SYSTEM FOR ANALYSIS OF VERBAL BEHAVIOUR OF SPEAKERS WITH NEUROLOGICAL DISEASES

by

Budi Rahardjo

A Thesis
Presented to the University of Manitoba
in Partial Fulfillment of the Requirements for the Degree of
Master of Science
in
Electrical Engineering Program

Electrical and Computer Engineering Department
University of Manitoba
Winnipeg, Manitoba

August 1990

ISBN 0-315-63242-9

SYSTEM FOR ANALYSIS OF VERBAL BEHAVIOUR OF SPEAKERS WITH NEUROLOGICAL DISEASES

BY

BUDI RAHARDJO

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Abstract

This thesis describes a system for analyzing the verbal behaviour of speakers with neurological disease. The system was developed on an IBM PC. A two-channel Analog to Digital Converter board and two low-pass filters were developed in order to sample voices from the speakers.

Several parameters may be extracted from the sample signal, with the most important parameter being the pitch period of the signal. A pitch period detection by means of modified autocorrelation function and simple time domain peak detector is described. This method uses center and infinite clipper in order to reduce the volume of computations and to avoid incorrect pitch period classifications.

Some elementary statistical analysis was done on the samples to analyze which parameters are relevant for classifying normal and pathological speakers. The parameters used in the analysis are the deviation of fundamental frequency (df0), degree of hoarseness (dh), frequency perturbation quotient (fpq), minima perturbation quotient (a2pq), jitter in fundamental frequency (fdlt), shimmer in minima (mindlt), directional perturbation factor of fundamental frequency (fdpf) and directional perturbation factor of minima (mindpf). The analysis was done with SAS.

A VLSI implementation of autocorrelation function is described. It could be used to improve the speed performance of the calculation. The implementation was done with Cadence Design Systems, which allowed for the design

in schematic level. Circuit simulation was done with SILOS.

Acknowledgements

I would like to express my sincere thanks to Prof. M. Yunik for his supervisor throughout my research.

I would also like to thank Dr. B. Boyanov of the Bulgarian Academy of Sciences for his initial research in this area and for his guidance especially at the preliminary stage.

A special acknowledgement as well to Dr. E. Cardoso of the Health Sciences Center for providing the necessary speech samples.

In conclusion I would also like to extend my sincere appreciation to the Inter-University Center (IUC) for Microelectronics in Indonesia for allowing me to further my studies abroad and to the World University Services of Canada (WUSC) for providing the neccesary support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Speech-related parameters such as average pitch period, degree of hoarseness, deviation in pitch period, frequency perturbation factor, and amplitude perturbation factor, have been used in various areas of study. For example, studies have been done in *speech communication* to analyze and recognize speech [13]. In other fields such as *phonetics and linguistics* studies have been done to analyze phonetic features [15]. In *education*, speech-related parameters are used to help the deaf in learning how to speak correctly, analyze speech disorders in children, and to assist in learning a foreign language [15]. Speech-related parameters are also used in *medicine, pathology, and psychology* to diagnose the condition of a patient, which is the main interest of this thesis.

It is known that some diseases change the speech quality of a patient. The results of several studies have shown that most laryngeal [1] [14] [19] [24] [36], neurological [6] [7] [11] [26] diseases, and changes in emotional state

[1] [8] cause significant changes in speech signal. If these changes can be formulated and measured, it may be useful in diagnosing the condition of a patient. The main interest of this thesis is the development of a computer-based system for speech signal analysis, which will be used in analyzing the presence of laryngeal and neurological diseases.

The correlation between laryngeal and neurological disease with voice parameters, such as *pitch period perturbation* was studied in [19] [24] [25], *amplitude perturbation* in [19] [21], *rate of unvoice* in [19], the functional status of larynx sources with a fuzzy approach was studied in [12], and *degree of hoarseness* in [19].

The availability of analog to digital converters open the possibility of performing the analysis of speech signal in the digital domain. The advantages of performing the analysis in the digital domain compared to analog domain are :

1. It opens the possibility of using digital devices, such as digital computers and digital signal processing chips.

2. It is more flexible in operation. Different methods can be implemented on the same system using different programs. Changing one method to another is a matter of running the proper program, whereas in analog the methods are hardwired for the implementation. Any alterations might require significant hardware changes.

3. The data is more insensitive to noise, since it is stored in digital format. The quality of the data does not change with time and the number of playbacks, whereas the quality of a recorded analog signal will reduce after a number of playbacks.

4. Digital circuits are more insensitive to noise, whereas analog circuits suffer from drift caused by temperature, humidity and time.

5. The data can be stored into disk for future analysis without reducing quality. Duplicating, copying, and transferring the data can be done easily.

6. On a networked system, where the data is stored in a shared disk, a number of computers or processing devices can share the same data with the same quality.

7. Digital components are relatively inexpensive compared to high-quality analog components.

The disadvantages of using digital domain are :

1. The quality of the data depends on the quality of the devices used to sample the analog data. For example, the number of bits or sampling rate of the Analog to Digital Converter (ADC) effects the accuracy of the digital representation.

2. High quality and fast ADC is relatively expensive.

3. Large memory or storage is needed to store the data. A system with 16 bits resolution and 20 kHz sampling rate would require 40,000 bytes of storage for every second of speech signal.

From the brief description above, analysis on digital domain has more advantages over analog.

Most of the work that has been done required special devices, such as a special microphone [19] [21], Digital Signal Processing (DSP) chips, or mini or mainframe computers [5] [14] [16] [24] [25]. The advantages of this method is fast and accurate. This method is however expensive and inflexible. Modification or fine tuning is also difficult and the users must have the same devices.

The availability of inexpensive and powerful personal computers or workstations, which sometimes are equipped with ADC, yields to an alternative method for analysis. This idea would allow physicians, clinicians or researchers to do analysis on their own computers in their own offices. A computer-based system for analyzing speech signal is proposed in this thesis. This implemented system has the following characteristics :

1. It requires a minimum additional hardware, namely a microphone and an Analog to Digital Converter. If the computer is equipped with an ADC board, no additional hardware is required. In the event that an ADC board is required, the additional ADC is relatively low priced.

2. It is compact, since only the computer and the microphone are used.

3. It uses two-channel ADC thereby avoiding special DSP chips.

4. It uses methods and procedures with reduced volume of computation. For example the FFT computation is done only on stable zones.

5. The influence of amplitude distortion is reduced since the pitch period is evaluated on band-limited signals using an autocorrelation function, which is relatively insensitive to external noise and phase distortion.

6. The results and data could be stored on floppy disk for future analysis.

7. The speech signal can be digitized directly into the computer's memory. This means the system does not require an expensive high-quality tape recorder.

8. It is safe to be used by untrained person.

In this thesis a computer-based system for analyzing verbal behavior of patients with neurological disease is proposed. The system only requires minimal additional hardware, namely ADC. The analysis is done in software with methods and procedures which reduce the volume of computation, thus the analysis can be done faster.

This thesis is organized as follows :

Chapter 2 discusses methods of digital signal processing of speech. Some theory and methods in estimating pitch period, pitch period perturbation, amplitude perturbation, and degree of hoarseness will be presented.

Chapter 3 discusses method, software, and hardware in the implemented system.

Chapter 4 discusses the application and preliminary result of the system in analyzing voices of normal speakers, speakers with laryngeal diseases, and speakers with neurological diseases.

Chapter 5 will discuss VLSI implementation of autocorrelation function, which is used to estimate the pitch period of a signal.

Chapter 6 contains the conclusions and suggestions for future work.

Chapter 2

Methods for Digital Signal Processing of Speech

There are several speech related parameters that can be extracted from speech signals, namely average pitch period, degree of hoarseness, deviation in pitch period, and pitch period perturbation. Some of these parameters depend on the pitch period of the signal. For example, *degree of hoarseness*, which can be viewed as a Noise to Signal ratio, is calculated based on the power of the harmonic components (fundamental frequency, sub harmonic, and harmonic components), and the inharmonic components. Therefore care must be taken in designing and implementing a method for pitch detection.

After a brief survey of methods for detection, two pitch detection methods, namely simple time-domain pitch detection and pitch detection with improved accuracy using an autocorrelation function, will be presented.

2.1 Survey of Methods for Pitch Detection

Pitch detection, which has been an interest for more than half century, seems to be an easy task, but in practice it is among the most difficult problems in speech analysis. Many methods have been proposed and implemented, each with its own strengths and limitations.

Pitch period detection / estimation can be divided into three categories, namely *time-domain analysis*, *short-time domain analysis*, and *hybrid analysis* which is the combination of the two methods.

The oldest and the simplest method in pitch detection, is manual pitch detection. In this method the signal is displayed visually and measurement is done by an operator. This approach is good only for a small amount of data. With large amount of data, such as a sample from a one-minute speech, this tedious task is not reasonable.

2.1.1 Time-domain Pitch Detection

As the name implies, in time-domain pitch detection, the analysis is done on the waveform itself by inspecting the waveform features in the time domain. The features can be peaks or valleys [9] [17], energy of the peaks [33], peak width [18], or a combination of them.

A typical block diagram of time-domain pitch detection is shown in figure 2.1. The *preprocessor* is used to preprocess incoming data, such as performing data reduction, in order to make pitch extraction easier. Low-

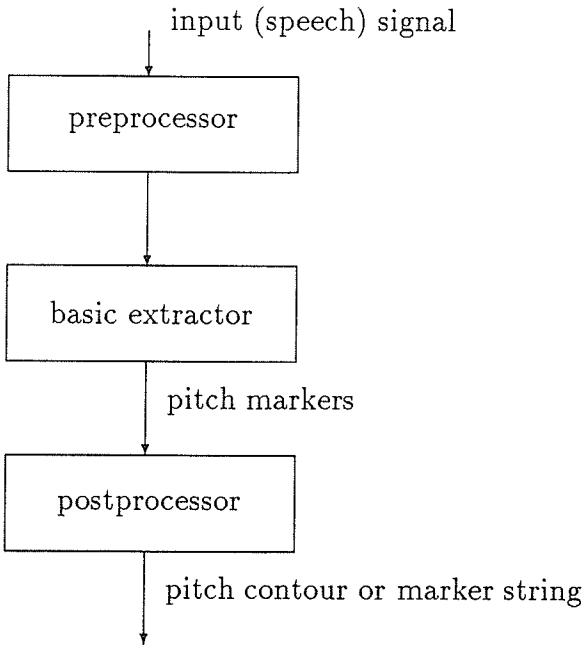


Figure 2.1: Block diagram of a typical time-domain pitch detection

pass filtering is an example of a preprocessor. *Basic extractor*, the main block, converts incoming signal into several pitch estimates. It is also able to process every pitch period (T_0) individually. By doing this, it is able to track rapid changes in T_0. However, if the signal is disturbed by noise or spurious signals, then this method may not be able to track the correct pitch period.

The output of basic extractors is usually a sequence of pulses or markers which indicate the appearance of a certain features of the signal. For example the output might be a sequence of pulses where the signal reaches the maximum values, or the signal crosses the zero value.

The *postprocessor* usually depends on the applications. This can be a

calculation of the distance between consecutive markers generated by basic extractor.

Time domain pitch extractors are based on the fact that for any periodic signal, a structured pattern appears exactly or at least approximately from period to period. This feature can be extracted in the time domain, along with the fundamental harmonic.

The advantages of time-domain pitch detection are simple to implement, able to perform real-time analysis, and able to locate any pitch period individually.

The disadvantages of time-domain pitch period extractors are :

1. The fundamental frequency must be present in the waveform. Therefore it restricts to the cases where the signal is band limited.

2. It is sensitive to low-frequency signal distortions. Therefore in an unpredictable environment, this system usually is replaced by more robust devices.

An example of time-domain pitch period detection is the *parallel processing* by Gold and Rabiner [9]. It uses six simple peak detectors, and processes different features of the signal, namely peak-to-peak value, peak-to-valley, and combination of peak and valleys.

Peak picker [17] is another example of time-domain pitch detection. Amplitude, energy of the peaks, sign of amplitude, and ratio of amplitude to energy are used as features of the pitch detector in [33].

2.1.2 Short-time Pitch Detection

In short-time pitch detection, the data is transformed into another domain. The transformation can be done with autocorrelation, Fourier transform, or inverse filtering [27]. This method is also known as spectral-domain pitch period detection.

The transformation is done in a short time interval or frame, which is then processed separately. The length of the frame depends on the application, however there is a constraint that the range is chosen so that at least two periods are on the frame. If this constraint is not met, the periodicity information is lost. If the frame length is too large, natural changes in the pitch may not be detected. For speech signals, typical frame length is 20 to 50 ms [15].

The operation is done in the following manner. After the signal is optionally preprocessed with a low-pass filter, or an adaptive center clipper, the signal is grouped into several frames or short segments. A transformation is done on this frame, for example by using autocorrelation function. The estimation is done by analyzing the peaks of the spectrum in the new domain. The output of this method is a sequence of average estimates of the fundamental frequency of the signal within the frame.

The disadvantages of this method are :

1. That it can not perform real-time calculation, since the transformation requires a large computing effort. Several methods have been proposed

to solve this problem. For example a non-linear processing such as center and infinite clipper [32] might be used.

2. It requires at least two periods of the signal inside the frame. Therefore it is unable to track the individual pitch period. As a result, rapid changes might not be detected.

The advantages of this method are :

1. Not sensitive to phase distortions

2. If the signal is periodic and regular, the indication of the pitch is strong. This makes it reliable under noisy environment.

From the overview above, it can be concluded that short-time analysis is not sensitive to phase distortion and less sensitive to noise and spurious signal. However, short-time analysis requires longer time to compute, unless it is implemented with a special trick or using a non-linear preprocessing.

Examples of short-time pitch detection are autocorrelation function with center clipper [32] and the Simplified Inverse Filter Tracking (SIFT) algorithm [27]

Hybrid analysis is developed to combine the advantages of both time-domain and short-time analysis. For example the short-time analysis is done first to get an accurate fundamental frequency, then time-domain analysis is

carried out, using the results from the short-term analysis. If the difference in the two successive fundamental frequency is too large then the short-time analysis is performed again.

2.2 Simple time-domain pitch detection

This simple time-domain pitch detection bases the detection only on one feature of the signal; namely the location of the highest peak. The algorithm starts from a peak and searches the next peak based on a prior value of the previous pitch period $\tau(i)$ and some tolerance value of k.

$$\begin{aligned} N_{max}(i+1) &= \max\{(N_{max}(i)+\tau_i - k\cdot\tau_i), \\ &\quad (N_{max}(i)+\tau_i - k\cdot\tau_i + 1), \\ &\quad \ldots, \\ &\quad (N_{max}(i)+\tau_i + k\cdot\tau_i + 1), \\ &\quad (N_{max}(i)+\tau_i + k\cdot\tau_i)\} \end{aligned} \tag{2.1}$$

where N_{max} is the location of the peak.

$$\tau(i+1) = N_{max}(i+1) - N_{max}(i) \tag{2.2}$$

From the experiment, it is found that the value of $k = 0.125$ or 12.5 % gives a good result for sustained vowel of normal speakers.

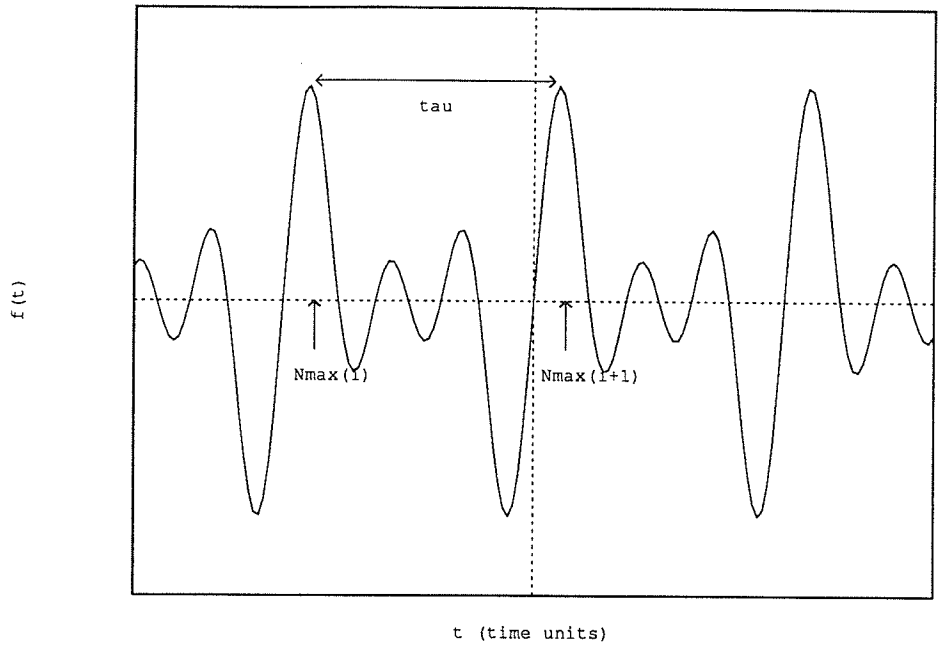


Figure 2.2: Simple Peak Detection

The advantage of this method is that it is fast and simple. However this method needs a prior knowledge of τ, which makes it unsuitable to be used by itself. Also if there is a rapid change, interruption, or discontinuity, the method might fail to detect the correct pitch period. If this occurs, a new value of τ must be given otherwise it will give incorrect results.

