Remitted
B

Posted
626

GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF RESEARCH ADMINISTRATION

RESEARCH PROJECT INITIATION

Date: April 24, 1973

Project Title: Research Initiation - Pitch Encoding in Speech Signals

Project No: E-21-627

Principal Investigator: Dr. T. P. Barnwell, III

Sponsor: National Science Foundation

Agreement Period: From April 1, 1973 Until March 31, 1975*
*18 month budget period plus 6 months for submission of required reports, etc.

Type Agreement:
Grant GK-37451

Amount:
$17,000 NSF Funds (E-21-627)
4,844 - GIT Contrib. (E-21-320)
$21,844 - Total

Reports Required:
Annual Letter Technical; Final Report

Sponsor Contact Person (s):

Administrative Matters
Mr. Wilbur W. Bolton, Jr.
Grants Officer
NSF
Washington, D. C. 20550

Technical Matters
Dr. M. S. Ghausi
Electrical Sciences & Analysis Section
Division of Engineering
NSF
Washington, D. C. 20550
Phone: (202) 632-5881

Assigned to: School of Electrical Engineering

COPIES TO:

Principal Investigator	Library
School Director	Rich Electronic Computer Center
Dean of the College	Photographic Laboratory
Director, Research Administration	Project File
Director, Financial Affairs (2)	
Security-Reports-Property Office	
Patent Coordinator	Other

RA—3 (6—71)

GEORGIA INSTITUTE OF TECHNOLOGY

OFFICE OF RESEARCH ADMINISTRATION

RESEARCH PROJECT TERMINATION

Date: May 12, 1975

Project Title: Research Initiation - Pitch Encoding in Speech Signals

Project No: E-21-627

Principal Investigator: Dr. T. P. Barnwell

Sponsor: National Science Foundation

Effective Termination Date: March 31, 1975 (18 month budget period plus 6 months for submission of required reports, etc.)

Clearance of Accounting Charges: March 31, 1975

Grant/Contract Closeout Actions Remaining:

None

Assigned to School of Electrical Engineering

COPIES TO:

- Principal Investigator
- School Director
- Dean of the College
- Director of Research Administration
- Office of Financial Affairs (2)
- Security - Reports - Property Office
- Patent and Inventions Coordinator
- Library, Technical Reports Section
- Computer Sciences
- Photographic Laboratory
- Terminated Project File No. _____
- Other _____

RA-4 (5/70)

E21-627

YEARLY LETTER REPORT TO NSF

Pitch Encoding In Speech

Introduction

This is a progress report on the NSF Research Initiaties Grant No. GK-37451, "Pitch Encoding in Speech." The progress report is as of April 1, 1974, and represents about half of the total effort.

The original project was divided into two parts: a "pitch detector" development portion; and a pitch information encoding study. All of the work thus far has been on the first task.

It was originally proposed that the work be done utilizing a Honeywell H-316 computer. However, since that time, a larger NOVA 820 computer has become available. Hence, work has progressed on this new facility.

Hardware Development

The first step in development of a speech research facility was the addition of specific hardware to the computer facility. These hardware devices included a double buffered D/A (16 hits), an 8 channel A/D, a programable interval timer (for sampling), four 10 bit D/A's (for refresh graphics), a line printer interface, a plotter interface, and a control interface for two analogue tape drives. These new items are used in conjunction with a mini-computer facility which already contains 24k of 16 bit memory, two 1.2 million word moving head disks, a cassette tape drive, and a Textronics 4010 graphics terminal. This whole hardware environment has been developed to give the user maximum flexibility and interaction at high data rates.

Software Development

Software development has been directed toward both general purpose speech and signal processing support and toward the precise goals of the project. The general analysis and support programs include:

PROGRAM	PURPOSE
HEAR	Sampler
SAY	Playback
ADPCM	Adaptive differential PCM
LPC	LPC transmitter
LPR	LPC receiver
SPCANA	Spectrum Analysis
LOOK	General Graphics
PLOT3D	3-D Plotter
HMVC	Homomorphic Vocoder
FILTER	Digital Filter
DFDP	Digital Filter Design Program

The development of these programs included extensive subroutine library development particularly in the areas of interactive graphics and speech (signal) processing.

