coefficients that minimize the mean square error. The mean square error is the squared difference between the original signal and the signal regenerated using the prediction coefficients, also known as the residual error (squared). The residual error $e(n)$ can be expressed as:

$$e(n) = y(n) + y(n)^* \quad (2.9)$$

where $y(n)$ is the speech signal under analysis and $y(n)^*$ is the resynthesized signal using the LPC filter. $y(n)^*$ is defined as

$$y(n)^* = - \sum_{k=1}^p (a_k y(n-k)) \quad (2.10)$$

This LPC approximation model is shown in figure 2-4: The residual signal whose energy is to be minimized is obtained by passing the signal $y(n)$ through the all-pole filter $H(z)$. The residual energy is given by:

$$E = \sum_{n=-\infty}^{\infty} [e(n)^2] \quad (2.11)$$

In other words E , the mean square error, is the total squared error in predicting $y(n)$ from its previous values using the calculated LPC coefficients. The coefficients a_k that minimize E are obtained by setting the partial derivative of E with respect to each of the a_k coefficients to zero.

$$\left\{ \frac{\partial E}{\partial a_k} = 0 \text{ such that } 0 \leq k \leq p \right. \quad (2.12)$$

Figure 2-4: LPC Basic Speech Synthesis Model

There are many methods for solving this system of p equations with p unknowns, autocorrelation and covariance are two such methods. Linear programming techniques have been suggested and are used. The Levinson-Durbin recursive method is one of them [RaSc 78].

LPC pitch extraction may be performed using a variety of techniques; the inverse filter approach, the partial correlation method (PARCOR) [RaGo 75], the simplified inverse filter method (SIFT) [Mark 72], autocorrelation, G-L and the harmonic sieve are some of the algorithms using LPC analysis as a tool for pitch extraction [WoFa 85].

In the inverse filter approach, speech is analyzed to establish the coefficients of the

all-pole filter $H(z)$. The speech is then "played back" through the inverse filter $1/H(z)$ to establish the input $u(t)$ using the equation [WoFa 85]:

$$U(z) = \frac{S(z)}{H(z)} \quad (2.13)$$

where $H(z)$ is the all-pole filter transfer function and $S(z)$ is the speech waveform. As a result one obtains the pitch pulses as given by $U(z)$.

A similar procedure to the inverse filter method is the partial correlation approach (PARCOR) [WoFa 85]. In the PARCOR method, instead of establishing the filter coefficients and inverse filtering the speech signal. The filter input is calculated in one operation by establishing the error at every iteration, e_p , using partial correlation, where p is the order of the LPC filter [Makh 84]. Then, as in the inverse filter method, the signal is low-pass filtered and a threshold is applied to obtain the pitch.

The computational complexity of the inverse filter and partial correlation methods presented put them at a disadvantage for real time applications. A large amount of processing is required to separate the excitation spectrum from the vocal tract inverse spectrum. The SIFT (simplified inverse filter) algorithm, based on the same principle as the other two presented methods, reduces the amount of processing by prefiltering the speech signal with a 1kHz cut-off low pass filter. The signal is then down-sampled and an inverse filter operation is performed on the signal [O'Sh 88]. Prefiltering the signal suppresses the effects of frequencies $\geq 1000Hz$; so, performing an inverse filtering operation with a fourth order filter ($p=4$) is sufficient. The SIFT algorithm then extracts the fundamental frequency F_0 from the existing waveform which also contains F_1 and F_2 . Note that formant frequencies greater than F_2 are not considered due to the prefiltering operation that was initially applied to the signal.

LPC also has formant tracking capabilities, the prediction coefficients provide frequency information of the speech segment under analysis. Furthermore, the formant center frequency and bandwidth can be determined accurately by factoring the prediction polynomial, the denominator of equation 2.8 [O'Sh 88]. A prediction analysis of order p , yields a minimum number of $\frac{p}{2}$ real formants, in the case where the roots of the prediction polynomial of equation 2.8 are complex conjugate roots. A disadvantage of this method is the difficulty in establishing the exact relationship between the roots and the expected resonances of the vocal tract [Pars 86]. In some cases LPC analysis, employing a prediction

filter of reasonable prediction order $p = 14$ fails to model the vocal tract resulting in an inaccurate representation of the vocal tract characteristics.

Linear predictive coding has been successfully implemented in real-time for use in the cochlear implant by Barszczewski and Morris [BaMo 89].

Figure 2-5: Block Diagram of SIFT[RaSc 78]

2.2.5 The Cepstrum

The cepstrum $c(n)$ of a speech sample $x(n)$ is by definition the inverse Fourier transform of the logarithm of the magnitude of the Fourier transform of the speech sample $x(n)$ [ScRa 70]. The cepstrum can be represented by the following equation:

$$c(n) = IDFT(\log | DFT(x(n)) |) \quad (2.14)$$

Cepstral analysis is based on the cepstral speech production model which considers speech as the time convolution of the vocal tract characteristics and a gain function with pitch pulse excitations. The vocal tract characteristics and the gain function may be modeled by a filter similar to the LPC filter model.

This method was first presented as a pitch extraction algorithm by Noll [NoSc 64], and developed into a speech analysis system by Schafer and Rabiner [RaSc 78, ScRa 70]. The cepstral speech analysis system, when used to extract pitch periods and track speech formants, separates the pitch period pulses of speech from the vocal tract characteristic function in speech. Cepstral analysis consists of three operations, as shown in figure 2-6. These operations are the following:

- 1- The speech sample is transformed into the frequency domain to render the convolution of the pitch pulses and vocal tract transfer function into a multiplication.
- 2- A logarithm function is applied to the input of the transformed sample, this changes the multiplication of the pitch period and the vocal tract characteristic function into an addition.
- 3- Applying an inverse FFT yields a scaled addition of the two waveforms, the pitch period and the vocal tract characteristic function, in the logarithm scaled time domain (quefrency domain). The pitch period waveform manifests itself as a single peak in the cepstrum [Rile 89].

Filtering the pitch from the cepstrum, leaves the time domain representation of the vocal tract characteristic function. An FFT would then produce the frequency response of the vocal tract. Cepstral analysis is considered by many experts as a bench mark for speech pitch calculations [WoFa 85].

Figure 2-6: Cepstral Analysis for Pitch Detection[ScRa 70]

2.3 The Current Spread Problem

The design of a speech processing strategy for the cochlear implant involves a clear understanding of the latest developments in cochlear implant research, in terms of the latest problems and/or solutions. This section discusses the electrical interactions between the channels of the stimulating electrode. The next section will present the various cochlear implant research groups and their findings.

The existence of conductive fluids in the cochlea, causing current spread as presented in Chapter One, has many implications in terms of signal processing for the cochlear implant. This is due to the fact that the output of the signal processing strategy is the stimulation input pattern to the electrode array. Therefore, it is of importance to discuss the problem of current spread and some of the proposed solutions. The cochlea, being of tonotopical nature (frequency mapped), causes different electrodes placed in different areas of the cochlea to give rise to different auditory sensations. The apical electrodes, electrodes placed in the area of the cochlea farther from the round window, upon stimulation correlate with low-frequencies. While the basal electrodes upon stimulation correlate with high frequencies. Upon electrical stimulation of the cochlea, current interactions among the various channels in the electrode array occur [WhCo 85]. If the current spread is large enough, all the channels may seem to produce the same auditory sensation to the patient upon stimulation.

The solution to the problem of current spread is in compensating for the non linearities in the damaged cochlea that result from electrical stimulation. Two routes have been suggested for solving this problem: improved electrode design to take care of excess current, and/or predicting the resulting current spread and compensating for current spread when stimulating the cochlea.

In all electrode designs, regardless of the manufacturing technology, a significant level of current spread will occur upon stimulation in the cochlea [LuSi 84]. It is important to note that bi-polar electrodes show a significant reduction in current spread over mono-polar electrodes [Town 87].

In conjunction with improved electrode design an attempt to compensate for the current spread produced in the cochlea is considered as a starting point for current spread correction. This approach involves the determination of a mapping function between the

channels of the electrode and the cochlea. There are virtually thousands of neurons in the cochlea so that creating a mapping matrix, between the neurons and electrode channels, would prove an impossible task [Town 87]. However grouping together sites of similar perceptual characteristics would render the problem of finding a spread matrix feasible. The spread matrix for each patient will be different from every other patient. Once a mapping function is obtained, the required patterns to produce localized maxima in the cochlea can be determined using matrix manipulation techniques (eg. matrix reduction).