2.3 Pitch detection with improved accuracy using the autocorrelation function

The method presented here is based on autocorrelation function, which is a special case of *correlation* function. The correlation of two discrete signals $x(n)$ and $y(n)$, is defined as :

$$corr(\tau) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} x(n)y(n+\tau) \tag{2.3}$$

For autocorrelation, the two input signals are identical. Therefore the autocorrelation is defined as :

$$R(\tau) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{+N} x(n)x(n+\tau) \tag{2.4}$$

provided that $x(n)$ are defined for all n. The parameter τ is the delay between the immediate and the delayed signal.

If the signal $x(n)$ is finite, i.e. samples of $x(n)$ are zero outside the interval $n \in [0, N-1]$, then

$$R(\tau) = \frac{1}{N} \sum_{n=0}^{N-\tau-1} x(n)x(n+\tau) \tag{2.5}$$

Notice that the upper summation index becomes $(N-\tau-1)$, since for $n = N-\tau$ the second term is defined as zero.

The autocorrelation function, $R(\tau)$, is the inverse Fourier transform of the power spectrum of the signal [15]. As a result, phase distortions are eliminated.

According to definition (2.4), $R(\tau)$ has its highest peak at $\tau = 0$ which equals the average power of the signal $x(n)$. For a finite signal, according to (2.5), this peak indicates the average power P_{AV} during the interval $n \in [0, N-1]$.

$$P_{AV} = R(0),\ where\ n\ \in\ [0, N-1] \tag{2.6}$$

For a periodic signal $x(n)$ with period T_0, $x(n + kT_0) = x(n)$, the autocorrelation function, $R(\tau)$, is also periodic.

$$R(\tau + kT_0) = R(\tau),\ k \text{ integer} \tag{2.7}$$

The equation above shows that for a periodic signal, the autocorrelation function has significant peaks at $\tau = kT_0$.

The difficulties of T_0 detection using the autocorrelation function are :

1. When a strong second or third harmonic is present in the speech signal, then it may be classified as T_0 [2].

2. When a strong subharmonic is present (especially in most pathological voices [36], frequently it is classified as T_0 leading to a drastic error in T_0 detection [2].

3. Large volume of computation.

Many solutions have been proposed to solve the problems mentioned above. One of the solutions is to use *Spectral flattening.* This technique equalizes the speech spectrum in such a way that the spectral peaks which represent the formants are removed. This can be done by using center-clipping and signal encoding [2] [32].

The method used here involves several steps :

- preprocessing with analog filters
- signal segmentation
- voice/unvoice detection using amplitude as the feature
- threshold calculation
- signal encoding to 1, 0, and -1, by using center and infinite clipper
- $R(\tau)$ evaluation
- voice/unvoice detection
- finding all peaks in $R(\tau)$
- approach T_0 detection by using amplitude selection and logical analysis in order to find the true T_0 and to reject subharmonic and harmonic components

Preprocessing

A low-pass filter with a cutoff frequency of 700 Hz is used as a preprocessing. This filter removes all high frequency components, including the formants for the vowel *"a"*, from the speech signal. This will increase the precision of T_0 calculation using autocorrelation [2]. Results from several researchers [2] [5] [15] show that a cutoff frequency of 700-800 Hz is optimal for T_0 detection using autocorrelation.

Segmentation

The digital signal is divided into 30 ms segments. Autocorrelation requires that at least two periods of the signal be used in the calculation, with the largest pitch period in human voice being around 15 ms [15].

Segmentation is carried out with a rectangular time window, since rectangular window preserves the signal's shape. A smoothing window must not be used since smoothing has been done by the analog filter.

Voice unvoice detection with amplitude

A screening is done on the signal in the frame to check whether the signal is too low for calculation. If the signal is lower than some threshold value, then it is more likely that there is no fundamental frequency present in this particular frame. If this is the case then the segment is marked as unvoiced. It is suggested [2] that the value of 12.5 % of possible maximum value of the ADC is taken as a threshold.

Threshold calculation

Threshold calculation is done by finding the segment's global minimum and maximum. The global minimum and maximum of the i^{th} segment are found by using :

$$AMAX(i) = \max\{A^i_{max}(t)\}, t = 1, \ldots, P \tag{2.8}$$

$$AMIN(i) = \min\{A^i_{min}(t)\}, t = 1, \ldots, C \tag{2.9}$$

where :
$A^i_{max}(t)$ is the local maxima amplitude in the i^{th} segment
$A^i_{min}(t)$ is the local minima amplitude in the i^{th} segment

P is the number of local maximas
C is the number of local minimas

Two thresholds are calculated on the basis of AMAX(i) and AMIN(i) :

$$P^{+}(i) = k_{max} \cdot AMAX(i) \tag{2.10}$$

$$P^{-}(i) = k_{min} \cdot AMIN(i) \tag{2.11}$$

Several experiments with pathologic and normal voice [2] [16] have shown that $k_{max} = k_{min} = 0.75$ is the optimal value for suppression of noise components, harmonics, and subharmonics.

Signal encoding

The signal is encoded into -1, 0, and +1 by using a combination of center clipper and infinite clipper.

$$x'(n) = \begin{cases} 1 & \text{if } x(n) \geq P^{+}(i) \\ -1 & \text{if } x(n) \leq P^{-}(i) \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

The procedure is carried out in order to [2] :

1. minimize the errors caused by formants in T_0 detection, because the encoding destroys the formant structure.
2. eliminate significant noise components.
3. perform significance compression before calculating $R(\tau)$.
4. significant volume computation reduction since $R(\tau)$ is evaluated without multiplication.

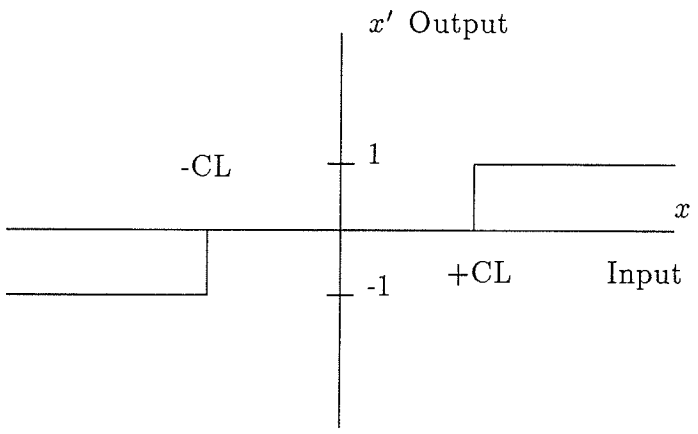


Figure 2.3: Input-output characteristic of the combination of center clipper and infinite clipper

Consider the following example. The autocorrelation function of a pure sine wave and a signal with a dominant third harmonic are shown in figure 2.4 and figure 2.5 respectively. As shown in figure 2.5, it is difficult to determine the actual fundamental frequency. If the detection is based on the highest peak, after some lag to ignore the peak at origin, then the third harmonic will be picked as the fundamental frequency.

Analysis using signal encoding (center and infinite clipper) on the signal with a dominant third harmonic produced result shown in figure 2.6. It is shown that the highest peak after the peak at origin is the correct fundamental frequency.

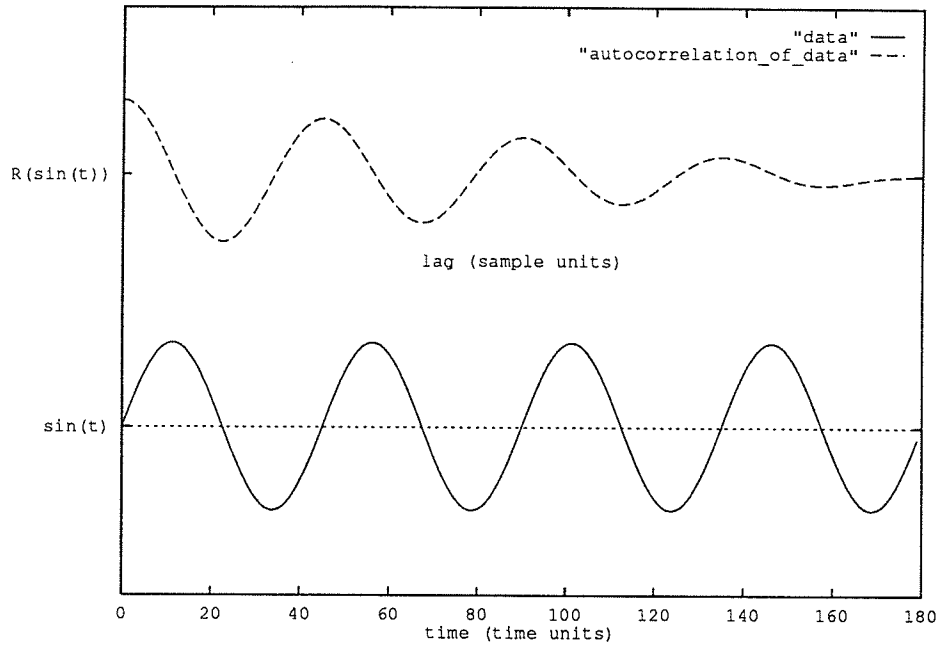


Figure 2.4: Autocorrelation of pure sine wave

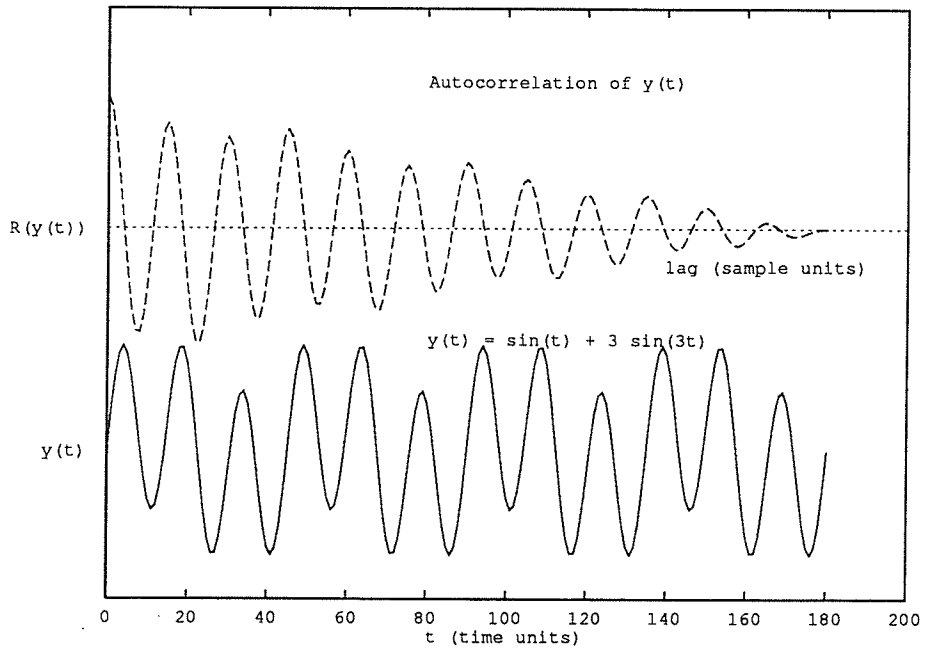


Figure 2.5: Autocorrelation of a signal with a dominant third harmonic

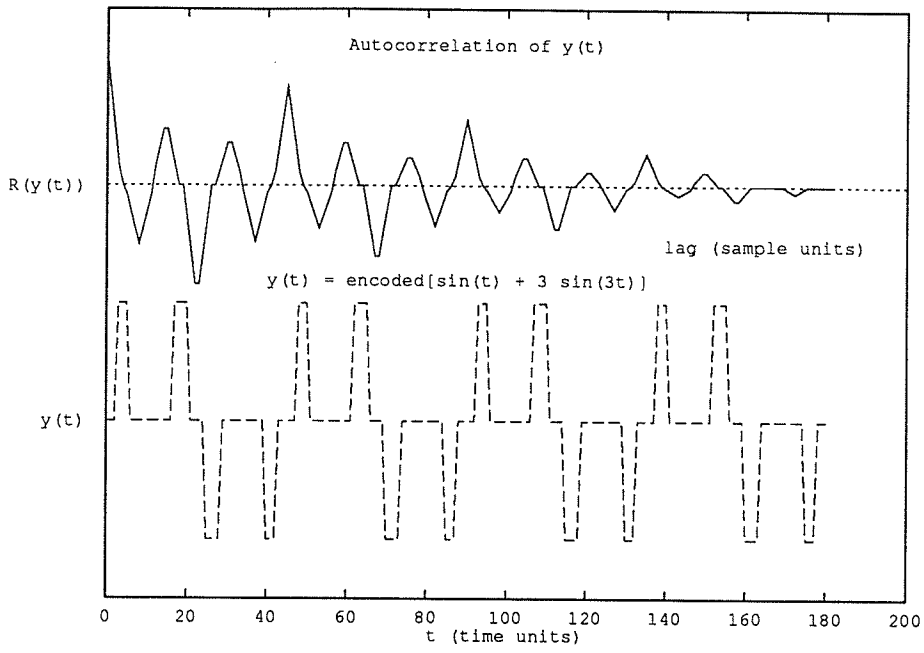


Figure 2.6: Autocorrelation of the encoded signal

Autocorrelation calculation

For every i^{th} segment, $R(\tau)$ is calculated with a modified form of equation 2.5. The equation must be modified since the signal encoding reduces the value of $R(\tau)$. In order to solve this problem the value of $R(\tau)$ is not divided by N (normalized). As a result equation 2.5 becomes :

$$R^i(\tau) = \sum_{n=1}^{N-\tau-1} x(n)x(n+\tau),\ \tau \in [T1 - T2] \tag{2.13}$$

where :

T1 = the lowest possible value of T_0

T2 = the highest possible value of T_0

Equation 2.13 leads to subharmonic suppression but has an undesirable effect on T_0 determination when strong secondary peaks, which is due to the

formant structure, are present [2]. However the preprocessing stage with the analog filter, signal encoding, and $R(\tau)$ evaluation for values of τ in the range of T_0 will destroy the formant structure.

Voice / unvoice detection

Voice segments are separated from unvoice segments by comparing the value of global maximum of $R(\tau)$, which is $RMAX(\tau_0)$, and $P_{AV}(i)$.

$$RMAX^i(\tau_0) = \max_{\tau}\{R^i(\tau)\},\ \tau \in [T1 - T2] \tag{2.14}$$

$TR(i)$, the threshold value of $P_{AV}(i)$ is evaluated with :

$$TR(i) = k_{uv} \cdot P_{AV}(i) \tag{2.15}$$

Experiments [2] have shown that the value of $k_{uv} = 0.3$ is the optimal value. Based on the value of $TR(i)$ and $RMAX^i(\tau_0)$, segment i^{th} is classified as voiced if $TR(i) \leq RMAX^i(\tau_0)$, otherwise as unvoiced.

Finding peaks in the autocorrelation function

The evaluation of $R(\tau)$ was done in the range of T_0, therefore the fourth and higher harmonics are eliminated. However 2^{nd} (and sometimes 3^{rd}) harmonics are not completely suppressed. This can cause errors in T_0 detection. In order to solve this problem, the following approach and algorithm are proposed.

In [2] [15] it is shown that the peak corresponding to T_0 in autocorrelation function is larger than the subharmonics and harmonic peaks, typically

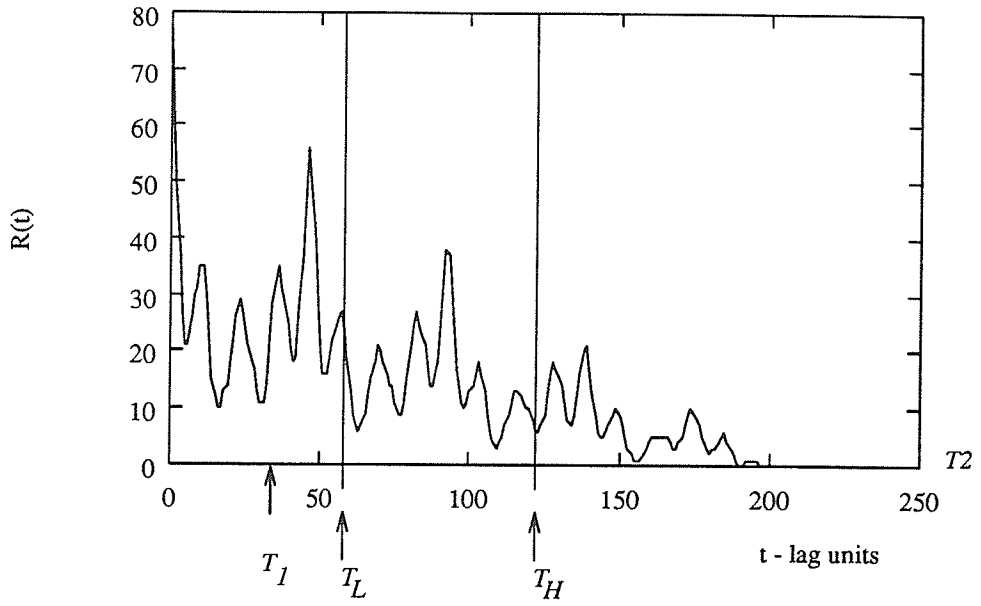


Figure 2.7: Frame division

on the order of four. Based on this, an attempt of T_0 detection is done by dividing the spectral frame into three regions. The division is done with two threshold values, namely T_L and T_H, as shown in figure 2.7. T_L and T_H are chosen to give

$$T_H = 2 \cdot T_L \tag{2.16}$$

- If τ, correspond to the highest R_τ, falls below T_L, then there is a possibility that the correct τ is 2τ or 3τ. Thus three temporary τs (mtau) are prepared for further analysis.

$$
\begin{aligned}
mtau(1) &= \tau \\
mtau(2) &= 2 \cdot \tau \\
mtau(3) &= 3 \cdot \tau
\end{aligned}
\tag{2.17}
$$

- If τ falls between T_L and T_H then

$$
\begin{aligned}
mtau(1) &= \tau \\
mtau(2) &= \tau/2 \\
mtau(3) &= 2 \cdot \tau
\end{aligned}
\tag{2.18}
$$

- If τ falls above T_H then

$$
\begin{aligned}
mtau(1) &= \tau \\
mtau(2) &= \tau/2 \\
mtau(3) &= \tau/3
\end{aligned}
\tag{2.19}
$$

To identify the correct τ, 15 fast autocorrelation functions are calculated, each calculation is done by moving the rectangular window in time-domain by one period. The fast autocorrelation function is basically the standard autocorrelation function, except it is calculated only on $mtau(1)$, $mtau(2)$, and $mtau(3)$ with some tolerance value. From these 15 fast autocorrelation functions, a voting is done to find the most probable value for τ.