Pitch Testing

In addition to the general purpose software, a great deal of task oriented software was developed. In the pitch area, a two program package for the testing of pitch detectors were developed. The first, called PTGTC was an interactive pitch painter which utilized the speech waveform, the autocorrelation function, and the cepstrum to set speech pitch periods. The second, called PCHECK, compare the ouput of pitch detectors under test to the hand painted pitch.

Pitch Detectors

In the original proposal, it was suggested that a cepstrum type detector

might be used for pitch contour extraction. Initial testing of that scheme indicated, however, that such a detector had local inaccuracies and used a great deal of computer time. Several alternate schemes are currently under study. These include an autocorrelation detector based on the hard limited speech signal, a minimum difference detector, and several zero crossing detectors. It is hoped that one of these will prove acceptable for doing the pitch extraction necessary in the rest of the study.

E-21-627

PITCH ENCODING IN SPEECH

Final Report

National Science Foundation Grant GK37451

by

T. P. Barnwell III

School of Electrical Engineering
Georgia Institute of Technology
March, 1975

ACKNOWLEDGEMENTS

The major funding for this project came from the National Science Foundation through Research Initiation Grant GK37451.

The School of Electrical Engineering at Georgia Tech provided assistance in the form of matching funds, graduate assistants, and extensive experimental facilities without which the work could not have been completed. Particular participants are listed in the section on personnel. Special thanks go to Dr. J. H. Schlag, Dr. A. M. Bush, Dr. J. E. Brown, and Dr. C. R. Patisaul.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	iv
LIST OF ILLUSTRATIONS	v
SUMMARY	vi
STATEMENT OF POTENTIAL APPLICATIONS TO ENGINEERING AND TECHNOLOGY	vii
DESCRIPTION OF THE RESEARCH AND RESULTS	1
PROJECT PERSONNEL AND THEIR CONTRIBUTIONS	26
PAPERS ACCEPTED OR IN PREPARATION	27
BIBLIOGRAPHY	28

LIST OF TABLES

Table		Page
1	NOVA 820 Speech Facility	4
2	NOVA 820 Graphics/Analysis Subprograms	8
3	NOVA 820 General Subsystems	10
4	Speech Processor Programs	11
5	Hand Painting Pitch Program Parameters	13
6	Structure of the Noun Phrases used in the Pitch Experiment	24

LIST OF ILLUSTRATIONS

Figure		Page
1	Layout of NOVA 820 Facility used in the Speech Algorithm Development and Testing	3
2	A Typical Pitch Contour	14
3	Zero Crossing Pitch Detector	17
4	Illustration of Operation of the Basic Algorithm for the Hard-Limited Autocorrelation Pitch Detector . . .	19
5	Illustration of Operation of the Basic Algorithm for the Minimum Difference Pitch Detector	20
6	A Block Diagram of the Proposed Adaptive Cepstrum Pitch Detector	21
7	Multi-Band Pitch Detector	22

SUMMARY

The work was divided into three phases.

Phase I was concerned with development of an interactive experimentation and simulation facility base on a Nova 820 minicomputer. This work included the development of a modular interfacing packaging scheme, construction of several special purpose interfaces, and the development of extensive graphics and speech simulation software.

Phase II concerned the development and testing of several pitch tracking algorithms. Algorithms based on the windowed autocorrelation function, the hard limited autocorrelation function, the cepstrum, the minimum difference function, and the inverse filtered speech waveform were investigated. Also several zero counting detectors and a "multiband" configuration were tried. The "multiband" was found most suitable for the experimental work.

Phase III investigated the way in which segmental and supersegmental information is encoded in the speech pitch contour. Two subjects, one male and one female, recorded phrases. The pitch contours for these utterances were then extracted. Analysis is still in progress attempting to relate the features of pitch contours to the syntactic and phonemic features of the utterances.

STATEMENT OF POTENTIAL APPLICATIONS TO ENGINEERING AND TECHNOLOGY

Understanding the encoding of the segmental and supersegmental features of speech in the pitch contour is useful in several areas of communications. First, it can be used directly in the segmentation and parsing problem in speech recognition. Second, it is essential in very low bit rate (less than 1000 bps) speech compression techniques. Third, it forms the basis for rules for controlling the pitch in computer generated speech systems.

The pitch detectors studied have direct application in numerous speech compression techniques, including Linear Predictive Coders, Delta Modulators, Adaptive Differential Pulse Code Modulators, etc.