The current spread problem arises from the electrical stimulation of inter-conductive sites in the cochlea (current spreads from the sites of stimulation). It results in the auditory perception of a stimulation pattern other than the intended [Rabi 89]. There are some hypotheses regarding the correction of current spread. Two such hypotheses are the threshold theory and the resistor model theory [Town 87, GeCh 87]. The threshold theory is based on the fact that hearing occurs when the stimulating current at a certain site exceeds the electrical threshold at that site causing an auditory neuron to fire. The current spread problem may be solved by defining the interactive relationship between the thresholds and calculating a compensating current pattern to cancel the effects of the interactions. Some research groups (such as a group from the department of Electrical Engineering at the University of Wisconsin at Madison [Geis 89]) are attempting to model the damaged cochlea in an effort to compensate for current spread.

The resistor model theory calculates the interaction between the channels of the electrode in the cochlea by modeling the current interactions as resistors and compensating for their activity. The resistor model is shown in figure 2-7.

The threshold theory will be discussed in the next sections, it has been implemented and tested with varying degrees of success [Town 87].

2.3.1 Theory

The theoretical basis of the threshold current spread correction scheme was developed by B. Townshend [Town 87]. The theory assumes the existence of a spread matrix S such that the following equation is true:

$$j^\infty = S^\infty i \quad (2.15)$$

where j^∞ is the current density vector in all directions, S^∞ is the current spread matrix

Figure 2-7: The Resistor Model of The Cochlea[GeCh 87]

in all directions and i is the current vector pattern input to the electrode. The theory is based on three assumptions:

- 1- A patient experiences an auditory sensation if at least one auditory neuron is firing, i.e. a site is stimulated at a level greater than its threshold.
- 2- Stimulating currents add linearly in the cochlea.
- 3- i is a unique vector and the direct cause of j^∞ , which is also unique for a given point in the cochlea.

With j_p^∞ being a unique current density vector associated with point p , a unique threshold vector t_p^∞ is defined for point p . Due to the initial physical damage of the cochlea or damage during electrode placement in the cochlea, $t^\infty = \infty$ for most points or sites in the cochlea. It is concluded therefore that only a finite number of excitable sites are present in an implanted cochlea. Therefore, the assumption of a unique S^∞ is valid.

2.3.2 Algorithm: Formulation and Implementation [Town 87]

S^∞ being unique the following may be inferred:

$$j_p^t = S_p^t i \quad (2.16)$$

where j^t is the normalized current density vector along each fiber by its threshold ($j_p^t = j_p^\infty / t_p^\infty$), i is an (N_e) by 1 current pattern vector (where N_e is the number of channels in the electrode) and S^t is the normalized current spread matrix of size 1 by (N_e) .

Let N_s be the number of excitable sites in the cochlea.

Testing a number of current vector patterns, approximately 1000, an $N_s = 1000$ is obtained. Expanding equation 2.16 into matrix form with n_s as a variable the following matrix equation is obtained:

$$\begin{pmatrix} j_1 \\ j_2 \\ \vdots \\ j_{N_s} \end{pmatrix} = \begin{pmatrix} S_{1,1} & S_{2,1} & \dots & S_{N_e,1} \\ S_{1,2} & S_{2,2} & \dots & S_{N_e,2} \\ \vdots & \vdots & \ddots & \vdots \\ S_{1,N_s} & S_{2,N_s} & \dots & S_{N_e,N_s} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \\ \vdots \\ i_{N_e} \end{pmatrix} \quad (2.17)$$

In its present form, the S^t matrix is of little use having a size of an N_s by N_e matrix. Applying matrix reduction techniques with a chosen error function, sites exhibiting the same stimulatory behavior are grouped together. It has been shown that for $N_s > 6$ the value of the error remains almost constant, this is shown in figure 2-8. The number of sites is typically set equal to the number of channels in the electrode to simplify the calculations. The reduced S matrix is solved to obtain the required current pattern vector to stimulate a specific site. Applying this technique, single site perception may be achieved using a correction matrix of size (N_s) by (N_e) .

The implementation of the above mentioned technique involves four basic operations as shown in figure 2-9:

- 1- Patient testing, using a multitude of current pattern vectors. The selection of an error function and applying matrix reduction techniques to determine the current spread matrix.
- 2- Once the spread matrix is determined, it is deconvoluted to produce a stimulation matrix. Deconvolution of the spread matrix requires the calculation of the i current pattern vectors that stimulate each site independent of the others.

Figure 2-8: Error vs n , [Town 87]

- 3- The stimulation matrix is coded as a routine in the speech processing system and the current pattern needed to stimulate the required channels is calculated.
- 4- The resultant current pattern is fed to the stimulator in the form of energy per channel and in turn sent to the electrode.

2.4 Signal Processing in Cochlear Implant Research Projects

2.4.1 A General View

Attempting to predict the behavior of a damaged cochlea has led to the adoption of a number of hypotheses by cochlear implant research groups. This section presents an overview of the work of these groups, including their speech processing strategies and results.

2.4.2 London

The London implant group uses only extracochlear electrodes in their cochlear implants. Extra-cochlear electrodes minimize the complexity of the surgery involved and the possible physical damage to the cochlea. Most of the group's research involved studying the effects

Figure 2-9: Threshold Theory Current Correcting Method

of cochlear implants on the lip reading abilities of the patients [King 87]. The London system was developed for research purposes only.

2.4.3 Los Angeles

The Los Angeles group is the most prolific of groups in number of implantees (approximately 1500). The Los Angeles group varied the specifications of their system widely to accommodate patient requirements, this practice prevented objective multi-patient testing. Using the minimum auditory capabilities battery, MAC battery, it was observed that the perception of vowels in general correlates with the first formant F1 and the duration of the vowel. The MAC battery contains a set of words, sounds, expressions and sentences that are used as speech samples to test the degree of auditory perception [OKRS 85].

2.4.4 Vienna

The Vienna implant system uses a single channel electrode (extra and intra cochlear electrodes), and has employed a large number of distinct speech processing strategies. The Vienna group attempts to vary the methods of mapping of speech to the electrode in the cochlea in terms of band-limiting and wave-shaping of speech [MiTC 84].

2.4.5 San Francisco

The San Francisco group was among the first to extract feature information from speech and map it to the channels of the electrode. Their implant uses a four channel electrode, the electrode channels are positioned to take advantage of the distinct frequency distribution of the cochlea. The electrode channels are used in the following stimulating strategy, a single channel handles the first formant, two channels are used to cover the region of the second formant, and the fourth handles the mapping of higher frequencies. Exhaustive testing by Owens and others using the MAC battery resulted in very encouraging results [MiTC 84]. This implant was also developed for research purposes. The 3M cochlear implant system was initially based on the San Francisco cochlear implant research group's design.

2.4.6 Paris

The Paris group employs a speech processor design which is based on a quantized frequency map of the electrically stimulated cochlea, correlated with the tonotopical nature of the cochlea. This frequency map is derived from unilaterally deaf implantees by means of audiometric comparison with the good ear. The speech signal is divided into frequency bands, the width of the band is varied from channel to channel in order to equalize the energy at each band when stimulating different channels of the electrode [MiTC 84].

2.4.7 Stanford

The cochlear implant research group at Stanford was formed in the mid 1960s. By 1968 several tests had already been performed and speech perception mapping methods had been suggested [Math 78]. The Stanford group tested many forms of the implant, the group used single and multi-channel electrodes. They conducted speech perception studies classifying speech as phonemes and determining the voicing of the speech. The Stanford group was the first to suggest the extraction of the first two formants and use their frequency

placement in the cochlea as an indication of the channel to be stimulated. They contributed to the cochlear implant system research by analytically formulating a correcting strategy for the cochlea current spread problem [Town 87]. This implant was developed for research purposes.

2.4.8 Melbourne

The Melbourne group, initially a research effort at the University of Melbourne in Australia, was formed in 1969. The system they designed and implemented was the first cochlear implant to be manufactured and put on the market. The marketing of the implant in North America is done through the American firm, Cochlear Incorporated. The first generation of the cochlear implant was implemented using a 10-channel electrode. By 1978 the group was using 22-channel electrodes, which is the electrode currently in use by Cochlear Incorporated. The speech processing strategy employed consists of filtering speech through a set of bandpass filters to determine the peak power frequency locations. The first and second formant frequencies are extracted, and their corresponding electrode channel locations in the cochlea are stimulated [Coch 85]. The Nucleus implant system is considered by many specialists to give the best overall patient speech perception results [Schw 87, Boot 89, Dorm 89].

2.4.9 Boston

The Boston group, a research group from the Massachusetts Institute of Technology, designed and implemented the Symbion Implant. The system uses four bandpass filters and a speech compression algorithm. The output is fed to a 4-channel electrode. The Symbion implant has produced encouraging results and was made available for public use in 1990 [Eddi 89]. Some work on creating an electronic model of the auditory process is currently underway at M.I.T. [Sear 89]. Current spread corrective algorithms are also being investigated [ASHA 86].