2.4 Voice Unvoice Detection

As in pitch detection, voice and unvoice detection is not an easy task. Voice unvoice detection is important in speech analysis. In pitch period detection, voice unvoice detection is useful to reduce the volume of computation since in unvoiced segment there is no pitch period information.

The main task here is to decide as to whether a segment is voiced or unvoiced. To do so, a degree of voicing must be defined. There are several parameters proposed for voicing determination, such as :

- Energy of the signal [15]. It is defined as the short-term root mean square (rms) value of the signal.

$$E_S = \sqrt{\sum_n x^2(n)} \tag{2.20}$$

  Frequently $\log E_s$ is used rather than E_s.

- Average Magnitude [15]

$$\left|\bar{X}\right| = \sum_n |X(n)| \tag{2.21}$$

- Peak-to-peak amplitude [15] [2]

$$\hat{X} = \max[X(n)] - \min[X(n)] \tag{2.22}$$

- Normalized short-term autocorrelation coefficient at unit sample delay [15]

$$r(1) = \frac{R(1)}{R(0)} = \sum_n [X(n) \cdot X(n+1)] \,/\, \sum_n X^2(n) \tag{2.23}$$

where R(d) are the short-term autocorrelation coefficients.

- Ratio between the autocorrelation coefficient at 0, $R(0)$, and at τ [2]

$$C = R(\tau)/R(0) \tag{2.24}$$

where R(d) are the short-term autocorrelation coefficients.

Typically for voiced segment, C is greater than 0.125.

- The ratio of the energy of the differenced signal and the energy of the ordinary signal [15]

$$Q_S = P[X(n+1) - X(n)]/2P[X(n)] \tag{2.25}$$

- The number of zero-crossing [15] [Reddy, 1966]. This is a simple parameter that provides a good measure of voicing. However if used by itself it is not reliable [8].

- Ratio of energy in high-frequency subband and low- frequency subband. This parameter is based on assumption that for voiced signals the energy is concentrated in the low-frequency subband and for voiceless signals the energy is concentrated in the high-frequency. Typically the low-frequency subband is the frequencies below 1 kHz [15] or 2 kHz [20]. The high-frequency subband is the frequencies above 2 kHz [15] or 4 kHz [20].

Most of the methods proposed use a combination of the above parameters.

2.5 Short survey of methods for degree of hoarseness evaluation

Degree of hoarseness (DH) is an index indicating how hoarse the speech of a patient is. Typically it is a measure of the power of the fundamental frequency with its harmonics and the other components, i.e. noise components. One way to calculate DH is to take the Fourier transform of the signal, calculate the power of the fundamental frequency and the harmonics, calculate the power of noise components, and calculate the ratio between those two.

Study [36] has shown that most laryngeal diseases cause an increase in the degree of hoarseness. Several methods for DH evaluation have been proposed, however they can be classified into two groups :

1. DH evaluation by analyzing temporal structure of the signal. The advantage of this method would be a reduction in the volume of computation and lack of errors caused by insufficient spectral resolution. However they have the following disadvantages :

   - small changes in pitch period, which are normally present in voiced speech, can cause a false increase of DH.
   - all changes in the amplitude of pitch period peaks, which are not caused by noisy components, can cause significant increase in DH.
   - It is sensitive to errors in T_0 detection. Such errors can lead to incorrect DH evaluation.

- distortion caused by external noise affects the accuracy of DH evaluation.

2. DH evaluation by analyzing the signal spectrum. The advantages of this method are :

   - distortions caused by amplitude changes of T_0 peaks are minimized
   - the influence of small changes in T_0 is minimized
   - the harmonic structure of the voiced signal could be displayed and visually analyzed.

   The disadvantages of this method are :

   - a large volume of computation for the Fast Fourier Transform (FFT)
   - errors in DH caused by insufficient spectral resolution
   - errors by spectral distortion, caused by pitch- synchronous spectral analysis. This could lead to false DH increase.

From the comparison above, DH analysis using spectral analysis is more accurate. As a result this method is chosen in this thesis.

Chapter 3

Implementation

The implementation of the system was done in two stages. The first stage was done to test the method. This was done on an IBM PC by developing an analog to digital board and several programs specifically written for the IBM PC.

The second stage was done to refine the method and programs used in the system. This was done by rewriting the program in C language, which will allow for the operation of the program on other computers. This was done on a Sun workstation.

One problem encountered with the last step is that the ADC board was designed specifically for an IBM computer. This board will not work on computers other than IBM. Therefore the IBM PC is still needed to digitize the data, which may be transferred to the Sun workstation using file transfer program if the IBM is connected to a network. The other solution is to make an independent ADC device or to use the ADC that comes with the

computers as in Sun SPARCstations or NeXT computers.

The description of the implementation can be divided into three sections. The first section will discuss the methods chosen for the analysis. The next section will give description of the hardware. The last section will give a brief description of the software.

3.1 Methods

The method of analysis can be divided into two stages, namely pitch period detection stage and parameter calculation stage. The result of pitch period detection is a sequence of pitch periods, location of minimas and maximas, the value of minimas and maximas, and the location of *stable zones*. The *stable zone* is defined as a region where at least five consecutive pitch periods were found without interruption. These results are then fed to a program which calculates the parameters of the speech.

3.1.1 Pitch Period Detection

The method used for pitch period detection is based on the improved accuracy of the autocorrelation function, described in section 2.3, and simple peak detection algorithm.

The autocorrelation function is used at the beginning of the pitch period detection to get an accurate pitch period. After the pitch period is found, simple peak detection, described in the previous section , is used to get the

pitch period individually. If the simple peak detection fails, the autocorrelation function is called again. This procedure is repeated until the end of the data is reached.

From the sequence of pitch period, a stable zone is defined. The stable zone is defined as a region which contains at least five consecutive fundamental frequencies with the difference between two consecutive periods being less than 12 %.

3.1.2 Calculation of Speech Parameters

There are several parameters used in the implemented system, namely :

- *average fundamental frequency (f0sr)*,
- *deviation of fundamental frequency (df0)*,
- *degree of hoarseness (dh)*,
- *frequency perturbation quotient (fpq)*,
- *jitter in fundamental frequency (fdlt)*,
- *directional perturbation factor of fundamental frequency (fdpf)*,
- *minima perturbation quotient (a2pq)*,
- *shimmer in minima (mindlt)*,
- *directional perturbation factor of minima (mindpf)*,
- *percentage of sign changes of frequency and minima.*

Average Fundamental Frequency (f0sr)

Average fundamental frequency is calculated by averaging the corresponding frequency of each pitch period on the stable zone. The reason for performing analysis only on the stable zone is that the pitch period of normal subject is relatively constant, or has little variation.

$$f0sr = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tau(i)},\ \tau(i) \in \text{ stable zone} \tag{3.1}$$

Deviation of Fundamental Frequency (df0)

The deviation of fundamental frequency factor is the average of the absolute value of the difference between each fundamental frequency and the average fundamental frequency.

$$df0 = \frac{\sum_{i=1}^{N}|f(i) - f0sr|}{N} \tag{3.2}$$

This parameter shows the ability of the speaker to produce a stable fundamental frequency.

Degree of hoarseness (dh)

The degree of hoarseness is calculated based on the frequency spectrums of the signal on stable zones.

$$dh = \frac{\sum_{m=1}^{k} \sum_{n=1}^{N} |X_m(n)|^2}{\sum_{m=1}^{k} \sum_{n=1}^{N} |W_m(n)|^2} \tag{3.3}$$

where :

k = the number of stable zones

$W_m(n) = n^{th}$ harmonic components on the m^{th} stable zone

$X_m(n) = n^{th}$ non-harmonic components on the m^{th} stable zone

The spectrums are obtained by performing 2048-point Fast Fourier Transform (FFT). Zero padding is applied when it is necessary. This parameter can be thought as a noise to signal ratio. The assumption used here is that pathological speakers generate more "noise" in the speech signal. The "noise" may resulted from the presence of an unstable fundamental frequency.

Frequency perturbation quotient (fpq)

There are several definition of frequency perturbation quotient. Davis (1976) in [16] defines it as standard deviation of f_0 in Hz divided by mean f_0. In this implementation, frequency perturbation quotient is calculated based on the *perturbation quotient (PQ)* [19]. This equation is a modified version of Koike's perturbation quotient [22].

$$PQ = \frac{100}{N-k+1} \sum_{n=1}^{N-k+1} \left| 1 - x(n+m) / \frac{1}{k} \sum_{r=1}^{k} x(n+r-1) \right| \% \tag{3.4}$$

where $m = (k-1)/2$ and k is odd number.

In the implemented system $k = 3$ and $m = 1$ are chosen as suggested by [19]. For frequency perturbation quotient (fpq), fundamental frequency sequence is used in the above equation.

This parameter can be used to measures rapid variations in pitch period and amplitude. It is shown in [19] that fpq and apq are often useful in detecting laryngeal diseases even in their early stage.

Amplitude perturbation quotient

As in the previous section, amplitude perturbation quotient (apq) is also calculated using Koike's equation. Two *apq*s, namely a1pq and a2pq, can be calculated using the sequence of maximas and minimas respectively. However inn the implementation, pitch detection is calculated based on the location of minima, therefore only a2pq (apq for minima) is used in the analysis.

The hypothesis used here is that normal speakers are more capable of producing stable amplitude.

Jitter in fundamental frequency

Jitter in fundamental frequency (fdlt) is defined as the sum of the absolute differences in consecutive data divided by the number of differences [16]. In this particular case the sequence of fundamental frequency is used as the data.

$$fdlt = \frac{\sum_{i=1}^{N} |f_0(i+1) - f_0(i)|}{N-1} \tag{3.5}$$

where $f_0(i)$ is the i^{th} fundamental frequency.

This method is similar to the magnitude of the difference in duration between adjacent periods described in [14] [25] [31], except they visually analyze the frequency distribution of fundamental frequency.

It is proven [14] [25] that this parameter differs for laryngeal and normal speakers. In this thesis analysis is done to compare neurological and normal speakers.

Shimmer in minima (mindlt)

Shimmer is defined similar to jitter. In this particular case the sequence of minima is used as the data.

$$mindlt = \frac{\sum_{i=1}^{N} |Amin(i+1) - Amin(i)|}{N-1} \tag{3.6}$$

where $Amin(i)$ is the i^{th} value of minima.

Directional perturbation factor (dpf)

Directional perturbation factor (dpf) is based on the method described in [14], which is based on Lieberman's method [25]. It is defined as the percentage of changes in algebraic sign (nsgn) between adjacent pitch or amplitude which

the magnitude of the difference is equal or greater than 0.5 msec. In this particular implementation it is defined as the percentage of changes in algebraic sign between adjacent pitches or amplitudes.

$$\text{nsgn} = \begin{cases} \text{nsgn} + 1 & \text{if } |X(i+1) > X(i)| \text{ and } |X(i) < X(i-1)| \\ \text{nsgn} + 1 & \text{if } |X(i+1) < X(i)| \text{ and } |X(i) > X(i-1)| \end{cases} \tag{3.7}$$

$$\text{dpf} = \frac{\text{nsgn}}{N} 100\% \tag{3.8}$$

3.2 Hardware

The hardware of the implemented system consists of a two-channel Analog to Digital Converter (ADC) board, two low-pass filters, and a microphone. The design was done so that the software does not depend too much on the hardware.

The ADC board was designed for the IBM PC in a form of a plug-in board. Beside the ADC board, other parts are designed for general system. Thus implementation in other types of computer is possible as long as there is a similar ADC board available for the computer.

3.2.1 Analog to Digital Converter

As the name implies, the Analog to Digital Converter (ADC) is used to convert analog signal into digital signal. In this particular case the ADC is used to convert analog speech signal into digital speech signal. The quality

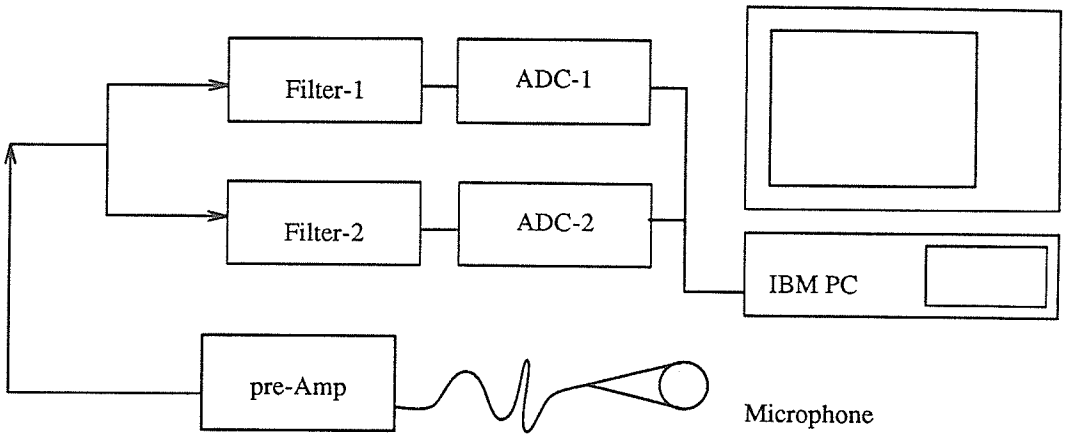


Figure 3.1: Block Diagram of the Implemented System

of the the digital representation depends on the quality, word length, and sampling rate of the ADC.

The sampling rate determines the discretization of the independent variable, i.e. time. According to Nyquist theorem or Shannon theorem, the sampling rate must be greater than twice of the maximum frequency in the signal.

$$f_S \geq 2 \cdot f_{max} \tag{3.9}$$

In the implementation, the signal is band-limited to 700 Hz for the first ADC and 8 kHz for the second ADC. Thus minimal sampling rates are 1400 Hz and 16 kHz for ADC1 and ADC2 respectively. The actual implementation uses 8 kHz sampling rate for ADC1 and 16 kHz for ADC2.

Two ADCs are used in the implementation. ADC1 is connected to low-

Table 3.1: Typical implemented ADC

	number of bits	sampling rate
Gold and Rabiner (1969)		10 kHz
Tucker (1978)	14 bits	10 kHz (speech)
	14 bits	20 kHz (music)
Horii (1979)	16 bits	20 kHz
Hess (1983)	12 bits	
Boyanov (1984)	12 bits	8kHz and 16 kHz

pass filter1 which has the lower band-width (700 Hz). The sampled data is used in pitch period detection to get faster and more accurate results.

ADC2 is connected to the filter with higher bandwidth. The resulting data is used to calculate the degree of hoarseness with Fast Fourier Transform (FFT).

The accuracy of the sampled data depends on the word-length of the ADC. If the ADC has k bits word-length, then the number of possible discrete steps is

$$K = 2^k \tag{3.10}$$

A large number of k will create higher resolution of the sampled data thus generating better representation of the signal. However high resolution ADC is expensive. For speech signal typically a 12-bit ADC is used.

AD 574 A, an ADC chip manufactured by *Analog Devices*, is used in the implementation. It is low-cost and is a 12-bit successive-approximation ADC with a typical 35 us conversion time. It is also equipped with 3-state output buffer for direct interface to an 8 or 16-bit microprocessor bus.

3.2.2 Filter

As shown in figure 2.1, filters can be used to preprocess incoming signals in order to make pitch detection easier and faster. Two low-pass filters with different cut-off frequencies are used.

The first low-pass filter has a 700 Hz cut-off frequency, while the second low-pass filter has 8 kHz cut-off frequency. The output of the first filter is fed into ADC1, which will be used in calculating pitch period. The output of the second filter is fed into ADC2 which will be used in calculating dh using FFT.

The pitch detection scheme produces stable zone marks of the signal sampled with ADC1 (channel 1). These marks are used to indicate the starting and ending region of stable zone in time domain. The same region is also used to calculate dh. However dh is calculated using data sampled with ADC2 (channel 2). Therefore the marks on channel 1 must be mapped to the same location on channel 2. To get the correct location, the delay and phase difference between the two low-pass filters must be minimized.

The filters are implemented as second-order Butterworth low-pass filters. A frequency response analysis of the implemented circuit was simulated with HSPICE, a circuit simulator. The components, such as op-amps, resistors, and capacitors, were entered into a spice data deck. Figure 3.3 and 3.4 show the results of the simulations. Similar response was observed on the actual harware using a spectrum analyzer.