DESCRIPTION OF THE RESEARCH AND RESULTS

1. Introduction

The overriding purpose of this study was to examine the relationship between the segmental and supersegmental features of speech and the resulting fundamental frequency contours. It is well known that the proper control of the frequency of the glottal excitation (hereafter called F_o) is central to the production of natrual and intelligible synthetic speech. This is because not only is F_o effected by many characteristics of the utterance, but also F_o has been shown to be a major cue in the perception of many features cf continuous speech. In particular, variations in F_o contours have been related to phonemic context (14), stress (3-4,9), emphasis (9), position in the utterance (14), juncture (3), speaker attitudes (intonation) (5-6), and structure (5-6). Further studies (7,11) have shown that when the cues supplied by the F_o contour and other acoustic attributes (phonemic duration, intensity, etc.) conflict, F_o is the more powerful cue in perception. This implies that not only could a well formed F_o contour add much to the intelligibility and naturalness of synthetic speech, but also that a poorly formed F_o contour could be highly detrimental to speech quality.

In structuring a study of fundamental frequency there are many pragmatic problems to be solved. First, an experimental environment must be created in which the relevant parameters may be easily measured and tested. Second, since the fundamental frequency of human speech is the quantity being studied, some technique for the extraction of pitch from recorded speech is a necessity. The development of such "pitch extractors" is an interesting

subject in its own right and is one of the major subjects of this study. Such algorithms are of interest not only for studies of pitch contours, but also as components in numerous speech compression systems.

This study can be divided into three parts. The first part consisted of developing a minicomputer based digital signal processing facility in which to do the experimental work. The second consisted of studying several pitch extraction algorithms and critiqueing them for use in this study and in conjunction with speech compression systems. The last part was devoted to studying the information content of the pitch contours in a carefully designed phrase set. This portion of the study is still in progress.

2. The Development of the Experimental Facility

In the original proposal it was intended that most of the experimental work be done on a Honeywell H-316 minicomputer. However, following the receipt of this grant, the School of Electrical Engineering of Georgia Tech committed funds for a larger and faster minicomputer system. The basic hardware of this new facility consists of a Nova 820 processor with 24K of 800 nsec memory, two 1.2 million word Diablo moving head disc drives, and a Tektronix 4010 graphics terminal. The first portion of this project was devoted to developing this basic system into a highly interactive and easily usable digital signal and speech processing facility.

2.1 The Hardware Development

Figure 1 shows a diagram of the Nova 820 system and Table 1 gives a list of the hardware items. The system bulk storage consists of the two previously mentioned Diablo moving head discs and a Nova cassette tape drive, with a capacity of 50K words per tape. The cassette is used primarily for backup.