2.4.10 Tokyo

The research conducted at Hokkaido University, Japan, was the first to incorporate a hardware model of the cochlea. The speech processor design takes into consideration lateral inhibition, in an attempt to compensate for the current spread in the cochlea. The modeling of current spread in the cochlea is encoded in the processor design by evaluating

optimal current stimulation patterns [IfWh 86]. No implants using this system have been produced. The Hokkaido speech processor is shown in figure 2-10.

Figure 2-10: Hokkaido's speech processor, an eight channel implant[IfWh 86]

2.4.11 Nimes

Genin and Charachon suggested a novel speech processing strategy that uses a resistor network to model the non-linearities resulting from electro-stimulation of the cochlea. This speech processing strategy was presented at ICASSP 86 (held in Tokyo 1986) [GeCh 86]. The speech processing algorithm used involves the implementation of a 12-channel filter bank with logarithmic compression and envelope analysis. The analog circuitry for a single

Figure 2-11: Genin-Charachon: Analog circuitry for 1 channel [GeCh 86]

channel of the Nimes Implant is shown in figure 2-11.

2.4.12 Ottawa-Sherbrooke

The Canadian cochlear implant project is a combined effort of researchers from two universities; Carleton University in Ottawa (the Department of Electronics and the Department of Systems and Computer Engineering) and Sherbrooke University (the Department of Electrical Engineering). This group is the first to propose a wearable microprocessor-based hardware. Their speech processing strategy uses LPC analysis for formant tracking and pitch extraction. One of the group's major contributions is the research conducted on thin film photolithography automated electrode production [PuMD 87]. A Carleton rolled electrode is shown in figure 2-12. To date no humans have been implanted using this system.

2.4.13 Summary

The previous section was an overview of the available data on the work done and results obtained by various cochlear implant research groups. It is seen that no dominant speech

Figure 2-12: The rolling process of the electrode made at Carleton

processing strategy for the cochlear implant is evident and that every group has its own strategy with acceptable recorded and documented results. One important factor to note is the attempt by many groups to compensate for the non-linear behavior of the cochlea in their speech processing strategy. Table 2-1 summarizes the data presented, in section 2.4, on the various cochlear implant research groups in terms of electrode design, speech processing strategy and cochlear implant product if available.

2.5 Suggested Signal Processing Strategy

In the following discussions only post-lingually deaf patients are considered as cochlear implant recipients. Post-lingually deaf people are people who were capable of intelligible conversation before losing their hearing abilities. These people constitute by far the majority of cochlear implant patients [Dorm 89]. There is general agreement among cochlear implant experts that patients stimulated with frequency representations of speech tend to perceive speech with a higher success rate than patients using other speech features [TGKT 87, BLMC 85, WDDE 87]. The problem of selecting a speech processing strategy for the

Research Group	Electrode	Signal Proc.	Implant Product
London	Single Channel (S.C.), Extr-Cochlear (EC)	FilterBank (F.B.)	not available(n.a.)
Los Angeles	S.C., Inter-Cochlear (IC)	Energy Detection (E.D.)	House
Vienna	S.C., EC & IC	E. D. & F.B.	Hochmair
San Francisco	Mutli Channel (M.C.)	Formants (Fs)	3M
Paris	M.C. & S.C., IC	F.B.	Bertin
Stanford	M.C., IC	F.B. & E.D. & Fs	n.a.
Melbourne	M.C., IC	F.B. & Fs	Nucleus
Boston	M.C., IC	F.B.	Symbion
Nimes	M.C., IC	F.B.	Mono/MultiSonic
Duren	M.C., IC & EC	F.B. & E.D.	Universal
n.a.	M.C., IC	F.B. & E.D. & Fs	Clarion

Table 2-1: The Various Cochlear Implant Groups

cochlear implant is focused on the choice of speech features to be extracted. In this research the extraction of the pitch period and the first three formant frequencies of speech is proposed. The pitch provides timing information, which is an important factor in speech perception for cochlear implant patients [Pfin 87, Pauk 87]. The pitch can also be used as the rate of stimulation in a pulsatile stimulating strategy [Town 87]. The formant frequencies are important for speech recognition because they contain abstracted spectral speech information [Boot 89].

The choice of a signal processing method for pitch extraction and formant tracking depends on the constraints imposed by hardware factors and speech processing considerations for the cochlear implant. The most important constraint is the execution time of the signal processing algorithm. The processing time of the signal should be less than 40ms as not to inject any noticeable delay in the stimulation process of the patient (this time constraint is related to the syllabic rate of speech). Another important consideration is the physical size of the system. It should be wearable and have a low power consumption.

Referring to the various speech processing methods presented in section 2.2, cepstral analysis and LPC are the best candidates for formant tracking and pitch extraction. LPC has been implemented on a wearable hardware with a 20ms execution time [HPDM 87, BaMo 89]. Cepstral analysis was the speech processing strategy chosen for implementation for the cochlear implant. Cepstral analysis is considered a bench mark for pitch extraction, and is a proven formant tracking speech processing method.

Chapter 3

Cepstral Analysis: Software-Hardware

Bogert was the first to observe banding (clustering) in the spectrograms of seismic signals [BoHT 63]. He realized that this banding was caused by periodic ripples in the spectra. This periodic ripple effect occurs in the spectra of any signal consisting of itself plus an echo[BoHT 63]. Tukey also suggested that this banding might be achieved by first taking the logarithm of the spectrum, a spectral analysis of the log spectrum itself could then be used to determine the frequency of the ripple. These two observations by Bogert and Tukey led to the definition of the cepstrum as the inverse Fourier transform of the logarithm of the Fourier transform of a speech signal.

3.1 Theory of Cepstral Analysis

3.1.1 Signal Cepstral Analysis

Consider a signal represented by a finite sequence $x(n)$ of length N , such that:

$$x(n) = p(n) * v(n) \quad (3.1)$$

where $*$ denotes the convolution operation. Cepstral analysis, as presented in Chapter Two, provides an algorithm capable of separating $p(n)$ from $v(n)$ from $x(n)$. Cepstral analysis involves three basic operations as shown in the following equations:

$$X(e^{j\omega}) = \sum_{n=0}^{N-1} x(n)e^{-j\omega n} \quad (3.2)$$

$$X'(e^{j\omega}) = \text{LOG}[X(e^{j\omega})] \quad (3.3)$$

$$x'(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X'(e^{j\omega}) e^{j\omega n} d\omega, 0 \leq n \leq N-1 \quad (3.4)$$

Equation 3.3 involves the calculation of the logarithm of a complex number, which is not a uniquely defined operation. For cepstral analysis purposes it is sufficient to define the complex logarithm operation as:

$$\text{LOG}[X(e^{j\omega})] = \text{LOG}|X(e^{j\omega})| + j\text{ARG}[X(e^{j\omega})] \quad (3.5)$$

The calculation of the real part of the complex logarithm poses no problems, however problems of uniqueness arise in evaluating the imaginary part of equation 3.5 [OpSc 75]. The imaginary part is simply the phase angle of the z transform evaluated on the unit circle. To ensure the uniqueness of the phase angle, the phase angle should be a continuous odd function of ω [RaSc 78].

The continuous integral in equation 3.4 makes the calculation of $x'(n)$ a complex one, especially for computerized applications. Using the discrete form of the Fourier transform, the complex cepstrum may be evaluated using the following equations (3.6,3.7,3.8):

$$X_d(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn} \quad (3.6)$$

$$X'_d(k) = \text{LOG}[X_d(k)] \quad (3.7)$$

$$x'_d(n) = \frac{1}{N} \sum_{k=0}^{N-1} X'_d(k)e^{-j\frac{2\pi}{N}kn} \quad (3.8)$$

the subscript d denotes the use of the discrete form of the Fourier transform. Equations 3.6-3.8 are the sampled versions of equations 3.3-3.5. Therefore, $x'_d(n)$ is in general an aliased version of $x'(n)$. This creates a problem for the uniqueness of the phase of the complex logarithm. As mentioned earlier, the phase angle should be a continuous function of ω for the uniqueness of the phase angle. It is therefore customary to oversample the logarithm function by appending zeros to the original signal $x(n)$ [RaGo 75].