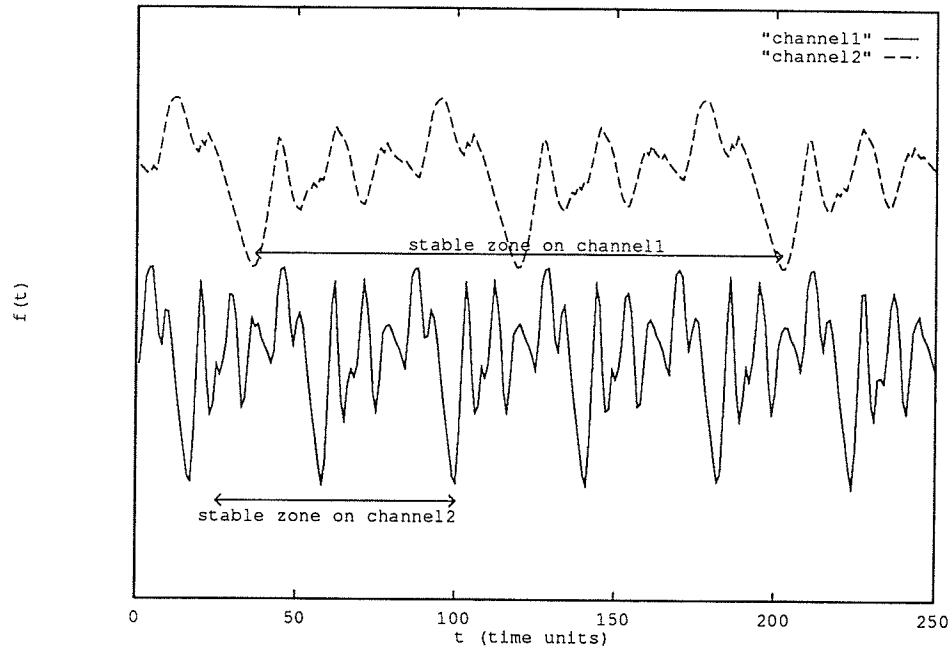


Figure 3.2: Example of stable zones on ch1 and ch2

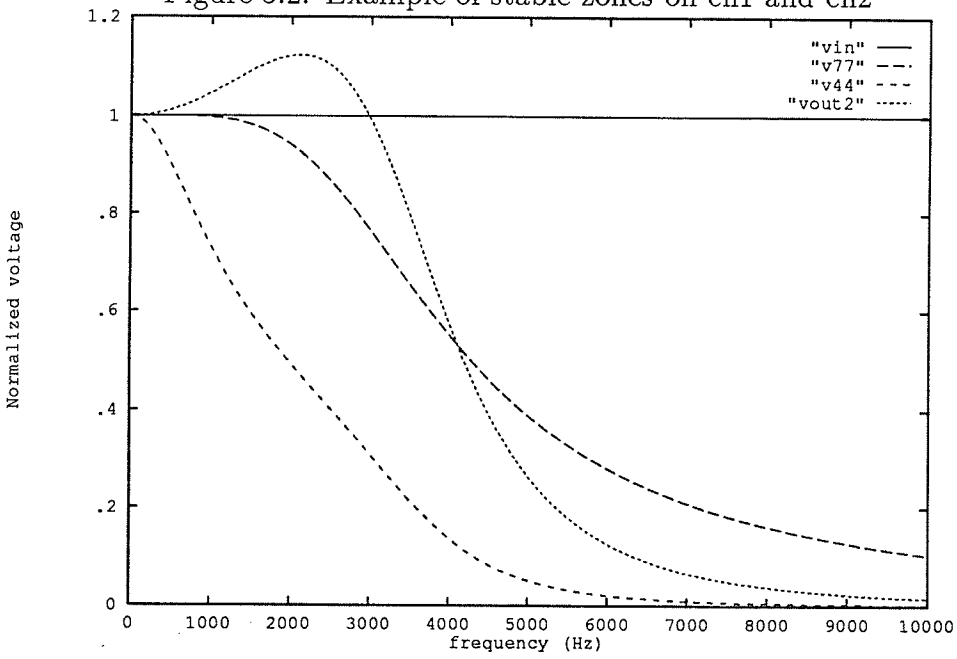


Figure 3.3: Frequency response of filter1

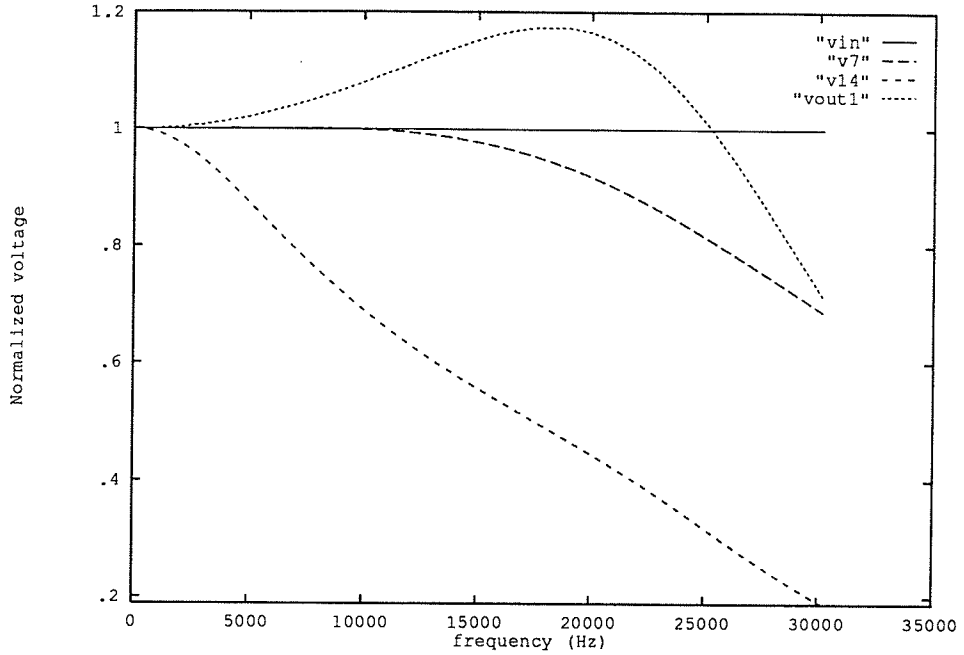


Figure 3.4: Frequency response of filter2

Table 3.2: Suggested cut-off frequency of low-pass filter for pitch detection

Gold and Rabiner (1969)	70-600 Hz
Markel (1972)	800 Hz
Boyanov (1984)	700 Hz
Jovanovic (1986)	700 Hz
Laver (1986), for male	660 Hz
Laver (1986), for female	800 Hz

Suggested cut-off frequencies of the low-pass filter is shown in table 3.2.2.

3.3 Microphone

The quality of the recorded or digitized signal depends on the choice of microphone. In music, most recorded vocals are recorded with a Condenser microphone, which can get expensive, in the order of $ 800.00. The usage

of electret microphone was suggested by [17] since it preserves low-frequency phase information. Another suggestion was using Brűel and Kjaer 4134 microphone which has 20 kHz bandwidth [33].

In the implementation, Syncron S-10 solid state/condenser microphone was used. This microphone has the following characteristic.

- Type: Pressure gradient, condenser
- Frequency Response: 40 - 20,000 Hz $\pm$ 3db
- Directional Characteristic: Cardioid at all frequencies, 20 db front-to-back ratio
- Price: $ 240.00

As mentioned in the Syncron manual, a condenser microphone does not have to move a voice coil as in dynamic microphone, also it is not surrounded by a large magnetic structure as in the ribbon. Therefore it is free to follow the sharpest audio transient.

3.4 Software

As mentioned earlier, the software was done in two stages. The result of the first stage, testing stage, is a collection of programs written in different languages for the IBM PC.

The program that controls the ADC board, digitizes the data and stores it in memory is done in 8088 assembly language. Assembly language is chosen

to achieve a fast and optimized program. The program that displays the signal and saves it in a file on a disk was done in Pascal using Turbo Pascal. The program that calculates speech parameters was done in C language using Microsoft C compiler.

The following problems arised due to the usage of mix languages :

- The resulting program is not portable. For example the program written in 8088 assembly can not be ported to other machines.
- It is difficult for an operator to operate a non-integrated and non-user friendly system. For example, the operator has to run the program to digitized the data and run another program to display or save the data.

As a result, most of the program has been rewritten in C language except the digitizing program which is still in 8088 assembly language. The advantages of C language compared to other high level languages are :

- more portable program
- C compiler is available on most computers. This was done on a Sun workstation.

The final program will let you select a data file and perform analysis on it. The output will be displayed on the screen and saved in files. The users can also graphically view :

- waveform of signal digitized by each ADC

- sequence of pitch period
- sequence of minima an maxima
- spectrum of signal in stable zones

The graphical display depends on the user's environment. Under *Sunview* (Sun's windowing) the graphic outputs can be viewed with *Sunplot*, a public domain plotting program. *Xgraph* will be used to view graphic outputs under Xwindow. Standard unix plot will be used in other environments.

The Sun workstation used to develop the program is accessible through a telephone line. Thus one can use a terminal or personal computer equipped with a modem to run the program from a remote place. This would allow several users sharing the same program and data.

Chapter 4

Application in analyzing pathological voices

Analysis of neurological diseases using voice parameters is becoming an attractive method [6] [7] [11]. However it is still not as popular as in analysis of laryngeal diseases.

This chapter will discuss the application and results of our system in analyzing normal, laryngeal, and neurological voices. In particular the differences between normal speakers and speakers with neurological diseases will be analyzed.

Sustained vowel is used in this analysis because it has the following characteristics :

- a stationary period of the speech is achieved by using a vowel [6].
- During sustained vowel phonation, pitch period T_0 is stable.
- Most laryngeal diseases generate additive and multiplicative noisy com-

ponents, which appear most clearly during vowel phonation.

- Sustained vowel phonation is relatively independent from accent, speaking habits, and native language.
- The formant's value are constant, allowing preliminary formant removal by means of analog filtering.
- Sustained phonation is easy and speakers do not have to be trained before testing.
- appropriate when more or less random perturbations caused by mechanophysiologic conditions of the vocal folds are in question [16].

The selection of the vowel determines the method used in pitch detection. Vowel `/i/` is used in [16], because the performance of their program was excellent for this vowel. In our application, vowel `/a/` is chosen because it has a higher first format.

4.1 Subject and data collection

The samples were obtained in three different ways. The first method is by digitizing the data directly into the computer. Each speaker was told to utter vowel `/a/` for six seconds. The sample is then saved directly into a file on a hard-disk. This method is done for normal speakers, who are Electrical Engineering graduates students at University of Manitoba.

The second way of obtaining the data was by recording the voice of the speakers on a tape. The recording was done at the Neurological Lab, Health Sciences Center. The recorded voice was then played back with the output connected to the input of the filters. Adjustment was made to get optimal amplitude. This was done for speakers with neurological diseases.

The third method was done by sampling the data directly into a computer with a different system. This was done for speakers with laryngeal diseases. The data was sampled with the same format and specification in Bulgaria and sent as a file on a disk.

4.2 Results

Analysis was done on each sample by running the program with the correspondence data file for each sample. The program will generate arrays of pitch periods, values of maxima, value of minima, location of maxima, location of minima, and spectrum of the signal on stable zones. These arrays are stored in files for further analysis. Based on the sequence of pitch periods and minima, other parameters such as $df0, dh, fpq, fdlt, fdpf, a2pq, mindlt$, and $mindpf$ are also calculated and displayed on the screen. These parameters are saved for statistical analysis.

Users are also allowed to view graphical results, such as viewing pitch period sequence (figure 4.1), sequence of maxima and minima (figure 4.2), and spectrum of the signal (figure 4.3).

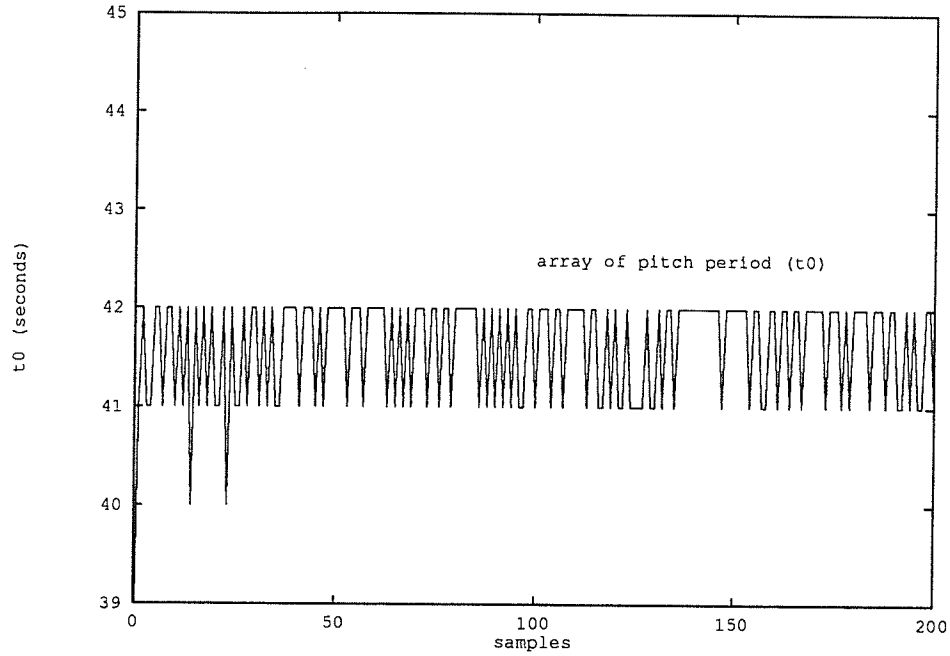


Figure 4.1: Sequence of pitch periods

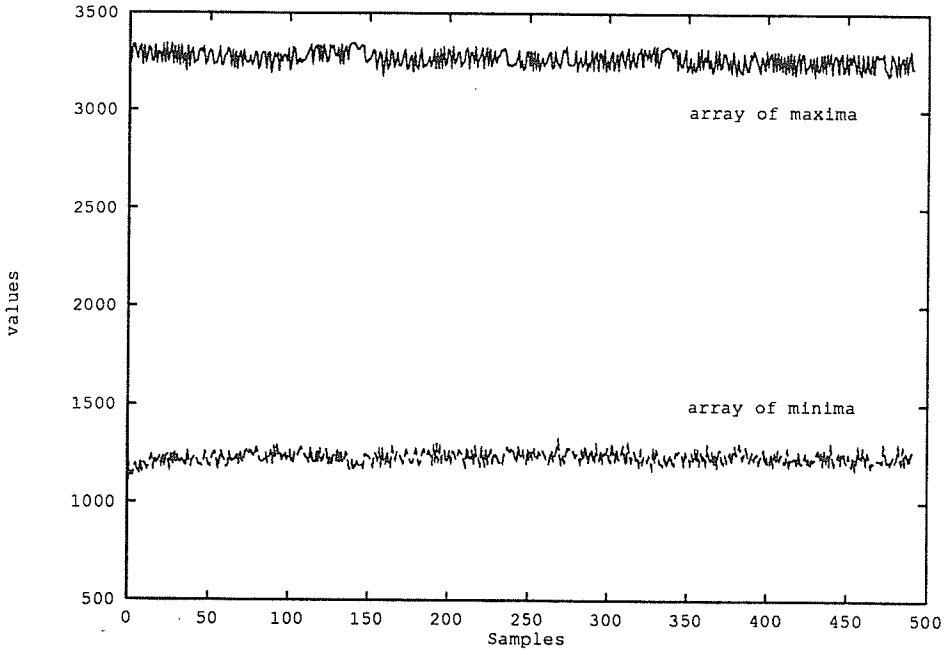


Figure 4.2: Sequence of maxima and minima

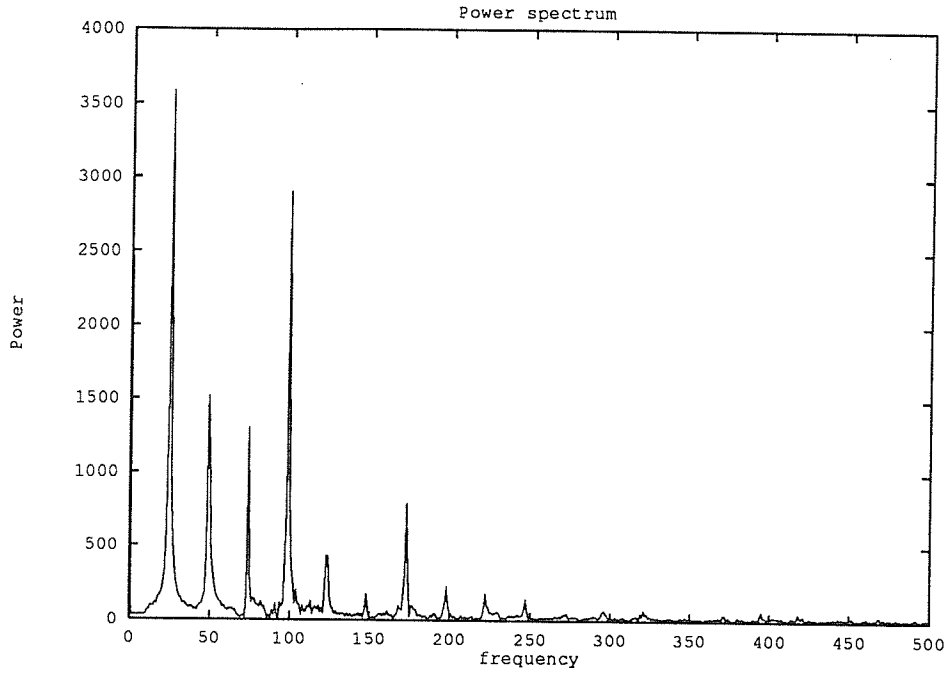


Figure 4.3: Spectrum signal

4.3 Analysis

Some elementary statistical analysis was performed based on the results of the program. The statistical analysis was done by grouping the results into four groups, namely normal speakers (group 0), neurological speakers (group 2), laryngeal speakers (group 3), and pathological speakers (group 1), which is a combination of neurological and laryngeal speakers.