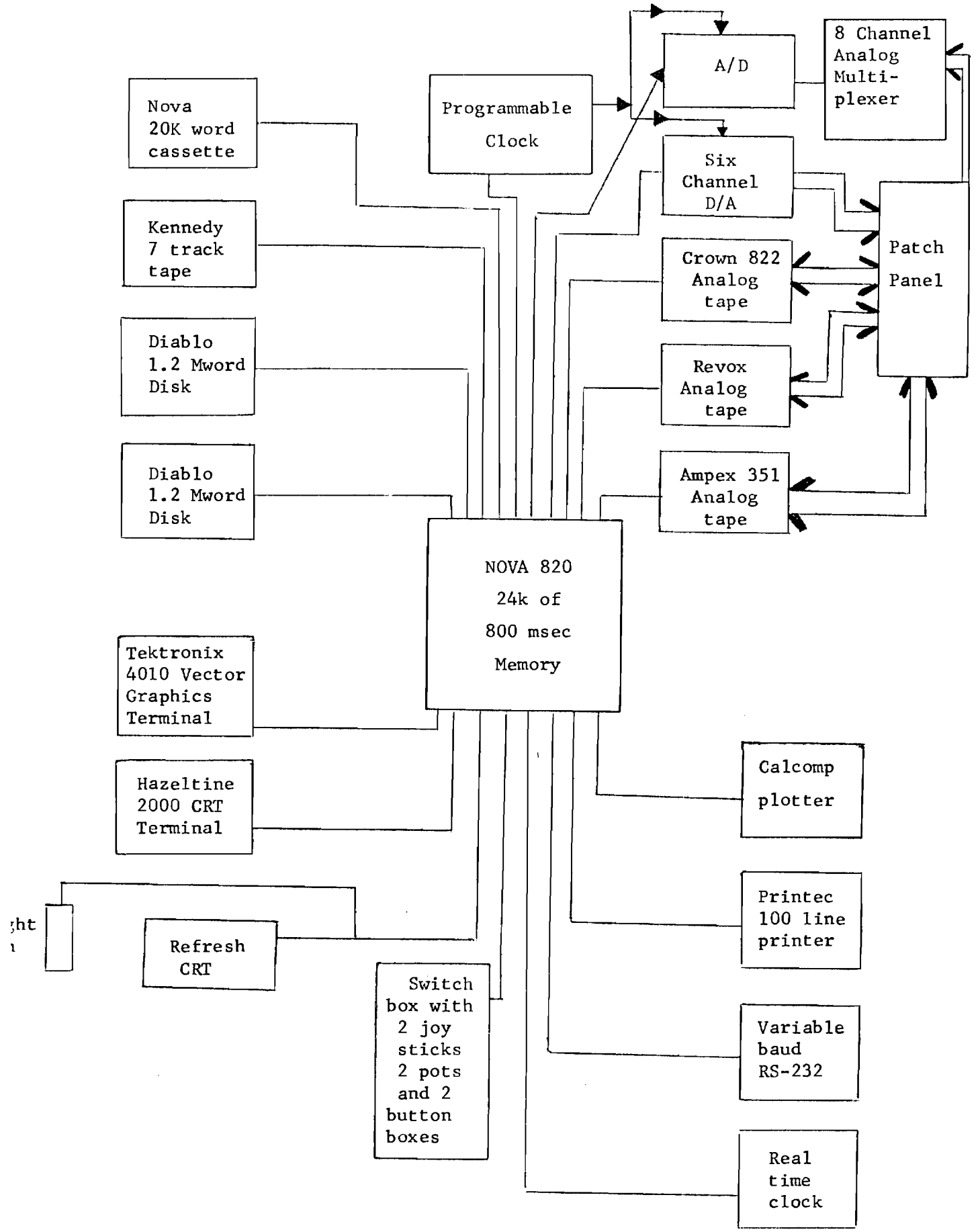


Figure 1. Layout of NOVA 820 Facility used in the Speech Algorithm Development and Testing

Table 1

NOVA 820 Speech Facility

MAINFRAME:

NOVA 820 Processor

24k of 800 μsec core memory

hardware Multiply/Divide

real time clock

2 RS-232 serial interfaces

GRAPHICS:

large screen refresh CRT

light pen

input box, 32 switch inputs, two joy sticks, and 2 pots

Tektronix 4010 vector graphics terminal

Calcomp plotter

STORAGE:

two Diablo moving head disks units totaling 2.4 M words of memory

Kennedy Incremental 7-track tape unit

NOVA cassette tape unit

ANALOG:

Crown 822 1/4" audio tape drive with 4 channel inputs, preamplifiers, and Crown D 60 power amplifier

Revox A77 1/4" audio tape drive with quarter track heads and Dolby preamplifiers

two Ampex model 351 1/4" audio tape drives with two channel preamplifiers

8 channel A/D

6 channel D/A

Crown IC150, Crown DC300, and Bose 901 listenint facility

7 position listening facility using Sennheiser HD414 headsets

MISCELLANEOUS:

Hazeltine 2000 CRT terminal (foreground terminal)

Programmable clock for sampling

Programmable rate RS-232 serial interface

Kay sonograph

Hewlett Packard/Altec real time spectrum analyzer

In order to make this basic system into a flexible speech processing unit, a number of hardware additions needed to be made. One of the perenial problems in developing minicomputer hardware systems is that interfaces developed for one type of computer can seldom be used on another. Contrastingly, most modern 16 bit minicomputers share many input-output characteristics. Hence, it was decided to develop a modular packaging system in which control funtions peculiar to particular computers could be separated from those functions common to most computers. This allows large portions of interfaces designed for one make of computer to be used, without modification, on another. This philosophy, in our environment, resulted in the construction of card cages in which the Nova 820's 16 bit data bus was physically separated from its device select bus and its control bus. All interfaces and special processors were built in two such chassis and consisted of approximately forty individual printed circuit cards.