In cases where phase information is not required, or where the application of cepstral analysis does not involve signal reconstruction, the real cepstrum $c(n)$ is used instead of the complex cepstrum. The real cepstrum employs the logarithm of the magnitude of the Fourier transform. The term cepstrum, as of now, unless otherwise specified will represent the real cepstrum throughout the remainder of this thesis. The real cepstrum is calculated using the following equation:

$$c_p(n) = \frac{1}{N} \sum_{k=0}^{N-1} \text{LOG}|X_p(k)|e^{j\frac{2\pi}{N}kn} \quad (3.9)$$

3.1.2 Speech Cepstral Analysis

As seen in the speech production model in figure 2-2, voiced speech is the product of the filter transfer function $H(z)$ and the impulse frequency response $P(z)$. The filter transfer

function $H(z)$ consists of two parts, the vocal tract transfer function and the glottal transfer function. For the proposed application the glottal effects are neglected. Therefore $H(z)$ may be approximated by $V(z)$ the vocal tract transfer function [Rile 89]. Assuming the transfer function $H(z)$ is constant over the processing interval, it may be defined using a set of coefficients of order p directly related to $V(z)$ such that [OpSc 75]:

$$V(z) = A \prod_{k=1}^p (1 - c_k z^{-1})(1 - c_k^* z^{-1}) \quad \text{where } |c_k| < 1 \quad (3.10)$$

Referring to the speech production model in figure 2-2, voiced speech $s(n)$ can be expressed as:

$$s(n) = p(n) * v(n) \quad \text{where } 0 \leq n \leq N - 1 \quad (3.11)$$

The effect of $p(n)$ in the cepstrum manifests itself as a peak at n_0 samples away from the origin in the cepstrum, where the n_0 th sample's period corresponds to the fundamental frequency of the processed speech sample [OpSc 75]. $V'(z)$, the complex cepstrum of $v(n)$, is calculated using the following equation:

$$V'(z) = \log[A] - \sum_{k=1}^p (\log[1 - c_k z^{-1}] + \log[1 - c_k^* z^{-1}]) \quad (3.12)$$

Using the reduction properties of $\ln(1+x)$ and $\ln(1-(-x))$, equation 3.12 can be reduced to

$$v'(n) = \begin{cases} 0 & n < 0 \\ \log[A] & n = 0 \\ \frac{1}{n} \sum_{k=1}^p [(c_k)^n + (c_k^*)^n] & n > 0 \end{cases} \quad (3.13)$$

using the polar notations for c_k , this results in an exponentially decaying $v'(n)$ as expressed by equation 3.14 (this result also applies for the real cepstrum),

$$v'(n) = \sum_{k=1}^p \frac{|c_k|^n}{n} 2 \cos \phi_k n \quad \text{for } n > 0 \quad (3.14)$$

The exponentially decaying term $\frac{|c_k|^n}{n}$ can be related to the bandwidth of the k^{th} formant peak [Noll 67]. If the bandwidths of the formant peaks are larger than the fundamental frequency of speech, $v'(n)$ would decay to small values at the cepstrum point representing the pitch period. Under this condition the cepstrum can be treated as a uniformly sampled version of the vocal tract transfer function [Brac 78]. If the bandwidths of the formant peaks are smaller than the value of the fundamental frequency (the pitch) aliasing will not occur in the cepstrum.

Following the cepstrum calculations, the frequency domain vocal tract transfer characteristics are obtained from the cepstrum by applying a Fourier transform to a filtered version of it. The filtering of the cepstrum, known as liftering, is achieved by zeroing all cepstral points representing frequencies less than a chosen cutoff frequency. The cutoff point (corresponding to the cutoff frequency) is chosen to eliminate the peak corresponding to the pitch frequency (usually $\leq 250Hz$) [OpSc 75].

The implementation of cepstral analysis as a speech processing method for pitch and formant extraction faces two problems, presented earlier in this section. These problems are the filtering operation of the cepstrum to obtain the vocal tract transfer function, and the aliasing in the cepstrum. The filtering operation is discussed in detail in section 3.2.5. The aliasing problem, that occurs due to the rate of decay of $v'(n)$ in the cepstrum and the uniqueness of the phase information when calculating the logarithmic function in cepstral analysis, will be discussed next.

The aliasing effects due to the decay of $v'(n)$ in the cepstrum are related to the bandwidths of the formant peaks. Thus they are unavoidable when they occur. The aliasing occurring due to the problem of uniqueness of the phase angle (explained in section 3.1.1) becomes unimportant when feature extraction (the pitch and formants) is the cepstral analysis application. Phase information is not required when implementing cepstral analysis for extracting either the pitch or the formants, because the real cepstrum can be used. Stated differently, for the proposed application, phase information and any related implementation problems may be ignored.

For cepstral applications where phase information is required, this aliasing may be minimized by oversampling the speech signal when transforming it from the time to the frequency domain. This is accomplished by appending zeros to the original sampled speech signal and using a large DFT. Traditionally large DFTs are used, the original Rabiner and Schafer implementation uses a 1024 point DFT for a 400 point speech frame. With the use of FFTs, cepstral analysis has been implemented successfully using 512 point FFTs for a 256 point speech frame [TrQu 79].

3.2 Software

The implemented cepstral system in this thesis is similar to the system proposed by

Figure 3-1: The Cepstrum of a Speech Frame

Schafer and Rabiner [ScRa 70]. The difference in the implementations is the author's use of a 256 point FFT implemented using fixed point arithmetic for a real time application. The Schafer and Rabiner implementation is a non-real time floating point implementation using a 1024 point DFT.

In the next section the algorithms used in the implementation of the cepstral analysis system, will be presented and discussed. These algorithms are: the FFT, the logarithm, the magnitude, the windowing routine, the feature extraction routine and the interrupt service routine. The program development was done using the Software Development System developed by Texas Instruments for the TMS320C25 microprocessor [TI 1986].

Figure 3-2: Cepstral Analysis, A Block Diagram

An important factor to note when considering programs coded in the assembler language for real-time applications is the use of internal (on-chip) memory. Executing a program which executes from code and data located in internal memory is much faster (approx. 1.20 times) than a program which executes from code and data located in external memory. Minimizing the use of external program and data memory was adopted as one of the programming policies for the implementation. A factor to consider is the scaling of intermediate values. This is an important and necessary operation due to the fixed point implementation of the algorithm. Another consideration for a real-time implementation, is to optimize the interrupt service routine (ISR) which is used for data acquisition. The ISR

is invoked 256 times during each execution frame of the cepstral analysis system.

3.2.1 The Fast Fourier Transform

The FFT is used three times in the cepstral system. It is therefore necessary that the FFT coding and optimization be given special consideration. There are many ways to implement an FFT. Radix 2 (R-2) FFTs were considered for implementation since they do not require special memory considerations in terms of intermediate storage and on chip memory. The following FFT algorithms were considered and evaluated:

- A R-2 straight line code FFT, this FFT required a large amount of program memory approximately 40kbytes.
- A looped code R-2 FFT, this FFT required less than 1kbyte of program memory.
- A R-2 FFT utilizing a macro library, this program required 13kbytes of program memory.

The implemented FFT, in the cepstral system, is a Cooley-Tukey R-2 complex FFT. The FFT is performed on a 256 point speech sample, with an execution time of 4.5 ms using the TMS320C25 digital signal processor. The FFT uses internal data locations for temporary storage and execution, and a program code size of 798 words. In the development system used, a word of data storage or a word of code storage is a 16-bit memory location.

3.2.2 The Magnitude Routine

The magnitude routine calculates the sum of the squares of the real and imaginary parts of the transformed speech sample. Let $X_k(\omega)$ and $Y_k(\omega)$ be the real and imaginary parts of the Fourier transform of the k th speech sample $x_k(n)$, and m a scaling factor. The magnitude routine scaling function as implemented in the cepstral system is

$$|FFT(x(n))| = \frac{(X(\omega)^2 + Y(\omega)^2)}{m} \quad (3.20)$$

Each 16-bit number corresponding to X and Y is squared and stored in a 32-bit accumulator. Storing the sum of squares (a 32-bit number) in a 16-bit memory location required scaling in order to retain a significant representation of the value. Also, it was necessary to avoid overflow in the addition of the numbers. The scaling adopted was optimized for a variety of speech samples. This routine does not require any temporary storage, it uses 15 words of program memory and has an execution time of 0.24ms for every analysis frame.

3.2.3 The Logarithm Routine

Most software implementations of the logarithmic function use expansion series, such as the Taylor expansion series. For microcomputer logarithm calculations this is achieved by representing a number I in the following form $I = (mantissa)(base^{exponent})$ taking advantage of the fact that $\log(a^n) = n\log(a)$.

$$LOG(I) = (exponent)LOG_{base}(mantissa) \quad (3.21)$$

Most logarithm routines use the method outlined above with a $base = 2$, since a divide by 2 operation is just a binary shift on the processor level. The processor calculation of the exponent and mantissa of a certain number in base two is not a difficult operation to program or perform [Hart 68]. An example of an implementation of the logarithm function using a TMS320 processor is found in a paper by Engebretson and O'Connell [EnOC 86]. A second order expansion series was implemented and used as a starting point. The second degree polynomial approximation did not provide the logarithm function with sufficient accuracy for the proposed application. The implemented logarithm routine uses the Taylor expansion series, the routine is an adaptation of a program written by Dr. R. Goubbran [Goub 88]. The program transforms a given number into its mantissa and exponent using binary arithmetic. The calculation of the logarithm base two is shown in equation 3.22.

$$LOG(x) = exponent + mantissa + \sum_{n=3}^{n=8} (A_n(mantissa)^{n-2}) \quad (3.22)$$

where $x = mantissa + 2^{exponent}$, A_n is a constant and $1/2 \leq mantissa \leq 2$. The actual implementation calculates the logarithm of $4096I$, this is done for scaling purposes. The logarithm routine is applied to 256 points in internal data memory, 23 temporary internal data memory locations and 530 bytes of program memory. The execution time for this routine is 4.9 ms using the TMS320C25 processor running at 40MHz.