The analysis was done with *SAS*, a statistical package. To begin with, averages of all parameters were calculated using a suggested procedure [29]. The result is shown in appendix B.

Beside fundamental frequency, maxima, and minima, other parameters are calculated from these three parameters. Thus it is reasonable to inves-

tigate the correlation between averages of fundamental frequency, maxima, and minima with other parameters.

4.3.1 Average fundamental frequency and other parameters

Sustained vowel phonation ideally generates one stable value of f_0. However in practice this is not the case. Even in normal speakers, there is always variations in f_0 as shown in figure 4.1. Average f_0 is then calculated and used as a representation of f_0 of the particular speaker.

Several parameters such as $df0$, fpq, dh, $fdpf$, and $fdlt$ are also calculated from the sequence of f_0. Thus it is reasonable to investigate the correlation of these parameters with average of f_0 ($f0sr$).

Deviation in fundamental frequency, $df0$

A visual correlation between $f0sr$ and $df0$ is shown in figure 4.4. A correlation factor r can be used to measure the strength of a relation between two variables. This correlation factor, which is the *Pearson product moment correlation coefficient*, is defined as :

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \tag{4.1}$$

The value of r ranges from -1 to +1. A value of '0' corresponds to a scattered points, '+1' corresponds to a plot of points that fall exactly on an

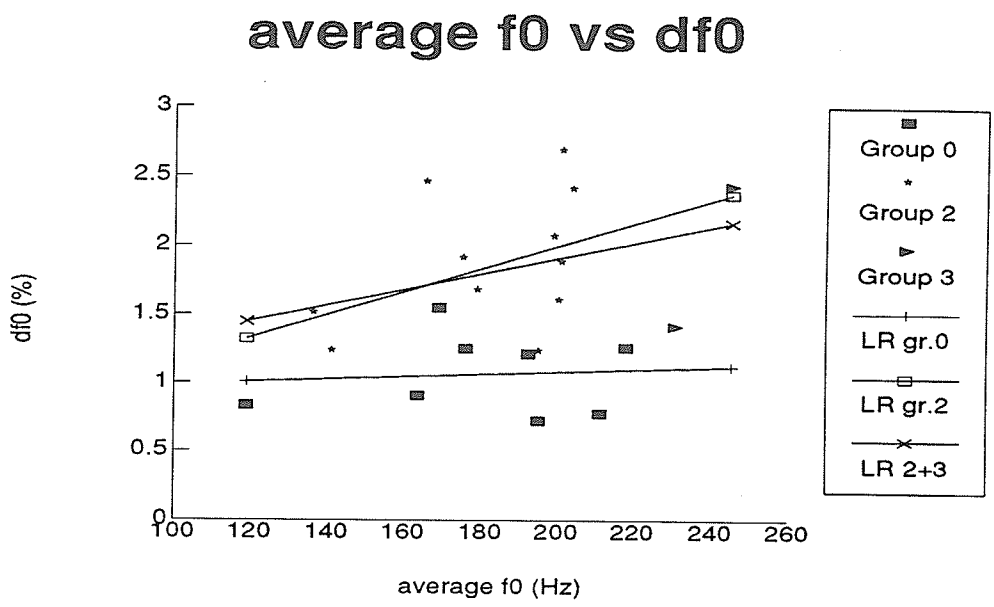


Figure 4.4: Deviation in f0 $df0$ vs average f0

upward straight line, and '-1' corresponds to a plot of points that fall exactly on a downward straight line.

For normal speakers, it is found that r is 0.09591. Since this value is small, the correlation between $df0$ and $f0sr$ is negligible. This result agrees with figure 4.4 which shows data from normal speakers is neither going up nor going down as the average f_0 increases.

A linear regression analysis was done and produced the following results :

VARIABLE	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|
INTERCEP	0.90467510	0.69041905	1.310	0.2380
F0SR	0.000891223	0.003776150	0.236	0.8213

F0SR = average f0

This result shows that the gradient of the line is close to zero (0.000891223) Thus the regression is a straight line almost parallel to x-axis (f_0). The p-

value, 0.8213 gives evidence that the probability of the parameter having a value of zero is high. This supports the fact that $df0$ does not depend on average $f0$.

The same analysis was also done using data from neurological and pathological speakers. The correlation factors r are found to be 0.41708 and 0.35805 for neurological and pathological speakers respectively. This result shows that there is only a small correlation between $df0$ and average $f0$.

Linear regression analysis using data from neurological and pathological speakers produced the following results :

GROUP 2 (NEUROLOGICAL)

VARIABLE	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|
INTERCEP	0.33847450	1.09715874	0.309	0.7647
FOSR	0.008246787	0.005990225	1.377	0.2019

GROUP 1 (PATHOLOGICAL)

VARIABLE	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|
INTERCEP	0.77098128	0.85805491	0.899	0.3882
FOSR	0.005665374	0.004454413	1.272	0.2297

Although the gradients of the resulting regressions are small, 0.008 and 0.006, the probability that the gradients are zero are small, 0.2019 and 0.2297. This suggests that higher f_0 tends to result in higher $df0$.

Other parameters such as frequency perturbation quotient (fpq), degree of hoarseness (dh), jitter in fundamental frequency (fdlt), and directional perturbation factor (fdpf) were also analyzed the same way. Correlation

Table 4.1: Correlation between f0 and other parameters

parameter	normal	neurological	laryngeal
df0	0.09591	0.41708	0.35805
fpq	0.14468	0.77359	0.46652
dh	-0.61752	0.32332	0.53261
fdlt	-0.19186	0.38278	0.50748
fdpf	0.02455	0.52404	0.55265

factors of these parameters and f_0 are shown in table 4.1.

Frequency Perturbation Quotient, fpq

As shown in table 4.1, the correlation between fpq and f_0 for pathological speakers, especially in neurological case, is stronger than for normal speakers.

A quadratic regression was also done using data from neurological speakers and gave the following result :

```
          PARAMETER
VARIABLE  ESTIMATE       PROB > |T|
INTERCEP  6.83279687     0.5036
FOSR     -0.08968551     0.4610
FOSQ      0.000324166    0.3647
FOSQ = FOSR * FOSR
```

Quadratic regression was chosen since it gives better representation of the data, numerically or visually as shown in figure 4.5. This result suggests that higher f_0 will result in higher fpq.

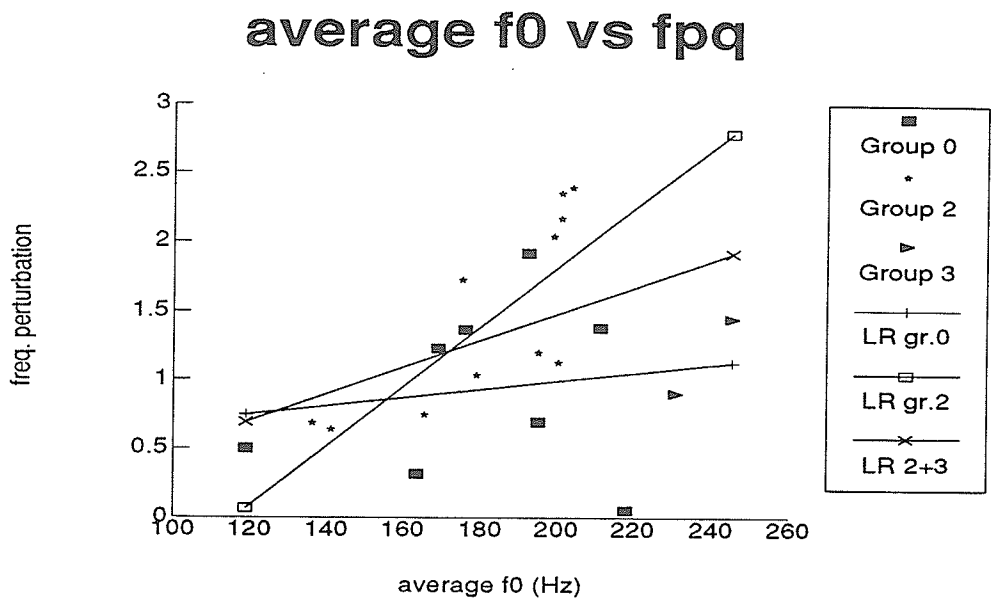


Figure 4.5: Frequency perturbation quotient fpq vs f_0

Degree of hoarseness, dh

Analysis of correlation between dh and f_0 is reasonable since the calculation of dh is based on f_0. The correlation factors in table 4.1 shows high correlation values, especially for normal speakers. Linear regression analysis produced the following results :

VARIABLE	PARAMETER ESTIMATE	PROB > |T|
Normal speakers:		
INTERCEPT	1.08944210	0.0249
F0SR	-0.003854534	0.1028
Neurological speakers:		
INTERCEPT	0.18488496	0.8383
F0SR	0.004924348	0.3321
Pathological speakers:		
INTERCEPT	-2.76213462	0.1964
F0SR	0.02176246	0.0609

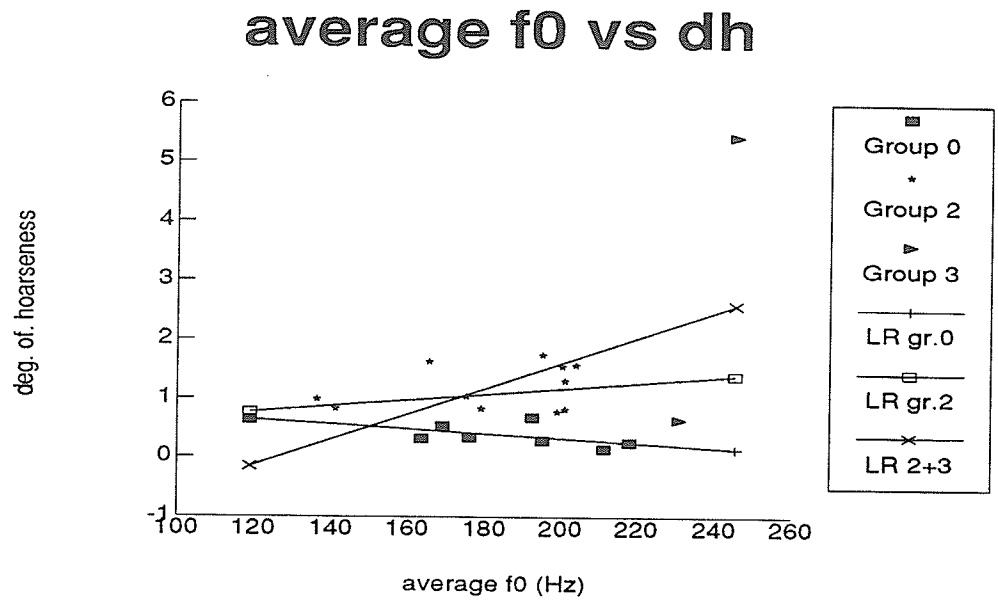


Figure 4.6: Degree of hoarseness dh vs f_0

For normal speakers the resulting regression is a straight line almost parallel to x-axis (f_0). Hence it proves that dh does not depend on f_0. This agrees with visual inspection on figure 4.6. For pathological speakers, the data are distributed above normal data. Linear regression of neurological samples shows a gradient close to zero. Therefore dh does not depend on f_0.

Jitter in fundamental frequency, $fdlt$

Correlation factors between $fdlt$ and f_0, shown in table 4.1, suggest a weak relation between the two parameters. This is reasonable since $fdlt$ is calculated from the differences between consecutive f_0s, not from the value of f_0 itself.

Linear regression analysis was calculated using normal samples only.

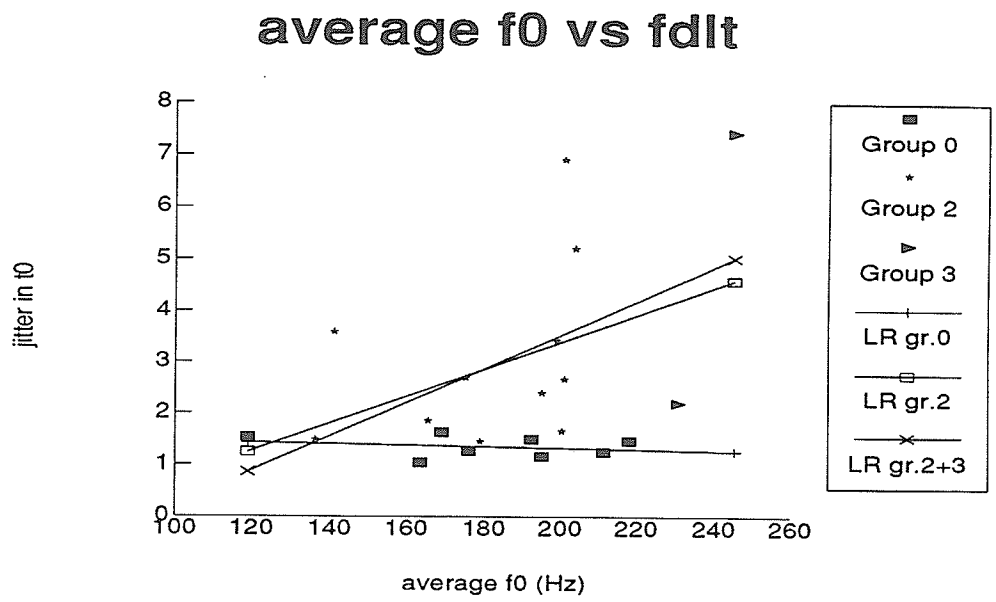


Figure 4.7: Jitter in fundamental frequency $fdlt$ vs f_0

The following result suggests that $fdlt$ does not depend on f_0. Linear regression was not done for neurological and pathological speakers since they are scattered as shown in figure 4.7.

VARIABLE	PARAMETER ESTIMATE	PROB > |T|
NORMAL		
INTERCEP	1.580292	0.0180
FOSR	-0.001283	0.6490

Directional Perturbation Factor of fundamental frequency, $fdpf$

Table 4.1 shows that the correlation factor of $fdpf$ and f_0 for normal speakers is small. This suggests that $fdpf$ does not depend on f_0. Although the correlation factors for pathological speakers are higher, they also show little correlation between the two variables. This result agrees with figure 4.8.

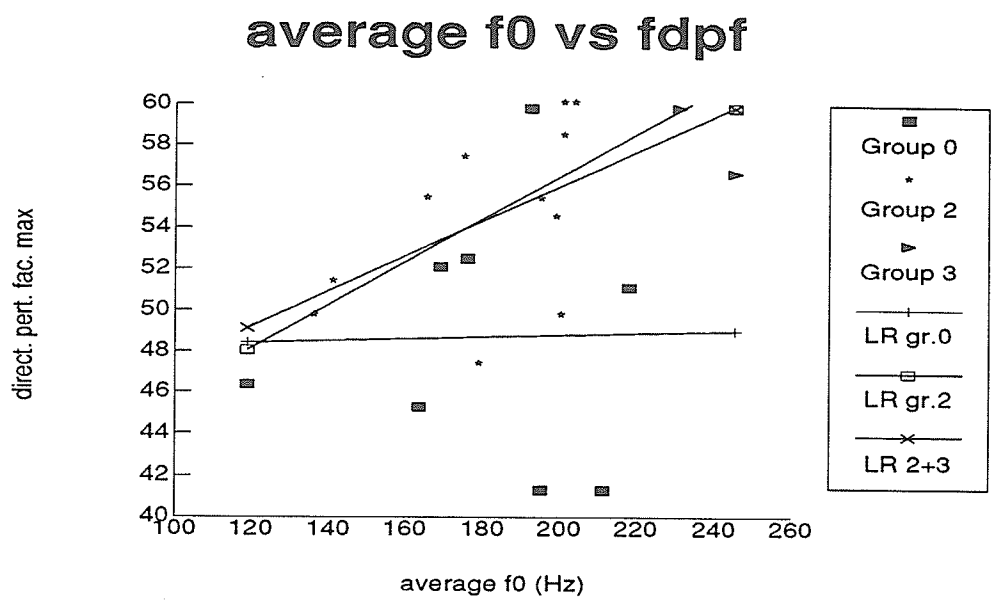


Figure 4.8: Directional Perturbation Factor $fdpf$ vs f_0

4.3.2 Average minima and other parameters

Minima perturbation factor ($a2pq$), shimmer in minima ($mindlt$), and directional perturbation factor of minima ($mindpf$) are calculated from the sequence of minima. Average minima ($minavg$) is also calculated from the same sequence. Therefore it is reasonable to investigate the relation between $minavg$ and the other parameters.

Correlation factors between $minavg$ and other parameters are shown in table 4.2.

The amplitude level depends on several factors, such as the distance between the microphone and the speakers, loudness of the speech, gain of the amplifier or pre-amplifier, quality of the tape if the sample is recorded on a tape before it is digitized by the ADC, and the record level of the tape

Table 4.2: Correlation between *minavg* and other parameters

parameter	normal	neurological	laryngeal
a2pq	-0.23712	0.86429	0.73082
mindlt	-0.77192	-0.57534	-0.49097
mindpf	-0.03760	-0.24603	-0.12973

recorder. These factors can be divided into two categories; namely speaker-dependent factor and non speaker-dependent factor.