The hardware additions specifically included digital-to-analog and analog-to-digital conversion capability. Data for both units are handled through accumulator transfers and are controlled by a programmable clock. The single Zeltex 10 bit A-to-D is preceded by an 8-channel analog multiplexer. Six channels of D-to-A were also included. The channel used for speech reconstruction is double buffered to prevent time jitter and has 16 bits resolution. The other D-to-A's are used in conjunction with a high persistence CRT for refresh graphics.

Three analog tape drives, a Crown 822, a Revox, and an Ampex 351 are also interfaced for computer control. The Crown drive has a fairly complex interface which allows the computer to position the tape as desired.

Other special devices include two programmable baud rate RS-232 serial interfaces, an interface for the Printec 100 line printer, 2 joy sticks,

2 button boxes, one 16 bit switch box, an interface for the Kennedy seven track incremental tape drive, and a light pen. The 7 track tape drive and the serial interfaces are used for communications with other computers while the joy sticks, button boxes and light pen are used to increase user interaction with the system.

2.2 The NOVA 820 Software System

The main criteria in the design of the NOVA 820 software system are experimental flexibility, ease of programming, and maximum interaction with the algorithms being tested. In order to meet these criteria, the following rules were adopted.

1) All programming was done in FORTRAN IV. A special feature of Data General's FORTRAN IV, which allows the direct dropping of assembler code into the FORTRAN instruction stream, is used to handle the special I/O functions of the non-standard hardware.
2) All inputs and outputs are disk files.
3) All systems conform to a NOVA system command template.
4) Any function to be used with two or more algorithms is programmed as a separate program module.
5) All algorithms being tested include "point-by-point" or "frame-by-frame" interactive graphics and interactive analysis subroutines for maximum speed in algorithm development.
6) All programs take control parameters either from the console or from a parameter file on the disk.
7) All programs can optionally output any of their internal parameters as disk files.

Following these rules results in a software system in which the experimenters can:

1. Concatenate many subsystems into larger systems.

2. Optionally watch every calculation or watch none.

3. Hear the result immediately.

4. Reprogram quickly.

5. Examine the total run data at his leisure.

6. Change parameters dynamically.

This degree of flexibility and interaction allows the examination of many more algorithms than would a less interactive system. Following is a description of some relevant programs and subprograms.

2.2.1 The LOOK Interpreter

When doing algorithm development under the NOVA 820 system, the experimenter has the option of dumping many internal parameters out as numeric time sequences in disk files. The LOOK interpreter is a program which allows the user to interactively examine several interrelated data files using the graphics terminal, the Calcomp plotter, and the refresh CRT. LOOK takes a program which causes it to maintain up to eight plots based on up to four input files. LOOK then allows the user to:

1. Look at any point in any file, while maintaining the interfile relationships.

2. Plot the display on the Calcomp plotter.

3. "Sweep" the data on the refresh CRT.

4. Change the plot characteristics and relationships dynamically.

Hence, a user may run a complex processor, and then use LOOK interactively to compile the results.

2.2.2 The In-Program Graphics Programs

The table of the in-program graphics programs available in algorithm production is given in Table 2. The basic graphics programs GRAF and IGRAF allow the straight plotting of available parameters. Other programs, like

Table 2

NOVA 820 Graphics/Analysis Subprograms

Name	Purpose
GRAF	Plots real data on vector terminal
IGRAF	Plots integer data on vector terminal
SPEC	Plots spectrum for LPC
ZPLOT	Plots z-plane poles using recursive polynomial solution
ZZPLOT	Plots z-plane poles using approximate polynomial solution

SPEC, ZPLOT, and ZZPLOT, allow the user on-line analysis during the data processing.

2.2.3 Additional Utility Programs

Table 3 gives a list of other utility programs available to the user. These include graphics capability, digital filter programs, spectrum analyzers, pitch detectors, and other useful "pocket" tools for speech analysis. These all combine to give the user considerable power for interactive processing.

2.3 Speed Related Software

Table 4 shows a summary of speech processors developed in support of this project. This group includes several vocoders, an ADPCM algorithm, and a speech synthesizer. All these programs are highly interactive and modular, and can realize many variations on these basic algorithms.