An alternative method, is to use a piecewise linear approximation of the logarithm. Dry coding for a piecewise linear logarithm approximation routine resulted in a 1.2ms execution time (also using the TMS320C25 @40MHz) and the requirement of 78 program memory locations and 10 temporary data memory locations.

3.2.4 The Feature Extraction Routines

3.2.4.1 The Pitch Extraction Routine

The pitch manifests itself as the maximum peak in the speech cepstrum [OpSc 75]. Therefore, pitch extraction involves locating the maximum peak which determines the pitch period of the speech sample. The pitch extraction routine is a comparison program executed over the cepstral points corresponding to the human voice pitch frequency. The range for human voice pitch frequency is between 50Hz and 400Hz. The pitch extraction routine requires one temporary word of internal data memory and 67 words of program memory. The execution time of the routine using a TMS320C25 processor running at 40MHz is 0.77 ms.

3.2.4.2 The Formant Tracking Routine

The formants of speech are the characteristic frequencies of speech for a certain speech interval. The presented cepstral analysis implementation locates and determines the first three formant frequencies and their respective amplitudes, which appear as peaks in the processed speech sample. The first three formant frequency ranges as defined by Schafer and Rabiner are [ScRa 70]:

$$\begin{cases} 200Hz \leq F_1 \leq 900Hz \\ 550Hz \leq F_2 \leq 2700Hz \\ 1100Hz \leq F_3 \leq 2950Hz \end{cases} \quad (3.23)$$

Note that the definition of the formant ranges in equation 3.23 opens the possibility of having more than one distinct formant in a given range due to the overlapping of ranges. The formant tracking routine requires six temporary words of internal data memory and 123 words of program memory. The average execution time of the routine using a TMS320C25 processor running at 40MHz is 0.83 ms.

3.2.5 The Windowing (Liftering) Routine

Removing the pitch period peak from the cepstrum is called cepstral filtering, also known as liftering. It is equivalent to applying a window such that:

$$w(n) = \begin{cases} w(n) & 0 \leq n \leq N \\ 0 & n > |N| \end{cases} \quad (3.24)$$

The purpose of cepstral filtering is to remove the pitch peak and minimize the effects of low frequency components ($\leq 400Hz$) from the cepstrum. The cepstral window attempts to

emphasize the points corresponding to the frequencies of interest. A symmetrical window is a window whose coefficient values are evaluated over half a period of the window generating function. In the case of a raised cosine window, the positive half period of a cosine function evaluated between $-\pi/2$ and $+\pi/2$ is applied to the data points under consideration. A single sided window is a window whose coefficients are calculated over a quarter of the period of the generating function. Usually symmetrical windows yield smoother spectra than single sided windows.

Four windows were chosen and tested:

- A rectangular window that zeros the cepstral peak. This window resulted in poor formant tracking.
- A single sided raised-cosine window, with a cutoff frequency of $250Hz$. This window resulted in smooth formant peaks, but poor formant tracking was observed.
- A symmetrical raised-cosine window was implemented with a cutoff frequency of also $250Hz$. The results indicated a strong dependence on the input speech patterns.
- A symmetrical Blackman-Harris window rendered the best results for the proposed application. The window weightings are of the form:

$$w(n) = \frac{1}{L} \sum_{k=0}^K a_k \cos\left(\frac{2\pi kn}{L}\right) \quad \text{where } n \leq L/2 \quad (3.25)$$

A third order Blackman-Harris window with the following specifications is implemented, $k = 3$, $a_0 = 0.3635819$, $a_1 = 0.4891775$, $a_2 = 0.1365995$ and $a_3 = 0.0106411$ [Nutt 81].

The choice of L affects the frequency content of the filtered cepstrum, as explained earlier in definition of the frequency of a point in the cepstrum. A normal practice when extracting the vocal tract transfer function is to adopt a cutoff of $250Hz$ for the cepstral filter, which is twice the average human voice pitch frequency. A lower cutoff frequency of $250Hz$, corresponds to a $4ms$ period. Therefore, the first 40 points in the cepstrum have to be windowed yielding an $L = 40$. It is also to be noted that a symmetrical window, such as the selected Blackman-Harris window, results in emphasizing chosen frequency peaks in the vocal tract transfer function (in this case the human formant frequency range). The program uses 20 temporary data memory locations, 74 words of program memory and has an execution time of 0.05 ms (implemented using the TMS320C25 @40MHz).

3.2.6 Interrupt Service Routine

Sampling at a frequency of 8kHz, gives a data acquisition rate of one sample every $125\mu\text{sec}$. Therefore the system as proposed for a real-time application, samples the speech data 256 times every 32ms processing interval. The interrupt service routine should not only have a short execution time but also employ a minimal number temporary memory storage locations to avoid lengthy stacks. Traditionally, interrupt service routines involve storing the values of the accumulator, auxiliary registers, program counter and stack pointers into a designated memory location prior to the data acquisition operation. This is done in order to execute the interrupt service routine without affecting the status of the main program. The stored values are restored and the main program execution is resumed after the is serviced. Interrupt service routines, similar to the one described above, require approximately 52 processor machine cycles to execute. The implemented, highly optimized, routine makes use of idle auxiliary registers to perform data acquisition operations, saving approximately 30 processor machine cycles. The execution time required to sample a 256 point frame is $192\mu\text{sec}$, only 10 words of program memory are used.

3.3 Hardware

As stated in Chapter Two, the hardware of the system is to be designed with two main objectives in mind. These objectives are the physical size and power consumption of the hardware. The complexity of the speech analysis system, cepstral analysis, and the execution time requirements suggest a digital signal processor based design. The first design decision to be considered is the selection of a suitable digital signal processor.

3.3.1 Processor Selection

The recent advances in VLSI technology have increased the availability of digital signal processors with short processor cycle time capable of executing signal processing algorithms such as cepstral analysis. Some of these processors available at the beginning of this project were the TMS320C25, the DSP56000, the ADSP-2100 and the $\mu\text{PD77230}$ [FLRB 86, Klok 86, Roes 86, Eich 86]. Keeping in mind the above mentioned requirements of speed, power consumption and size, the options were narrowed down to the Motorola DSP56000 and the Texas Instruments TMS320C25 digital signal processors. The implementation of cepstral

Timing & Memory Diagram

Routine	P. Memory in bytes	Time (msec)	Times Used	Total Time (msec)	P. Memory Total (bytes)
FFT	784	4.5	3	13.5	798
LOG	530	4.9	1	4.9	530
MAG	15	0.24	2	0.48	30
ISR	85	0.007	256	1.8	85
LNK	30	0.25	3	0.75	30
PP1	67	0.77	1	0.77	67
PP3	123	0.83	1	0.83	123
BHW	74	0.05	1	0.05	74
Used	---	---	---	23.08	1737
Remaining	---	---	---	8.92	2263
Total	---	---	---	32.0	4000

Table 3-1: Timing and Memory Information for The Routines

analysis is of the same complexity using either processor from a hardware designer's point of view. The TMS320C25 processor was chosen for the following reasons:

- 1- The availability of the TMS320C25 software development system at the author's research facilities.
- 2- The compatibility of the TMS320C25 with other previous TMS320 digital signal processors, thus providing access to a large program library of DSP algorithm implementations using the TMS320 family of processors.
- 3- A comparison between the TMS320C25 and the DSP56000, conducted by S. Romagnino of the Department of Electronics at Carleton, found the DSP56000 to be technically superior to the TMS320C25 [Roma 88]. However, in light of the proposed application

the author found the TMS320C25 adequate.

3.3.2 The TMS320C25

The TMS320C25 digital signal processor is a member of the TMS320 family of VLSI digital signal processors and peripherals developed by Texas Instruments. The TMS320C25 has the following relevant features:

- 100nsec instruction cycle time
- 544 words of on-chip data RAM
- 4K words of on-chip mask programmable ROM
- Single cycle multiply/accumulate instructions
- Object code compatibility with the TMS32020
- 16-bit instruction and data words
- 32-bit ALU and accumulator
- Block moves for data/program management
- Eight auxiliary registers and a dedicated arithmetic unit
- Bit-reversed indexed addressing mode for R-2 FFTs
- Three external, maskable user interrupts
- 40-MHz external clock frequency

3.3.3 The Board Level Implementation

The hardware was designed for use in two types of hearing aid systems developed in the Department of Electronics at Carleton University, a tactile hearing aid and a cochlear implant [Rahr 90]. The system has been implemented on a printed circuit board 12x10cm. The board consists of two sections, a digital section and an analog section. The analog section an ADC and the analog supporting hardware. The digital part contains the digital signal processor, memory and the required supporting hardware.