The gain of the amplifier or pre-amplifier, quality of the tape, record level of recording, and the distance between the microphone and the speakers are non speaker-dependent factors. The effect of these factors must be minimized since our main interest is only the speaker-dependent factor. Reducing the effects of non speaker-dependent factors can be done by making a fix amplifier gain and record level, using the same type of tapes, and sampling the data in the same environment. Thus noise and other non speaker-dependent factors are integrated in the same way in all data.

In practice, it might be difficult to reduce the effect of non speaker-dependent factors. For example it is difficult to get the same amplitude level for all speakers. One speaker might speak softly, another might scream, while a third might start softly and scream afterward. In this case an operator is needed to set the gain of the amplifier in order to get an optimum amplitude level, which is defined as the level that the peak of the signal is between 80 % and 100 % of the maximum allowable input voltage of the ADC.

There were some cases where in one frame the voice signal has an opti-

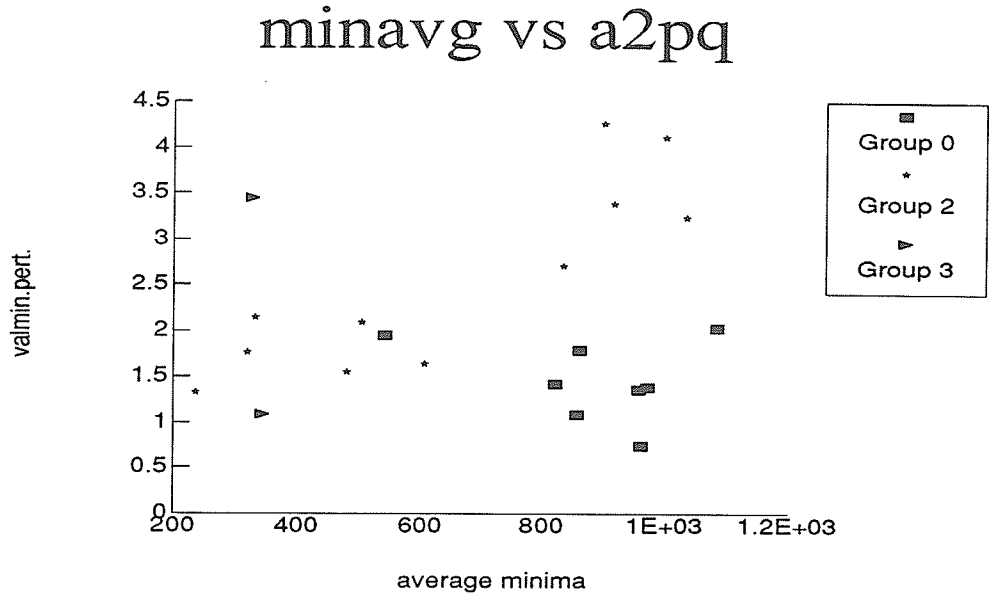


Figure 4.9: Minima perturbation factor *a2pq* vs *minavg*

mum level and in another frame the signal is too low or too high (clipped). This situation, which occurs often in the neurological cases, might be due to the inability of the speaker to hold a sustained vowel.

Minima perturbation factor, *a2pq*

A visual inspection of figure 4.9 shows that normal data is distributed on the lower-right corner. This suggest that normal speakers generate higher minima and lower *a2pq* compared to nuerological speakers.

Regression was done on all samples. For normal speakers quadratic regression was found to give better representation while linear regression was done for neurological data. The regression shows that higher *minavg* tends to produce higher *a2pq*.

```
VARIABLE  PAR ESTIMATE  PROB > |T|

Normal (Quadratic Regression)
INTERCEPT   7.248175    0.0731
MINSQUARE 0.000008506   0.1526
MINAVG     -0.014270    0.1398
R-square = 0.3983 or r = 0.63111

GROUP 2 (Linear)
INTERCEPT   0.501943    0.2636
MINAVG      0.003060    0.0006

GROUP 1 (Linear)
INTERCEPT   0.856192    0.1138
FOSR        0.002660    0.0045
```

Shimmer of minima, *mindlt*

As in previous section, figure 4.10 shows that the data from normal speakers are distributed at the lower-right corner while data from neurological speakers are scattered. The figure also shows higher *mindlt* value for lower *minavg*.

Directional perturbation factor of minima, *mindpf*

As shown in table 4.2, the correlation between *mindpf* and *minavg* are small, which suggests weak relations between the two parameters. This result agrees with figure 4.11 which shows scattered data.

4.3.3 Effect of diseases on each parameters

This section will discuss the effect of diseases on each parameter. The analysis is done by investigating the differences between each parameter for normal,

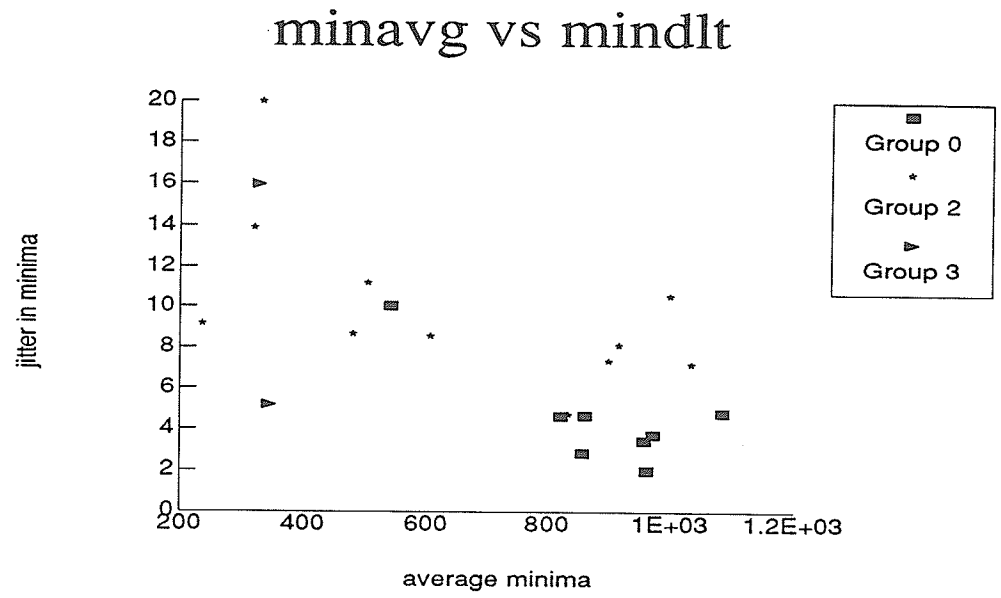


Figure 4.10: Shimmer of minima *mindlt* vs *minavg*

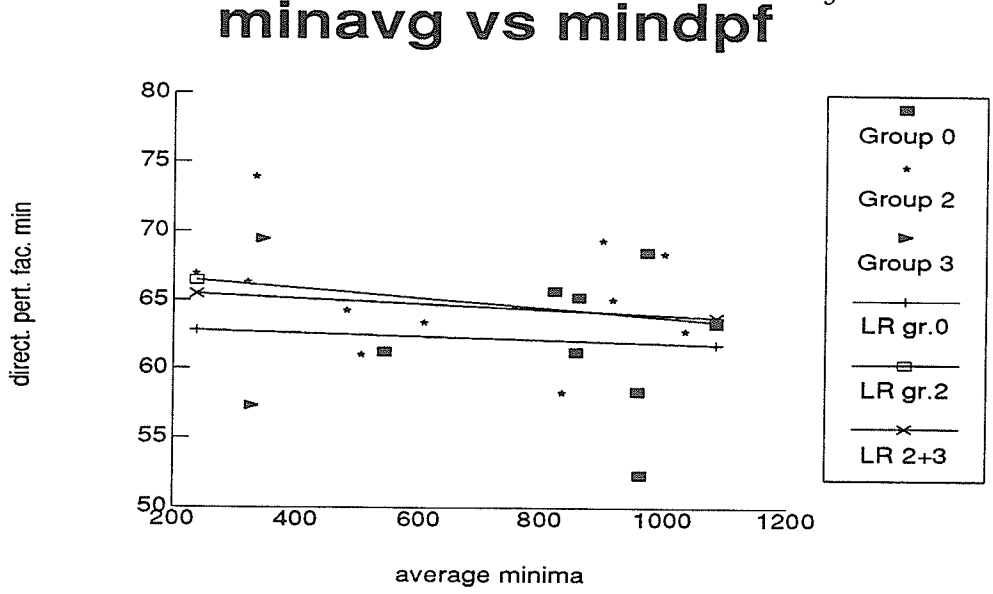


Figure 4.11: Directional perturbation factor of minima *mindpf* vs *minavg*

neurological, and pathological speakers. If a parameter is significantly different for normal and pathological speakers, then it can be used in further applications such as in pattern recognition.

The effect of a disease on each parameter could be determined by investigating the *mean* of each parameter for each group. The *mean* gives a descriptive statistic for each group. A *hypothesis-test* can be used to examine the significance of the difference between the means of the two groups.

To perform a test of hypothesis, two hypotheses are defined. The first hypothesis is the null hypothesis, meaning that the two means are the same. The other hypothesis is the alternative hypothesis, which is that the two means are different.

$$H_o : \mu_A = \mu_B \tag{4.2}$$

and

$$H_a : \mu_A \neq \mu_B \tag{4.3}$$

The samples from one group are independent from the other groups, therefore a *two-sample t-test* is a reasonable choice. There are two possible results of t-tests, namely that the p-value is lower than the predetermined reference probability, or it is not. If the p-value is less than the reference probability, then the result is statistically significant, hence the null hypothesis is rejected. In this particular case where the two groups are independent it can be concluded that the averages are significantly different. If the p-value

Table 4.3: T-test results

parameter	Unequal var.	Equal var.	Probability var. equal	Significance less than 5 %
df0	0.0002	0.0007	0.1761	yes
fpq	0.1377	0.1342	1.0000	no
dh	0.0166	0.0436	0.0000	yes
fdlt	0.0063	0.019	0.0000	yes
fdpf	0.0285	0.0152	0.3767	yes
a2pq	0.0088	0.0231	0.0245	yes
mindlt	0.0021	0.0057	0.1341	yes
mindpf	0.2318	0.2181	0.7776	no

is greater than the reference probability, the averages of the two groups are not significantly different.

Table 4.3 shows the results of t-test for each parameter. In this particular analysis, 5 % significance level is used as a reference.

The t-test result for $df0$ shows a significance level less than 5 %. This means the mean of $df0$ from normal and pathological speakers are significantly different at more than 95 % significance level. Figure 4.4 shows that data from normal and pathological speakers are separated with *means* of 1.0655000 and 1.8491538 for normal and pathological speakers respectively. Notice that $df0$ for pathological speakers is almost twice the value for normal speakers. Hence it is recommended to use $df0$ as a feature to distinguish normal and pathological speakers.

The t-test result for fpq shows a significance level greater than 5 %. A visual inspection on figure 4.5 will show that data from both groups are

mixed together. It is difficult to classify a data with given $f0$ and fpq. Hence this parameter is not relevant in classifying normal or pathological speakers.

Degree of hoarseness, dh, is another good feature to distinguish between normal and pathological speakers. The t-test result shows the difference between the mean of dh of both groups are significantly different at more than 95 % confidence level. As shown in figure 4.6, it can be seen that the value for normal speakers are lower than from pathological speakers.

T-test results for $fdlt$ show that the means of $fdlt$ of normal and pathological speakers are different at more than 95 % confidence level. This suggests that $fdlt$ is a good distinguishing feature between the two groups. However in figure 4.7 it is shown that some data from pathological speakers is located in the normal region. Thus despite the high confidence level that the means of $fdlt$ of the two groups are different, this parameter is not recommended to be used as a distinguishing factor by itself.

Results for $fdpf$ shows that the mean of $fdpf$ of the two groups are different at more than 95 % confidence level. However visual investigation on figure 4.8 shows that the data is mixed and scattered. It is difficult to classify a data given $f0$ and $fdpf$. Hence this method is not recommended to be used as a feature.

T-test for minima perturbation factor, $a2pq$, shows that the mean of $a2pq$ of the two groups are different at less than 5 % significance level. This suggests that $a2pq$ is a good distinguishing feature. As shown in figure 4.9,

this parameter is good especially in the high value of average minima. The data from normal and pathological speakers are separated as the value of *minavg* increases. This suggest that better separation is obtain when the sample has high value of minima.

T-test result for *mindlt* is similar to result for *a2pq*. This result shows that the means of the two groups are significantly different at more than 95 % confidence level. This is true especially for high value of *minavg*. For a lower value of *minavg*, the behaviour of this parameter is not known since there is no available data for this range.

T-test result for *mindpf* shows that the means of *mindpf* are not different at 5 % significance level. This result agrees with a visual inspection on the correlation between *mindpf* and *minavg* as shown in figure 4.11. It is shown that it is difficult to separate the data between the two groups.

4.4 Discussion

From the analysis, it is shown that *df*0, *dh*, *a2pq*, and *mindlt* could be used as features in classifying normal and pathological speakers. Other parameters, namely *fpq*, *fdlt*, *fdpf*, and *mindpf* are not recommended to be used as features in the classification.

The elementary statistical analysis was done to show the function of method and system developed in this thesis. However it was not meant as a complete pattern recognition or classifier. Further analysis might be needed

to perform such a task.

New parameters might be derived and calculated using the system. For example one might be interested in calculating the Walsh transform of the signal instead of using Fourier transform [35].

Chapter 5

VLSI Implementation of Autocorrelation Function

As described in the previous section, the speed of the system suffers from the large amount of calculation needed to calculate the autocorrelation function. Center and infinite clipper reduces the number of computations drastically. However the number of calculations is still dependant upon the speed of the computer used to perform the calculation.

Specially designed hardware might improve the performance of the system. In this chapter, a VLSI implementation of autocorrelation function is presented. The implementation is done in a systolic array structure since it provides a method of parallel or pipe-line processing [3] and the autocorrelation function has a regularity which is suited to be implemented as a systolic array.

Testability plays an important role in designing an integrated circuit (IC). The designer might want to test or pinpoint the error should the IC fail to

Table 5.1: Representation of the encoded value in 2-bit binary

Value of X	Binary Rep.
0	00
1	01
-1	10

work. The regularity of the implementation allows testing method with scan techniques, such as Level Sensitive Scan Design (LSSD) or Scan Path.

5.1 Canonic VLSI Implementation of Autocorrelation Function

Implementation of autocorrelation function as a digital VLSI circuit is relatively straight forward since the signal is encoded with center and infinite clipper into three levels, namely -1,0,and 1. These values can be represented in a 2-bit binary format as shown in table 5.1.

The block diagram of the cannonical implementation is shown in figure 5.1. The operation is carried out serially by shifting the incomming data one by one and adding or subtracting the result register *R(m)*.

The autocorrelation function is defined as :

$$R(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} x(n)x(n+m) \tag{5.1}$$

To calculate a specific lag m, it takes $(N-m)$ operations. Therefore the total number of operations required to calculate autocorrelation function with initial M_i and final lag M_f is :

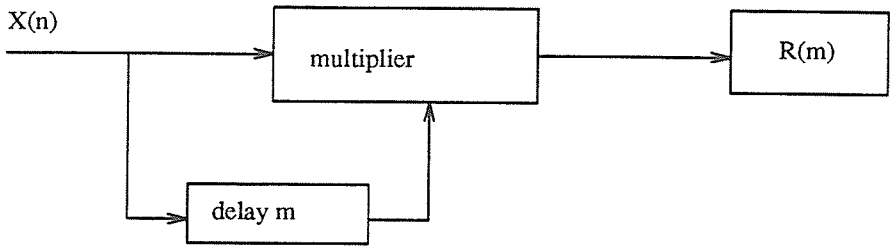


Figure 5.1: Block diagram of canonic implementation

$$\text{N.operation} = \frac{(M_f - M_i + 1)(2N - M_f - M_i)}{2} \tag{5.2}$$

where:

N = the number of data
M_i = initial lag
M_f = final lag

If one operation requires two memory reads, one multiplication, one addition, and one memory store, then on an IBM PC XT with 4.77 MHz it will take $(2 \cdot 10 + 70 + 3 + 10) = 103$ clock cycles or 2.16 msec. For $N = 1024, M_i = 40$, and $M_f = 200$ the number of operation required is 145544 or 3 seconds on the IBM. For a 90 K file this would take 4.5 minutes.

5.2 Implementation as a Systolic Array

Systolic array is an alternative choice of a regular design [23]. Since auto-correlation function has a regular structure, implementation as systolic array is an attractive implementation. Figure 5.2 shows the block diagram of this

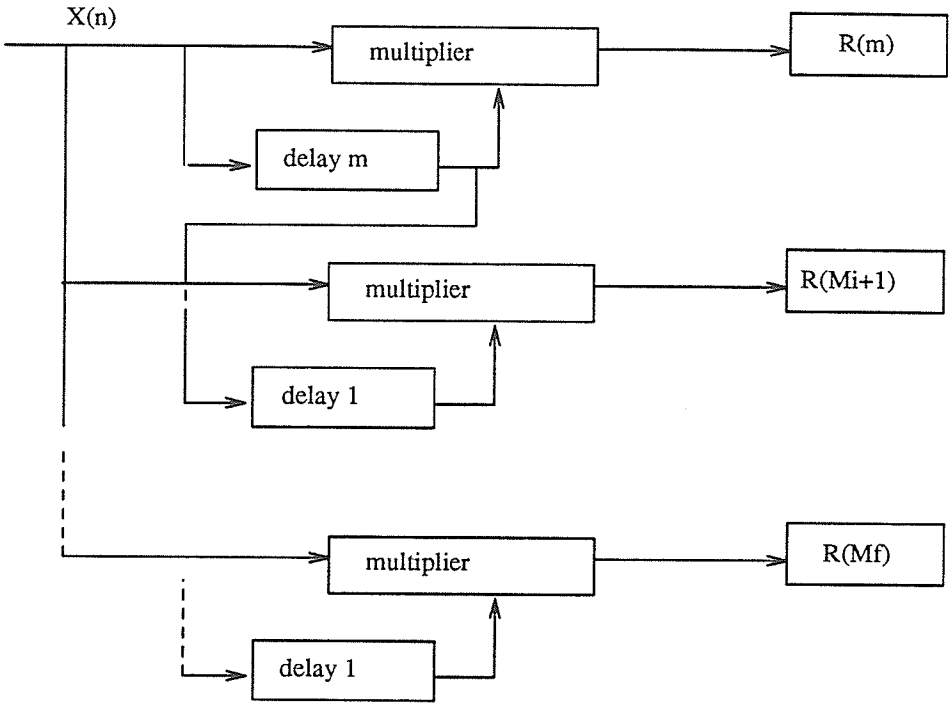


Figure 5.2: Block diagram of systolic array autocorrelation function chip

approach.