The analog section consists of a 14-bit analog to digital converter (TLC32041C) and interfaces to the processor using two 8-bit address decoders (74AS299). The digital section consists of the digital signal processor (TMS320C25) and of fast access time memories (CY7C29k-25) as not to incorporate any wait states in the hardware design.

Figure 3-3: The Board

3.3.4 Hardware Summary

The physical size of the hardware is approx. 12x10x3cm and weighs 400 grams. These specifications give the hardware wearable/portable capabilities. The estimated average power consumption of the system is 0.6 Watts. Two 3.0 volt "C" batteries would operate the system for approximately 24 hours before recharging would be required [EAC 989].

3.3.5 The Cochlear Implant System Hardware

The hardware of the cochlear implant system is divided into four parts:

- 1- The speech processing unit which has two distinct parts, the digital and the analog.

- 2- A miniature, omnidirectional microphone.
- 3- An encoder/transmitter as well as a receiver/electrode stimulator hardware designed at Sherbrooke University, Quebec, Canada [SADM 88].
- 4- The electrode, a multi-channel bi-polar electrode is to be used. The electrode was designed and manufactured in the Department of Electronics at Carleton University [PFSK 89, PPFS 89].

Chapter 4

System Evaluation and Performance

Evaluating any engineering product consists of testing the performance and the suitability of the product for the proposed application. It is important at this stage, before introducing the evaluation strategy and its implementation, to restate the defined scope of research for this project. This project is intended to implement a real time speech feature extraction system to interface with an existing stimulator/electrode for a cochlear implant system. As outlined in Chapter Two, the required speech features are the pitch (F0) and the formant frequencies (F1, F2, F3).

The evaluation of the cepstral speech processing system has been performed on four separate levels. These evaluation levels are the integrity of algorithm development, the software, the hardware and the suitability of the system for the proposed application. The integrity of the developed cepstral algorithm is evaluated using the speech sample "We were away a year ago". The system's integrity was tested by comparing the tracked formant frequencies (using the system) to the spectrogram of the speech sample. The software is evaluated by testing the validity of the implemented cepstral analysis system using a non-speech signal. The hardware is evaluated by assessing the conformance of the hardware to the preset physical constraints discussed earlier in Chapter Two. As for the suitability of the cepstral system for use in the cochlear implant system, that is tested by evaluating the discriminatory capabilities of the implemented system.

4.1 Integrity of Cepstral Analysis Development

In this test, the speech sample "we were away a year ago" is used to evaluate the integrity of the method of implementation of cepstral analysis. The strategy behind this test is to see the correlation between the spectrogram of the speech sample and the tracked formant frequencies using the suggested implementation for cepstral analysis. The tracked formant frequencies as given by the implemented cepstral analysis system are seen in Figure 4-3. It is observed that the tracked values are similar to the spectrogram of the test sample.

The above figure shows the speech sample, the test sample consists of 15001 points sampled at 8kHz.

Figure 4-1: The Test Speech Sample "We Were Away A Year Ago"

4.2 Validity of The Cepstral Implementation

Due to unavoidable modifications to the theoretical algorithm used, and the fact that the algorithm has been implemented in fixed point arithmetic, speech resynthesis was not employed to validate the implementation of the cepstral algorithm. This is because the cepstral analysis system has been implemented for feature extraction only and many parameters needed for speech resynthesis were not stored or tabulated throughout the implementation of the cepstral system. Therefore, other validating tests had The tests presented below can be grouped as the functional verification tests and the performance evaluation test.

The figure shown above displays the spectrogram of the speech sample (obtained using the Fourier Transform), this waveform is also the result of the first stage of cepstral analysis.

Figure 4-2: The Spectrogram of The Test Sample

4.2.1 Non-Speech Data

This test consisted of applying a sampled $x(t)$ waveform to the cepstral system, such that

$$x(t) = 10\sin(150t) + \sin(600t) + \sin(1700t) + \sin(2200t) \quad (4.1)$$

$x(t)$ was chosen to have a carrier frequency comparable to normal human pitch, and to contain dominant frequency representation corresponding to the first three formant frequency ranges (as given earlier in equation 3.23). The result of the analysis of $x(t)$ using the implemented cepstral analysis system is shown in table 4-1.

The figure shown above displays the tracked formant frequencies, F_1 , F_2 , and F_3 of the speech sample (obtained using the implemented cepstral analysis system). For comparison refer to [Flan 72] pp 179 Figure 5.3.0 panel B.

Figure 4-3: Tracking of The First Three Formant Frequencies

4.2.2 Cepstrum Comparison

The cepstrum of a speech sample is an evaluation of the pitch period of the speech sample. The pitch period is measured by calculating the distance from the origin of the cepstral peak along the time axis. The pitch period calculation of the implemented cepstral system was compared to that of the a commercial digital signal processing software package (The Interactive Laboratory System 'ILS' digital signal processing software package was used as a reliable commercial processing tool for all the tests). For a speech frame of the vowel 'a', spoken by an English speaking male Canadian sampled at 8 kHz for 32 ms, the implemented cepstral system detected a pitch value of 117.65 Hz (the location of the

Pitch = 156Hz

Formants	F1	F2	F3
Frequency	625Hz	1688Hz	2160Hz
Amplitude	117X	106X	102X

where X is a relative unit of measurement

Table 4-1: Non-Speech Data Test

cepstral peak was at 8.5 ms from the origin), ILS detected a value of 119 Hz. It is observed that the variance in the result is less than 1 percent.

4.3 The Cochlear Implant System Hardware

In Chapter Two, two essential restrictions were imposed upon the cochlear implant target hardware. These restrictions are the power consumption and the physical size of the target hardware. The hardware of the system has an average power consumption of 0.6 Watts, a physical size of 12x10x3 cm and weighs approximately 400 grams. The above mentioned physical specifications render the system wearable, and the electrical specifications present a number of cost effective power supply options. The physical size of the cochlear implant target hardware in Figure 4-4, is compared to a Canadian 5 cent coin.

4.4 Discriminatory Capabilities of Cepstral Analysis

An important test for the suitability of a speech processing strategy in cochlear implants is the discriminatory capability of the speech processing system [MiTC 84, King 87]. As lip reading is important for speech perception in post implant patient training, a cochlear implant speech processing strategy should be capable of differentiating between voiced and unvoiced speech (voicing of speech cannot be detected using lip reading techniques). Two speech samples "atta" and "adda" are used for this test. Figure 4-5 shows the results of processing the two speech samples. Figure 4-5 consists of two three dimensional plots (time vs. frequency vs. amplitude) of the vocal tract characteristics as produced by the cepstral

Figure 4-4: Target System Hardware

system. It is observed that the results of the analysis differed between the speech samples at the voicing interval. This capability of the cepstral analysis system would make it possible for a cochlear implant patient to distinguish between the two sounds that only differ in their voicing.

Spectral envelope of "ATTA" using Cepstral Analysis

Spectral envelope of "ADDA" using Cepstral Analysis

energy
time
frequency

Figure 4-5: Discriminatory Test Using "atta" and "adda"

Chapter 5

Conclusions

5.1 Achievements and Contributions

This thesis proposes a novel speech processing strategy, cepstral analysis, for the auditory prosthesis. The speech processing strategy is real time cepstral analysis and has been implemented using fixed point arithmetic on the TMS320C25 digital signal processor. The digital signal processing system performs favourably when compared to other known speech processing methods such as LPC and power spectrum calculations. The cepstral system results set it as a candidate for pitch extraction and formant tracking in a variety of applications [Daniloff 1989]. Execution time is 73 percent of the available time, permitting the addition of other programs that may cater the implementation of the cepstral analysis system to other applications. The size of the supporting hardware makes it wearable and may be further reduced when an EPROM version of the TMS320C25 is used. Taking into consideration recent breakthroughs in rechargeable lithium battery technology, the system can operate for 24 hours prior to recharging.

5.2 Future Work

The field of cochlear implant research is relatively new. There remains a large number of untouched areas for research and development. It is important to note that the vague understanding of the auditory process has led to a loose definition of the signal processing requirements for the cochlear implant, thus making the problem of defining clear signal processing requirements for the cochlear implant exceedingly difficult.