As shown in figure 5.2, this approach is similar to the canonic implementation, except that the results for all lags (M_i through M_f) are available at the same time. This is done by replicating the processing block. The number of these blocks depends on the initial and final lag. With this approach, the number of operations required is only N, and it does not depend on the choice of initial and final lag.

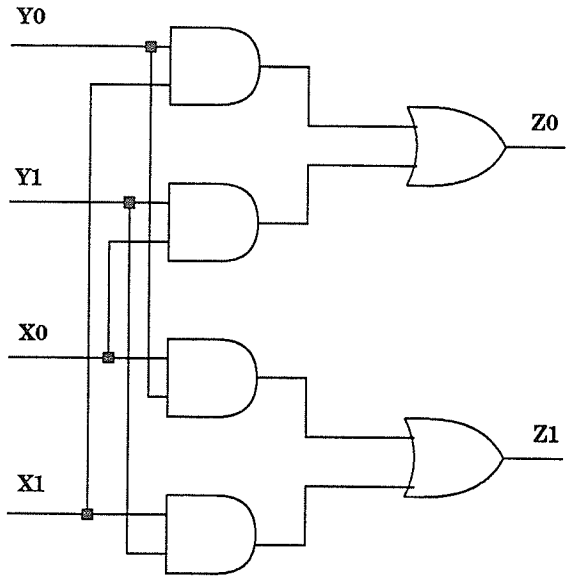
The trade-off of this implementation is an increase in the number of components which means an increase in chip area. However since the com-

Table 5.2: Truth table of the multiplier

	y1y0			
x1x0	00	01	11	10
00	00	00	00	00
01	00	01	xx	10
11	00	xx	xx	xx
10	00	10	xx	01

plexity of the circuit is not high, and the structure is regular, the designing process can be simplified.

The general description of the circuit is as follows. The incoming signal, encoded in -1, 0, and +1, is fed into a delay block and multiplier. The delay blocks are implemented with D-flipflops. The incoming signal is then multiplied with the delayed signal using a 2-bit multiplier, which is implemented as a combinatorial circuit as shown in figure 5.3. The output of the multiplication is accumulated in result register R, which is implemented as up-down counters. The counter will count up if the output of the multiplier is 1, or the LSB is 1, and will count down if it is -1, or the MSB is 1. The output of result register R is then multiplexed to get one parallel output of a particular lag.



Truth Table :

	Y1Y0			
X1X0	00	01	11	10
00	00	00	00	00
01	00	01	xx	10
11	00	xx	xx	xx
10	00	10	xx	01

Figure 5.3: Multiplier

5.3 Implementation in VLSI using Cadence Design System

Implementation in VLSI was done by designing the chip using a *Cadence Design System.* It allows the IC designer to enter the design in a schematic level or the exact physical layout. Physical layout can be generated automatically if the design is entered in the schematic level. Cadence also allows hierarchical design, which means a library can be built in schematic level based on some standard cells.

SILOSII and *APLSIM* were used as circuit simulators. The simulation was done to check the design before it is fabricated. Most of the simulation was done with *SILOSII* since at that time the standard cells did

not have *SPICE* representations which are required by *APLSIM*.

Some parts of the circuit were tested with *APLSIM* by extracting the transistors from the actual physical layout. However this simulation was limited by the size of the circuit *APLSIM* can handle in a non-hierarchical environment [28].

5.4 Circuit Structures

The highest level of the schematic, shown in figure 5.4, is almost identical to the block diagram shown in figure 5.2. The *autocorrelation with delay 1 pulse* contains *multiplier*, *delay 1 pulse*, and *R(M)* as in figure 5.2.

The symbols in this schematic represent another circuit or schematic representation. For example the *delay 5 pulses* is constructed of 10 D-flip flops, shown in figure 5.5. By using this top down design methodology, modification can be done easier. For example if the exact layout of the *AND* gate must be modified, or if the up-down counter needs modification, the final schematic representation will be the same, the modification is done only on the schematic level. However if the layout representation of one standard cell is changed then placement and routing must be done again.

On the first block, the input signal, *y0* (LSB) and *y1* (MSB), is connected to a delay line which will delay the signal 5 clock cycles. The input and the output are connected to the first block of autocorrelation function. The output of the autocorrelation function is the number of matching found so

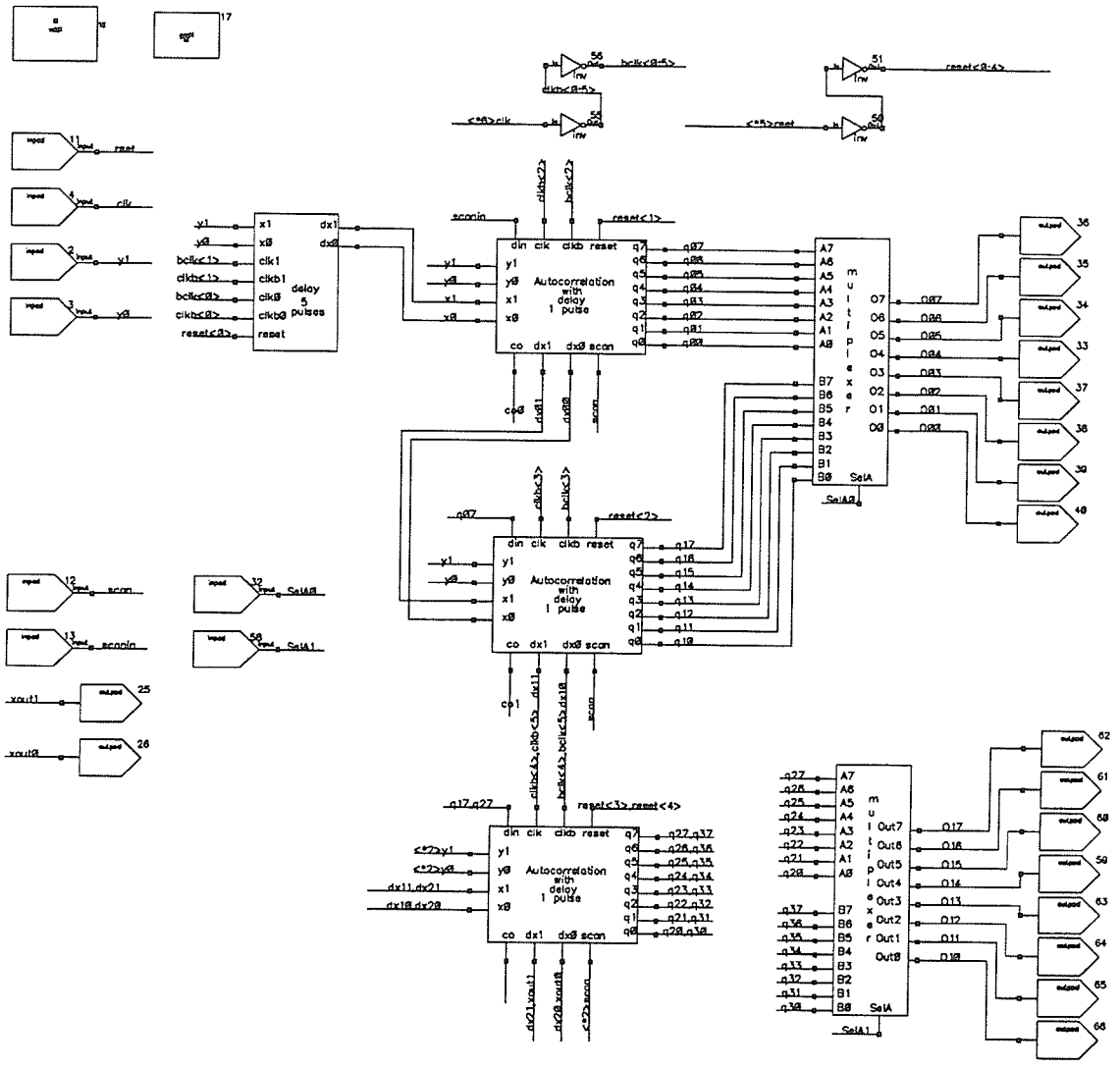


Figure 5.4: The top level of the autocorrelation chip

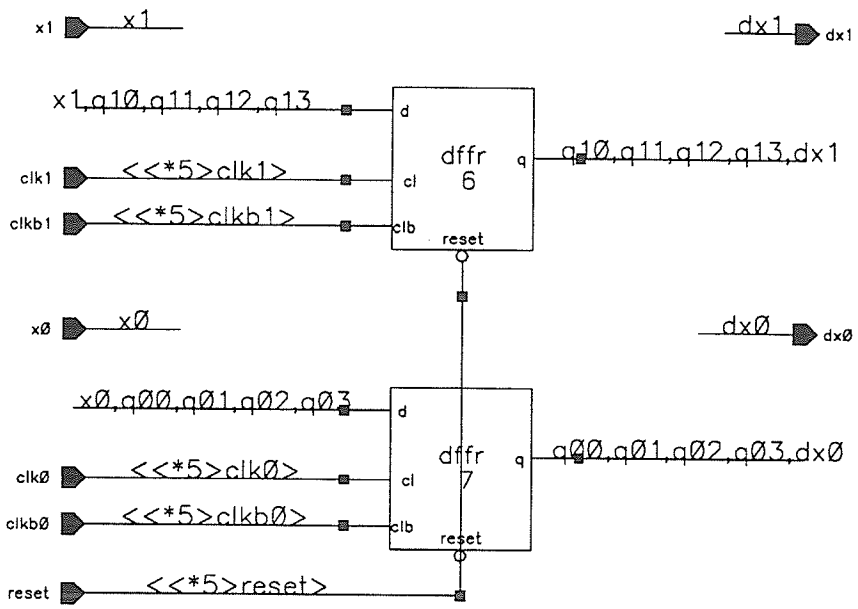


Figure 5.5: Delay 5 pulses circuit

far for the lag equal to 5. The output of this autocorrelation function is then multiplexed with the output of the other block, which in this case has a lag equal to 6.

In the actual implementation the counter, which will be used to count the number of matchings, is implemented as eight-bit counter. Therefore the number of matching must be between -127 and +127. Input signal must be 2 bits as described in table 5.1, i.e. after it goes through the center and infinite clipper. The length of the input is chosen so that the number of matching does not exceeding 127, otherwise the result will be incorrect.

The initial lag was chosen as five clocks, using the *delay 5 pulses* block. To get different inital lag, the circuit must be modified internally or by adding a delay line outside the chip. Five pulses was chosen to make debugging using short input possible.

5.5 Discussion

The final result shows that for a large number of lags the actual physical layout requires a large area so that they can not be implemented in one chip. This problem is solved by parting the design into several chips.

Furtunately this solution is not difficult since the next chip is actually exactly the same as the first chip. The difference is that in the first chip the *delay 5 pulses* is used to get the initial lag, which is 5 pulses, whereas in the next chips the *delay 5 pulses* is by-passed. The connection can be done by

connecting the outputs of the *delay 5 pulses* block; namely *dly1* and *dly0*; to the *x0* and *x1* of the first autocorrelation block. This connection can be done from the pin outside the chip.

Connection to the next chip is done by connecting *dx1* and *dx0* of the first chip to *x1* and *x0* of the second chip. This cascading can be repeated until the final lag is reached.

The number of the autocorrelation block represents the number of autocorrelation functions that will be evaluated. The outputs of the autocorrelation blocks are multiplexed so that only the outputs of one particular *lag* or one particular autocorrelation block will be observed.

5.6 Testability

Testability plays an important role in designing an integrated circuit. The regularity of this kind of implementation allows testing method with scan techniques, such as Level Sensitive Scan Design (LSSD) or Scan Path. The circuit then becomes testable.

Several testing approaches have been developed to make a circuit testable [34]. Scan Path is an attractive method for testing the ACF chip since it has a considerable number of latches. By configuring the latches into a Scan Path or LSSD scheme or chaining the latches into one or several long chains, the circuit becomes more observable. Testing techniques, such as Built-In Logic Block Observation (BILBO), Linear Feedback Shift Register (LFSR), Cellu-

lar Automata (CA) then can be implemented.

In this implementation, all flip-flops are connected in one big chain. By doing this, the value of all the flip-flops can be shifted out. This is a simple implementation of scan techniques. The scan is done by pulling the "scan" pin high, and the data will be scanned out, or scanned in by inputting the data into "data in" or "scan in" pin.

Chapter 6

Conclusion

As the price of personal computers or workstations drops and the performance improves, personal computer applications will continue to expand. In this thesis an application of the computer in signal processing and medical is presented. In particular the computer is used to process speech signal in order to find some parameters that can be used to determine the condition of a speaker.

Pitch period detection is one of the most important analysis in speech signal processing. It is also one of the most difficult tasks. A method of pitch detection using simple time domain and modified autocorrelation is presented. The method was tested and it works well.

A system for analysis of voice of speakers with neurological diseases was developed. Some hardware, namely ADC board and low-pass filters were developed and added to an IBM PC XT in order to sample the speech signals.

Several speech parameters were calculated from the sampled signals. While some parameters do not show a significant difference between normal and neurological speakers, others show a significant difference. These parameters are deviation in fundamental frequency ($df0$), degree of hoarseness (dh), minima perturbation factor ($a2pq$), and shimmer of minima ($mindlt$).

Speed was the major concern in the implemented system which suffers from a large volume of computation. As the price of DSP chips and custom VLSI goes down, the hardware may become useful in future application. As a start, the autocorrelation function was implemented using *Cadence System Design*. This tool allows hierarchical and schematic level, which is useful in reducing the design time.

Future Work

Some new personal computers or workstations, such as NeXT and Sun Sparcstation, are equipped with DSPs chip and/or ADC boards. This opens the possibility for a researcher, clinician, or doctor to sample and analyze the speech of a patient directly on his or her computer. Although current ADCs on these machines are of low quality and do not meet our standard, it is becoming a new standard in new computers. Future personal computers might have even better ADCs.

Although DSP is not required in our implementation, it certainly will help the performance of the system. This will reduce the time required to

calculate the autocorrelation function.

Some new parameters, such as Walsh spectrum of the speech signal, might be calculated and observed. These new parameters might provide better separation of normal and neurological speakers.

As the verbal behaviour of neurological speakers is better understood, a complete pattern recognition system might be developed based on method described in this thesis. This system will be a useful tool for clinicians and doctors in determining the condition of a patient.

Bibliography

[1] Boyanov, B., "Methods for analysis the speech of patients with laryngeal diseases and people under stress", *Ph.D. Thesis*, 1984, (in Bulgarian).

[2] Boyanov, B. "Pitch detection by means of autocorrelation function", *Automatica and Computer Systems*, 21, 1986, pp. 38-43, (in Bulgarian).

[3] Bridges, G.E., Pries, W., McLeod, R.D., Yunik, M., Gulak, P.G., and Card, H.C., "Dual Systolic Architecture for VLSI Design Signal Processing Systems", *IEEE Transaction on Computers*, vol. C-35, No. 10, pp. 916-923, October 1986.

[4] Irwin, J.V., Marge, M., ed., "Neurological aspects of language disorders in children", *Principles of Childhood Language Disabilities*, Appleton-Century-Crofts, Education Division, Meredith Corporation, New York, 1972.

[5] Dubnowski, J., Shafer, R., Rabiner, L., "Real-Time Digital Hardware Pitch Detector", *Transactions on Acoustic, Speech, and Signal Processing*, vol. ASSP-24, No. 1, Feb. 1976, pp. 2-8.

[6] Gath, I., Yair, E., "Comparative Evaluation of Several Pitch Process Models in the Detection of Vocal Tremor", *IEEE Transactions on Biomedical Engineering*, vol. BME-34, No. 7, July 1987, pp. 532-538.

[7] Gath, I., Yair, E., "Analysis of vocal tract parameters in parkinsonian speech", *J. Acoust. Soc. Am.*, 84, 5, November 1988, pp. 1628-1634.

[8] Georgiou, J.V., "A Parallel Pipeline Computer Architecture for Speech Processing", *UMI Research Press*, Ann Arbor, Michigan, 1984.

[9] Gold, B., Rabiner, L. "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain", *J. Acoust. Soc. Am.*, vol 46, 1969, pp. 442-448.

[10] Goldstein, J.L., "An optimum processor theory for the central formation of the pitch of complex tone", *The Journal of the Acoustical Society of America*, vol. 54, No. 6, 1973, pp. 1496-1516.