One of the latest findings in speech processing research for the cochlear implant is the importance of relaying the temporal information of speech to the cochlear implant patients, temporal information of speech is thought to give cochlear implant patients improved speech perception [Eddi 90]. Therefore, incorporating temporal information calculations in the implemented speech processing strategy is a task worth considering. The implementation of

a current spread corrective routine, as proposed in 2.5, would give way to multiple stimulation strategies that may account for the physical characteristics of the cochlea. Current spread corrective measures would give an improved imitation of the natural auditory process occurring in the natural cochlea. Other proposed work in this area is accounting for background noise [Dill 88], a VLSI implementation of the algorithm and an improved VLSI stimulator design to interface with the hardware.

References

- [BoHT 63] B. Boger M. Healy J. Tukey, "Proceedings of the Symposium on Time Series Analysis", John Wiley and Sons, New York, 1963, pp 209-243.
- [Berg 69] G. Bergland, "A Guided Tour of the Fast Fourier Transform", IEEE Spectrum, July 1969, pp 228-239.
- [BIMC] P. Blamey L. Martin G. Clark, "A Comparison of Three Speech Coding Strategies Using an Acoustic Model of a Cochlear Implant", JASA, January 1985, pp 209-217.
- [Boot 89] A. Boothroyd, Private Communications, Engineering Foundation Conference on Auditory Prostheses, August 1989.
- [BuEs 81] C. Burrus P. Eschenbacher, "An In-Place In-Order Prime Factor FFT Algorithm", IEEE ASSP, ASSP 29, 1981, pp 806-817.
- [Carl 86] Carleton University, "Summary of TMS 320 Software for a Cochlear Implant", Internal Report, Carleton University, August 1986.
- [Coch 86] The Nucleus Group, "A Sound Improvement", Cochlear Inc., 1986.
- [Dill 88] N. Dillier, "Programmable Master Hearing Aid with Adaptive Noise Reduction Using a TMS320-20". IEEE Proc. ASSP, 1988, pp 2508-2511.
- [Dorm 89] M. Dorman, Private Communications, Engineering Foundation Conference on Auditory Prostheses, August 1989.
- [EnAC 87] Promotional Pamphlet, Engineered Assemblies Corporation, P.O.Box 360, Clifton, NJ USA, 1987.
- [Eddi 90] D.Eddington, Private Communications, 2nd International Cochlear Implant Symposium, June 1990.
- [Eich 86] B. Eichen, "NEC's μ PD77230 Digital Signal Processor", IEEE, Micro, December 1986, pp 60-69.
- [EnOC 86] A. Engerbretson M. O'Connel, "Implementation of a Microprocessor-Based Tactile Hearing Prosthesis", IEEE, BME, V-33, No-7, July 1986, pp 712-716.

- [Feig 87] E. Feigenbaum, "Cochlear Implant Devices for The Profoundly Hearing Impaired", IEEE EMB, June 1987, pp 10-21.
- [Flan 72] J. Flanagan, "Speech Analysis Synthesis and Perception", 2nd Edition, New York, 1972.
- [FLRB 86] G. Frantz K. Lin J. Reimer J. Bradley, "The Texas Instruments TMS320C25 Digital Signal Microcomputer", IEEE, Micro, December 1986, pp 10-28.
- [Gede 88] I. Gedeon, "Speech Processing for the Cochlear Implant", Thesis Proposal, Electronics, Carleton University, 1988.
- [Geis 87] D. Geisler, "Coding of Acoustic Signal on the Auditory Nerve", IEEE EMB, June 1987, pp 22-28.
- [Geis 89] D. Geisler, Private Communications, Engineering Foundation Conference on Auditory Prostheses, August 1989.
- [GeCh 86] J. Genin R. Charachon, "Sound Processing for the Cochlear Implant", IEEE ICASSP 86, Tokyo 1986, pp 621-623.
- [GeCh 87] J. Genin R. Charachon, "Electrical Interactions - Measurements and Correction", Int'l Cochlear Implant Symposium, Duren, September 1987, pp 359-364.
- [Goub 88] R. Goubran, Private Communications, Dept. of Computer and Systems Engineering, Carleton University, 1988.
- [Hart 68] J. Hart, "Computer Approximations", Huntington, Kreiger, New York, 1968.
- [HPDM 87] R. Harrison P. van der Puije F. Duval R. Morris H. Kunov, "Technical Development of an Implantable Cochlear Prosthesis in Canada", The Journal of Otolaryngology (16:5), 1987, pp 311-315.
- [IfWh 86] T. Ifukube R. White, "A Processor with Lateral Inhibition for an Eight Channel Cochlear Implant and Its Evaluation by Subject Testing", IEEE ICASSP 86, Tokyo 1986, pp 617-620.
- [King 87] A. King, "Sound Pattern Tests for Cochlear Implant Patients", Int'l Cochlear Implant Symposium, Duren, September 1987, pp 319-322.
- [Klok 86] K. Kloker, "The Motorola DSP56000 Digital Signal Processor", IEEE, Micro, December 1986, pp 29-48.

- [LuSi 84] H. Lusted F. Simmons, "Interaction of Cortical Evoked Potentials to Electric and Acoustic Stimuli", JASA, August 1984, pp 449-455.
- [Makh 84] J. Makhoul, "Linear Predictive Coding Electronic Speech Synthesis", Granada Publishing Ltd., 1984.
- [Mark 72] J. Markel, "The SIFT Algorithm for Fundamental Frequency Estimation", IEEE Trans. on Audio and Electroacoustics, December 1972, pp 367-377.
- [Math 78] R. Mathews, "Sound Processing for an Auditory Prosthesis", Stanford University Phd. Thesis, June 1978.
- [MiTC 84] J. Millar Y. Tong G. Clark, "Speech Processing for Cochlear Implant Prostheses", ALSHA, June 1984, pp 280-296.
- [Mor1 85] L. Morris, "A Comparative Study of Time Efficient FFT and WFTA Programs for General Purpose Computers", IEEE, ASSP, ASSP 26, 1978, pp 141-150.
- [Mor2 85] L. Morris, "Digital Signal Processing Software", DSPS Inc., 1985.
- [NoSc 64] A. Noll M. Schroeder, "Short-Time Cepstrum Pitch Determination", JASA., vol 36, 1964, pp 1030.
- [Noll 67] A. Noll, "Cepstrum Pitch Determination", JASA, vol. 41, 1967, pp 293-309.
- [Nutt 81] A. Nuttall, "Some Windows with Very Good Sidelobe Behavior", IEEE, ASSP, V-29, No-1, February 1981, pp 84-91.
- [OKRS 85] E. Owens D. Kessler M. Raggio E. Schubert, "Analysis and Revision of the Minimal Auditory Capabilities Battery", Ear and Hearing, The Williams and Wilkins Co., 1985, pp 280-290.
- [O'Sh 88] D. O'Shaughnessy, "Linear Predictive Coding, One Popular Technique for Analyzing Signals", IEEE Potentials, February 1988, pp 29-32.
- [Pars 86] T. Parsons, "Voice and Speech Processing", McGraw-Hill Book Company, 1986, pp 197-210.
- [Pauk 87] C. Pauka, "Vowel-Pitch Discrimination in Cochlear Implant Patients", International Cochlear Implant Symposium Duren, Duren, September 1987, pp 101-113.
- [Pfin 87] B. Pfingst, "Stimulus Parameters Affecting the Discrimination of Temporal Frequency Cues", International Cochlear Implant Symposium Duren, Duren, September

ber 1987, pp 51-65.

- [PFSK 89] P. van der Puije Y. Fujimoto R. Shelley T. Kwasniewski, "In-Vitro Tests of a Thin Film Cochlear Electrode", Proc. International Symposium on Cochlear Implants, June 1989, Toulouse, France, under publication.
- [PPFS 89] P. van der Puije A. Petosa Y. Fujimoto R. Shelley, "Thin-Film Stimulation Electrode: Mechanical and Electrical Test Results", Proc. 3rd Int'l Vienna Workshop on Functional Electrostimulation, September 1989, under pp 238-242.
- [PuDM 87] P. van der Puije F. Duval R. Morris, "The Canadian Cochlear Implant Project: Design Objectives and Development", International Cochlear Implant Symposium Duren, Duren, September 1987, pp 381-394.
- [Rabi 89] W. Rabinowitz, Private Communications, Engineering Foundation Conference on Auditory Prostheses, August 1989.
- [RaGo 75] L. Rabiner B. Gold, "Theory and Application of Digital Signal Processing", Prentice-Hall, 1975.
- [Rahr 90] T. Rahrer, "Signal Processing for a Tactile Hearing Aid", M.Eng. Thesis, Carleton University, 1990, in preparation.
- [RSc 78] L. Rabiner R. Schafer, "Digital Processing of Speech Signals", Prentice-Hall, 1978.
- [Rile 89] M. Riley, "Nature of Speech Processing Signals", MIT Phd Thesis, January 1989.
- [Roes 86] J. Roesgen, "The ADSP-2100 DSP Microprocessor", IEEE, Micro, December 1986, pp 49-59.
- [SADM 88] M. Sawan R. Abtahi F. Duval F. Mailhot, "A Programmable Neural Stimulation Signal Generator", Proc. CCVLSI, Halifax, 1988, pp 238-245.
- [Schw 87] J. Schwartzman, "Comparative Results with the House 3M vs. the Nucleus Cochlear Implant", Proc. Int'l Cochlear Implant Symposium, Duren, September 1987, pp 645-651.
- [ScRa 70] R. Schafer L. Rabiner, "System for Automatic Formant Analysis of Voiced Speech", JASA, vol 47, 1970, pp 634-648.
- [Sear 89] C. Searle, "Modelling of the Human Auditory Perception System", submitted

to JASA, April 1989.

- [CoHB 86] H. Sorensen M. Heidman C. Burrus, "On Computing The Split-Radix FFT". IEEE ASSP, February 1986. pp 152-156.
- [STI 986] Signal Technology Inc., "ILS Command Reference Guide", ILS V6.0. STI. 1986, pp CPT1-CPT5.
- [StHo 87] P. Stypulkowski C. van den Hornet, "Coding of Electrical Stimuli on the VIIIth Nerve", IEEE EMB, June 1987. pp 33-35.
- [TeIn 81] Texas Instruments, "Digital Signal Processing Applications with the TMS320 Family", Texas Instruments, 1981.
- [TeIn 86] Texas Instruments, "The TMS320C25 Software Development System", Texas Instruments, 1986.
- [Town 87] B. Townshend, "Electrical Interaction in Auditory Prostheses". Phd. Thesis, Stanford University, July 1987.
- [TGKT 87] N. Tye-Murray B. Gantz F. Kuk R. Tyler, "Word Recognition Performance of Patients Using Three Different Cochlear Implant Designs", Proc. Int'l Cochlear Implant Symposium, Duren, September 1987, pp 605-612.
- [TrQu 79] J. Tribolet T. Quantieri, "Computation of the Real Cepstrum and Minimum Phase Reconstruction", Programs for Digital Signal Processing, IEEE Press, John Wiley and Sons, New York, 1979, pp 7.2.1-7.2.6.
- [WDDE 87] B. Wilson L. Dent N. Dillier D. Eddington I. Hochmair-Desoyer B. Pfingst J. Patrick W. Surth J. Walliker, "Round Table Discussion on Speech Coding", Proc. Int'l Cochlear Implant Symposium, Duren, September 1987, pp 693-704.
- [WhCo 85] R. White D. van Compernelle, "Current Spreading and Speech Processing Strategies for Cochlear Prosthesis", Proc. Int'l Cochlear Implant Symposium, Melbourne, 1985, pp 22-24.
- [Whit 87] R. White, "System Design of a Cochlear Implant", IEEE EMB, June 1987, pp 42-46.
- [WoFa 85] W. Woods F. Fallside, "Computer Speech Processing", Prentice-Hall, 1985.

Bibliography

- Atlas, Auditory Coding in Higher Centers of The CNS, IEEE, EMB June 1987, pp 29-32.
- Ball and Rosen and Walliker and Fourcin, A Speech Pattern Approach to Electrical Hearing, International Cochlear Implant Symposium 1987, Duren, 1987, pp 241-261.
- Belanger, Digital Processing of Signals, John Wiley and Sons, 1984.
- Bertin et Compagnie, The BERTIN Cochlear Implants, Promotional Pamphlet.
- Blamey and Dowell and Brown and Clark and Seligman, A Formant-Estimating Speech Processor for Cochlear Implant Patients, Speech Communication by Elsevier Science Publishers B.V., North and Holland, 1987, pp 293-298.
- Borraja and Aguilera and Pardo and Munoz, An Efficient Pitch Extraction Method for Diagnosis and Education, MELCOM 85 vol.1, IEEE Bio-Engineering 1985.
- Cater, Electronically Hearing Computer Speech Recognition, Howard Sams and Co. 1984.
- Clark and Brimacombe and Webb and Dowell and Meck, Speech Recognition Abilities in Profoundly Deafened Adults Using The Nucleus 22 Channel Cochlear Implant System, Internal Report by Cochlear Corp. (UK) and The Royal Victorian Eye and Ear Hospital (Australia).
- Cochlear Inc., Nucleus Mini 22 Cochlear Implant System, Promotional Pamphlet.
- Conway, UCH/RNID Cochlear Implant Performance (Description of Sound Processing and Implant), International Cochlear Implant Symposium 1987, Duren, 1987, pp 279-282.
- Dormer and Phillips, Auditory Prosthesis: Implantable and Vibrotactile Devices, IEEE, EMB June 1987, pp 36-40.
- Droucher, LSI Implementation of Speech Analysis (A Review of Three Integrated Circuits), Internal Report, Dept. of Electronics, Carleton University, 1987.

- Gold and Rabiner, Parallel Processing Techniques for estimating Pitch Periods of Speech in The Time Domain, JASA, vol. 46 no.2, 1969.
- Gold, Note on Buzz Hiss Detection, JASA, vol. 36 no.9, Sept. 1964.
- Hermes, Measurement of Pitch by Subharmonic Summation, JASA 83(1), 1988, pp 257-264.
- Hess, Algorithms and Devices for Pitch Determination of Speech Signals (Automatic Speech Analysis and Recognition), D. Reidel Publishing Co., 1982, pp 49-67.
- Jovanovic, A New Algorithm for Speech Fundamental Frequency Estimation, IEEE ASSP, vol. 34 1986, pp 626-630.
- Krishnamurthy and Li and Moses, Format Estimation from Noisy Voiced Speech, Internal Report by Dept. of Electrical Engineering at Ohio State University.
- Kunov and Hong and Pandey, Digital Speech Processor for Single Channel Prosthesis, International Cochlear Implant Symposium 1987, Duren, 1987, pp 667-672.
- Maksym, Real-Time Pitch Extraction by Adaptive Prediction of The Speech Waveform, IEEE, Trans on Audio and Electroacoustics, vol. 21, June 1973, pp 149-154.
- Mikami and Ohba, Fast Algorithm for Real-Time Formant Extraction by a Microprocessor, IECON 1984, 1984, pp 1148-1151.
- Morris, Digital Signal Processing Microprocessors: Forward to The Past?, IEEE, Micro, Dec. 1986, pp 6-8.
- Mourikis and Koukoutsis and Halkias and Carayannis, Two New Algorithms for Pitch Detection, IEEE Mediterranean Electro-Technical Conference, Athens 1983, pp 1.04-1.05.
- Nair, A Mixed Digital/Sampled Analog Signal Processing Architecture, Doctorate Thesis, Dept. of Electronics, Carleton University, 1980.
- Pearce and Whitaker, Reference Formant Analysis, Internal Report by GEC Research Ltd. (UK), 1986.
- Shaw, A Block Demodulator for Frequency Division Multiplexed NCFSK Signals, Master Thesis, Dept. of Electronics, Carleton University 1986.
- Sondhi, New Methods of Pitch Extraction, IEEE, Trans. on Audio and Electroacous-

tics, vol.16, June 1968, pp 262-266.

- Texas Instruments (Digital Signal Processor Products), First Generation TMS320 User's Guide, Texas Instruments 1987.
- 3M, Hear The Word, Promotional Pamphlet, 3M Vienna Cochlear Implant Program.
- Tribolet and Quantieri, Computation of The Complex Cepstrum (Programs for Digital Signal Processing), IEEE Press, John Wiley and Sons, 1979, pp 7.1.1-7.1.11.
- Tye-Murray and Gantz and Kuk and Tyler, Word Recognition Performance of Patients Using Three Different Cochlear Implant Designs, Internal Report by Dept. of Psychology at Iowa University.
- Tyler and Gantz and Preece and Lansing and Lowder and Otto and McCabe, Cochlear Implant Program at The University of Iowa, Proc. of The Scott Reger Conference, University Park Press, Baltimore 1984.
- Ulriksson, Conversion of Frequency Domain Data to The Time Domain, IEEE, ASSP vol. 34, Jan. 1986.
- Verhelst and Decker and Francq and Steenhaut, An Adaptive Signal Clipping Processor (ANUSC) for Real-Time Autocorrelative Pitch Detection, IEEE ICASSP Tokyo, 1986, pp 121-124.
- Whitaker and Pearce, Larynx Synchronous Formant Analysis, Internal Report by GEC Research Ltd. (UK).
- White and Finley and Wilson, electrical Stimulation Model of The Auditory Nerve: Stochastic Response Characteristics, IEEE, 9th Annual Conference of EMB, 1987, pp 1906-1907.
- White and Herndon and Mathews, The Use of Computers in The Development of a Cochlea Prosthesis (from Uses of Computers in Aiding The Disabled), North-Holland Publishing Co., 1982, pp 147-163.

END

1 2 0 4 9 1

FIN