[11] Gottschalk, L.A., Eckardt, M., Pautler, C.P., Wolf, R.J., and Terman, S.A., "Cognitive Impairment Scales Derived from Verbal Samples", *Comprehensive Psychiatry*, vol. 24, No. 1, January/February, 1983, pp. 6-19.

[12] Gubrynowicz, R., "A Fuzzy Approach of Pitch Period Analysis for Evaluation of Functional Status of Larynx Source", *Journal of Phonetics*, 14, 1986, pp. 525-530.

[13] Hadjitodorov, S., Boyanov, B., Rahardjo, B., "Recognition of Isolated Words in Bulgarian by Means of HMM", *Proc. of IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing*, Victoria, B.C., June 1989, pp. 216 - 217.

[14] Hecker, M.H.L., Kreul, E.J., "Description of the Speech of Patients with Cancer of the Vocal Folds. Part I: Measures of Fundamental Frequency", *J. Acoust. Soc. Am.*, vol. 49, No. 4, 1971, pp. 1275-1282.

[15] Hess, W., "Pitch Determination of Speech Signals : Algorithm and Devices", *Springer-Verlag*, Berlin, 1983.

[16] Horii, Y., "Fundamental Frequency Perturbation Observed in Sustained Phonation", *Journal of Speech and Hearing Research*, 22, March 1979, pp. 5-19.

[17] Howard, D.M., "Peak-picking fundamental period estimation for hearing prostheses", *J. Acoust. Soc. Am.*, vol. 3, September 1986, pp. 902-910.

[18] Jovanovic, G.S., "A New Algorithm for Speech Fundamental Frequency Estimation", *IEEE Transaction on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, No. 3, June 1986, pp. 626-630.

[19] Kasuya, H., Masabuchi, K., Ebihara, S., and Yoshida, H., "Preliminary experiments on voice screening", *Journal of Phonetics*, 14, 1986, pp. 463-468.

[20] Knorr, S.G., "Reliable Voiced/Unvoiced Decision", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, No. 3, June 1979, pp. 263-267.

[21] Koike, Y., "Vowel Amplitude Modulations in Patients with Laryngela Diseases", *J. Acoust. Soc. Am.*, vol. 45, No. 4, August 1968, pp. 839-844.

[22] Koike, Y., "Application of some accoustic measures for evaluation of laryngean dysfunction", *Studia Phonologica*, VII, 1973, pp. 17-23.

[23] Kung, H.T., "Why Systolic Architecture? ", *Computer (IEEE)*, vol. 1, No. 15, pp. 37-46, January 1982.

[24] Laver, J., Hiller, S., MacKenzie, J., and Rooney, E., " An acoustic screening system for the detection of laryngeal pathology", *Journal of Phonetics*, 14, 1986, pp.517-524.

[25] Lieberman, P. "Perturbation in Vocal Pitch", *J. Acoust. Soc. Am*, vol. 33, No. 5, May 1961, pp. 597-603.

[26] Ludlow, C.L., Bassich, C.J., Connor, N.P., and Coulter, D.C., "Phonatory characteristics of vocal fold tremor", *Journal of Phonetics*, 14, 1986, pp. 509-515.

[27] Markel, J.D., "The SIFT Algorithm for Fundamental Frequency Estimation", *IEEE Transaction on Audio and Electroacoustics*, vol. AU-20, No. 5, December 1972, pp. 367-377.

[28] Rahardjo, B., "Implementation of the Autocorrelation Function in a Testable Systolic Array", *Project Report for VLSI Design Methodology*, Electrical and Computer Engineering University of Manitoba, October 1989.

[29] Schlotzhauer, S.D., Littell, R.C., "SAS System for Elementary Statistical Analysis", SAS institute Inc., Cary, NC USA, 1987.

[30] Skinner, P.H., Shelton, R.L., ed., "Speech, Language and Hearing : Normal Processes and Disorders", John Wiley & Sons, Inc., 1978.

[31] Smith, W.R., Lieberman, P., "Computer Diagnosis of Laryngeal Lesion", *Computers and Biomedical Research*, 2, 1969, pp. 291-303.

[32] Sondhi, M. "New methods of pitch extraction", *IEEE Tr. Audio and Electroacoustic*, AU-16, 1968, pp. 262-266.

[33] Tucker, W.H., Bates, R.H., "A Pitch Estimation Algorithm for Speech and Music", *IEEE Transaction on Acoustics, Speech, and Signal Processing*, vol. ASSP-26, No. 6, December 1978, pp. 597-604.

[34] Williams, T.W., "VLSI Testing", *Computer (IEEE)*, vol. 10, No. 17, October 1984, pp. 126-136.

[35] Yunik. M. Boyanov, B., MacDonald, R., Rahardjo, B., "Spectral Analysis of Vowels and Musical Sounds by Means of the Fast Walsh Trans-

form", *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, June 1-2, 1989, pp. 206-207.

[36] Yunik, M., Boyanov, B., "Methods for Evaluation of the Noise-to-Harmonic-Component Rations in Pathologic and Normal Voices", *Acustica*, vol. 70, No. 1, January 1990, pp.89-91.

Appendix A

Raw data

The following is a description of the raw data. Name that ends with "0" is from a normal speaker, "2" is from neurological speaker, and "3" is from laryngeal speaker.

Table A.1: Description of speakers

Name	disease
flher2	hydrocephalus
flhrn2	hydrocephalus
hiper2	alzheimer and hydrocephalus
hyper2	alzheimer and hydrocephalus
jofat2	hydrocephalus
palac2	normal pressure hydrocephalus
palak2	normal pressure hydrocephalus
rooru2	alzheimer
schum2	hydrocephalus at young age
sicum2	hydrocephalus at young age
wisan2	idiopatic hydrocephalus

Table A.2: Raw data 1

Name	f0sr	df0	fpq	minavg	a2pq
burah0	192.51	1.215812	1.914362	819.84	1.422382
felex0	175.91	1.251609	1.364899	970.3	1.385834
felix0	163.43	0.908288	0.317916	542.67	1.950793
jadia0	118.9	0.833531	0.503033	860.13	1.784244
norml0	195.13	0.726269	0.693414	955.22	1.361687
tafik0	217.84	1.263707	0.047807	856.32	1.084254
tufik0	168.9	1.541563	1.22478	1085.14	2.02559
washa0	211.	0.781884	1.37295	960.57	0.748991
flher2	201.	1.834957	2.120975	332.66	2.072873
flhrn2	200.47	1.562617	1.07775	236.95	1.258117
hiper2	195.25	1.95544	1.149894	480.18	1.47439
hyper2	203.76	2.361424	2.344201	606.45	1.575336
jofat2	179.11	1.63416	0.987044	915.92	3.311933
palac2	136.28	1.471267	0.641293	833.88	2.643964
palak2	141.31	1.190995	0.597177	898.51	4.194314
rooru2	175.36	1.866866	1.674377	999.2	4.037556
schum2	201.05	2.642883	2.303004	505.28	2.016344
sicum2	199.04	2.021521	1.985757	330.26	1.693771
wisan2	165.45	2.417578	0.702811	1033.82	3.163351
lipav3	245.41	2.422256	1.444645	326.33	3.439832
majan3	230.52	1.416128	0.9023	343.7	1.085832

Table A.3: Raw data 2

Name	dh	fdlt	fdpf	mindlt	mindpf
burah0	.681345	1.49	59.8	4.62	65.71
felex0	.341629	1.26	52.5	3.77	68.5
felix0	.323084	1.03	45.28	10.04	61.26
jadia0	.631873	1.53	46.38	4.68	65.22
norml0	.273421	1.15	41.29	3.46	58.38
tafik0	.242202	1.47	51.12	2.84	61.22
tufik0	.525404	1.63	52.10	4.85	63.32
washa0	.132649	1.23	41.31	1.98	52.34
flher2	.71646	2.54	62.28	19.7	73.46
flhrn2	1.443649	1.52	49.53	8.8	66.46
hiper2	1.634409	2.29	55.16	8.35	63.79
hyper2	1.455609	5.08	61.04	8.24	62.96
jofat2	.732237	1.33	47.11	7.85	64.62
palac2	0.874761	1.37	49.47	4.45	57.89
palak2	0.707296	3.46	51.11	7.07	68.89
rooru2	.927996	2.56	57.2	10.26	67.9
schum2	1.196439	6.77	58.25	10.87	60.57
sicum2	.669619	3.3	54.27	13.55	65.85
wisan2	1.515498	1.74	55.19	6.95	62.34
lipav3	5.4215	7.41	56.67	15.96	57.33
majan3	.638058	2.19	60.52	5.21	69.43

Appendix B

Average of all parameters

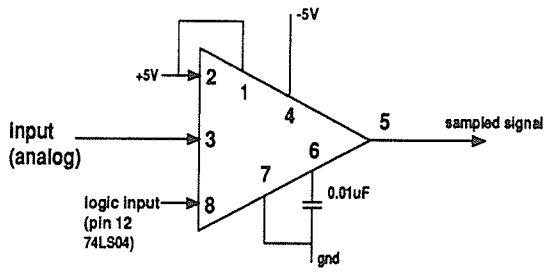
Table B.1: Parameter average for each group

Parameter	group0	group2	group3
f0sr	180.4525	181.643636	237.965
df0	1.065333	1.905428	1.919192
fpq	0.929895	1.416753	1.173472
minavg	881.27375	652.100909	335.015
a2pq	1.470472	2.494723	2.262832
dh	0.393951	1.079452	3.029779
fdlt	1.34875	2.905455	4.8
fdpf	48.7225	54.600909	58.595
mindlt	4.53	9.644545	10.585
mindpf	61.99375	64.975455	63.38

Appendix C

Hardware

C.1 Sample and hold



SAMPLE AND HOLD (LF398A)

1. V+	5. OUTPUT
2. OFFSET	6. HOLD TIME
3. INPUT	7. GND
4. V-	8. LOGIC INPUT

Figure C.1: Sample and hold

C.2 Low-pass filter

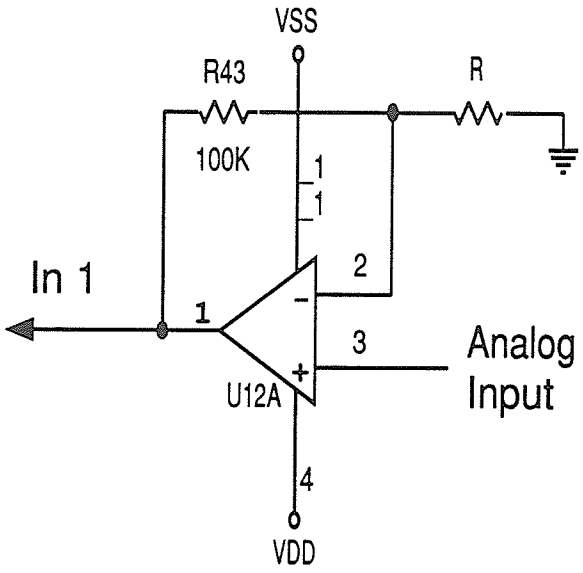


Figure C.2: Pre-ampifier

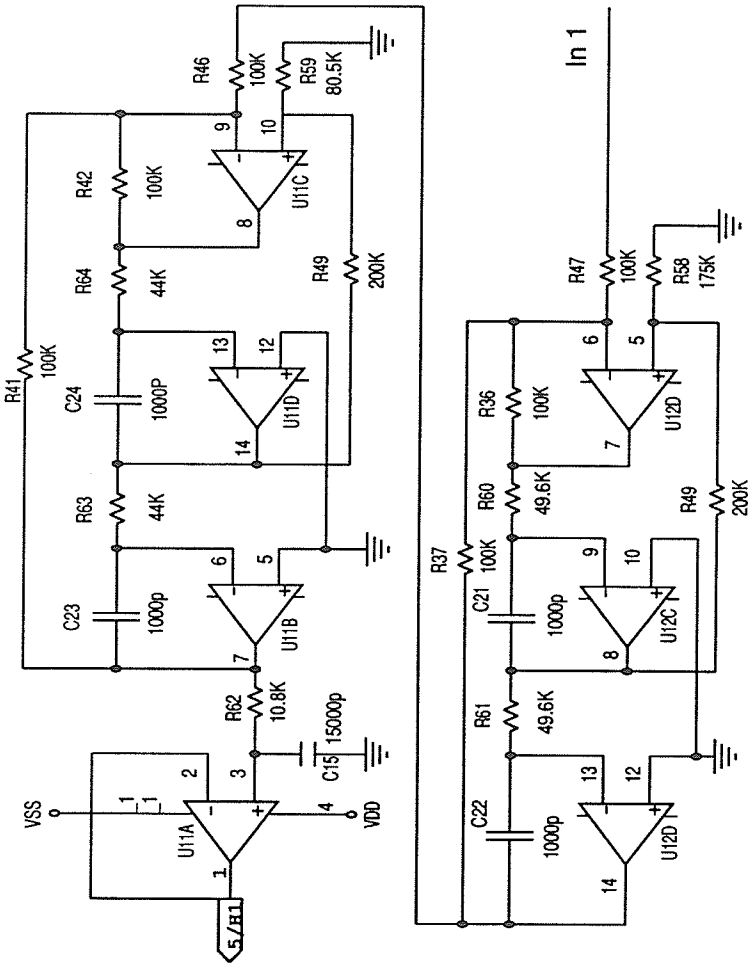


Figure C.3: Low-pass filter 1

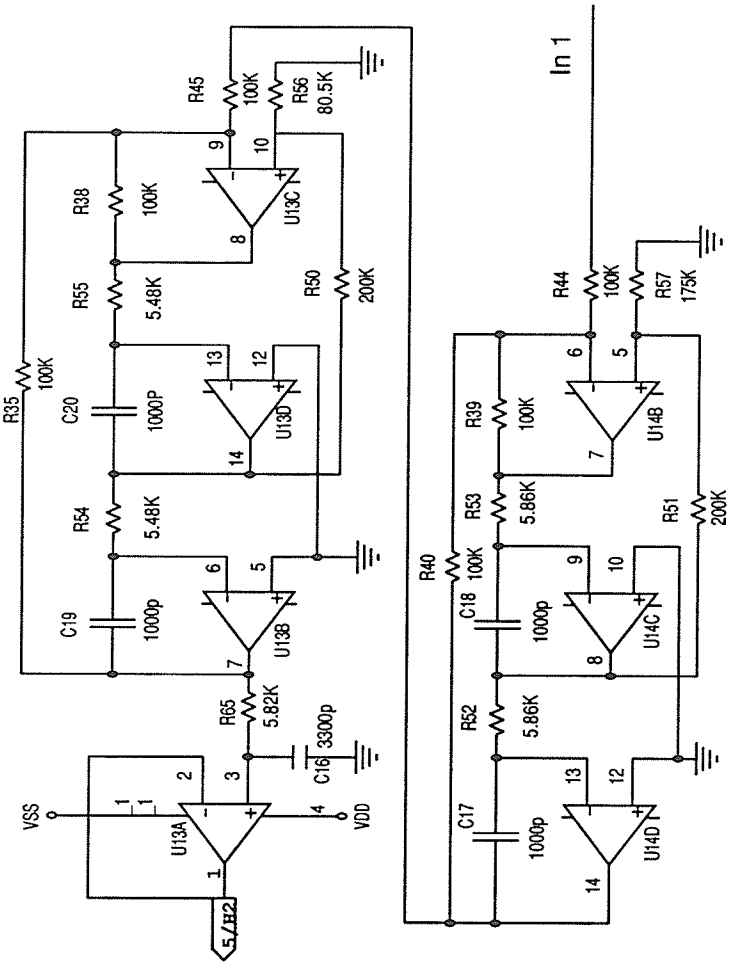
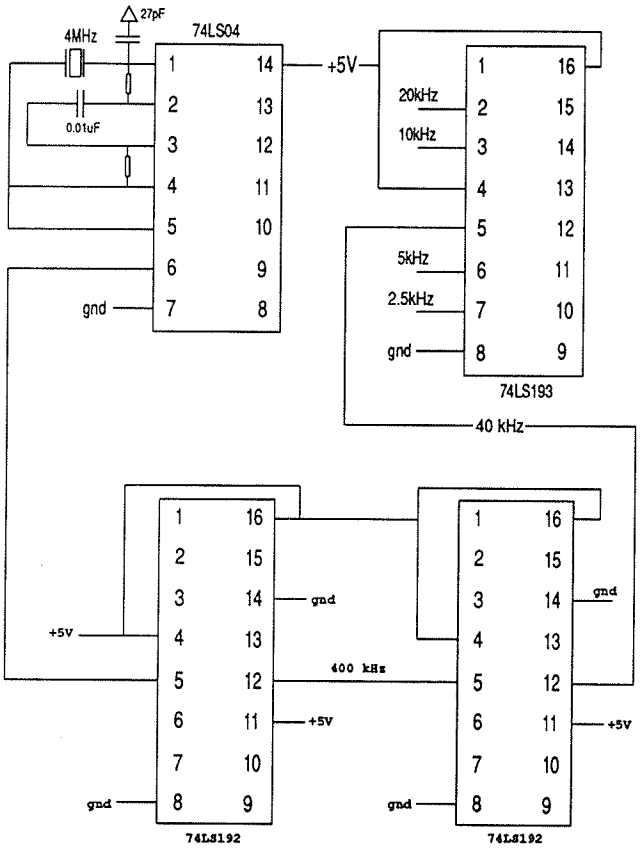


Figure C.4: Low-pass filter 2

C.3 Clock



Sampling Clock

Figure C.5: Sampling clock

C.4 Analog to Digital Converter