3. The Pitch Detection Algorithms

A number of pitch extraction algorithms were investigated as part of this study. The desire was to obtain a reasonably compact algorithm which was consistent enough so that its output might easily be used with synthetic speech.

3.1 The Pitch Testing Environment

In an environment where new pitch extraction algorithms are being developed, it is important to be able to quickly evaluate the results. It was decided that, for the most part, the goodness of any pitch detection scheme would be measured by its ability to reproduce the pitch contour of the utterance being tested. To test this, clearly, the pitch contours for the utterances must be available. The pitch painting techniques using the NOVA facility are discussed here.

Table 3

NOVA 820 General Subsystems

Name	Purpose
	Pitch Related
PTGTC	Interactive pitch handpainter
HLPD	Hard limited Autocorrelation pitch
MDPD	Minimum difference pitch detector
ZCPD	Zero crossing pitch detector
PCHECK	Pitch testing algorithm
	Speech Related
HEAR	Speech sampling Program
SAY	Speech playback Program
SPCAND	Creates spectrum for PLOT3D using LPC or FET
	Graphics Related
PLOTN	Plot several plots on Calcomp plotter from disk files
SPECPLOT	Plots magnitude and phase for digital filter
PLOT3D	3-D plotter for spectrums
	Digital Signal Related
DFDP	Designs digital filters interaction
FILTER	Canonical form digital filter
SF	Canonical form adaptive digital filter
	General
DECK	Controls Analogue tape units
DATASTART	Creates initial data file
DATAMAKE	Creates sequential data files
UVI	Inputs 7-track tapes from Univac 120
UVO	Outputs 7-track tapes from Univac 120
NORM	Normalizes a data file
ACONTS	Concatenates speech files

Table 4

Speech Processor Programs

Name	Purpose
LPC	Linear Predictive Coder Transmitter
LPR	Linear Predictive Coder Receiver
LSQ	Least Squares Linear Predictive Coder
LMS	LMS Linear Predictive Coder
HIRE	Homomorphic Impulse Response Extractor
XMTR	Homomorphic Vocoder Transmitter
RCVR	Homomorphic Vocoder Receiver
ADPCM	Adaptive Differential Pulse Code Modulator
SPEAK	Speech Synthesizer

3.1.1 The Pitch Contours Preparation Program (PTGTC)

One method of controlling the pitch contour of the test utterances is to generate the utterances synthetically. This method, however, is not acceptable in the current environment because it also controls the other characteristics of the speech. Hence, it was decided to extract the pitch contours from real speech.

The program which was designed to aid in the pitch extraction problem is called PTGTC (pitch get with cepstrum), and a table of its commands are given in Table 5. PTGTC is a highly interactive program running on the Tektronix 4010 graphics terminal. The user, interactively, steps through the speech, frame by frame, and chooses a pitch period for each frame.

PTGTC has two input files. These may be anything, but are generally the original speech and a digitally low pass filtered version of the input speech. The user chooses and inputs the pitch period by using the cross hair capability of the graphics terminal. At each frame, the period can be set from the original speech, the filtered speech, an autocorrelation function for either the original or filtered speech, a hard-limited autocorrelation function for either the original or filtered speech, or the cepstrum of either the original or the filtered speech. The auxiliary functions are calculated and plotted only on request, so the user may move quickly in easy regions of the speech, and move more slowly in difficult regions.

At each frame, the user has four options: he may mark a pitch period; he may mark the frame unvoiced (0 pitch period); he may mark the frame silent (-1 pitch period); or he may mark the frame with an optional pitch period (negative pitch period). In the last case, the user is saying that he will accept either a voiced or an unvoiced decision. An example pitch period plot is shown in Figure 2.

Table 5

Hand Painting Pitch Program Parameters

Name	Purpose
	Input Files
I	Input speech file
F	Input filtered speech file
	Input Parameters
R	Autocorrelation window size
F	FFT window size
B	Current frame number
T	Current pitch period
X	X axis hash mark interval
Y	Y axis hash mark interval
S	Scale factor
P	Current sample power
J	Frame length
	Commands
F	Do FFT
R	Do autocorrelation
C	Do cepstrum
N	Go to next frame
1	Set right interval boundary
2	Set left interval boundary
H	Do hard limited autocorrelation
T	Type current period
I	Type current information
M	Change to parameter recode
Q	QUIT

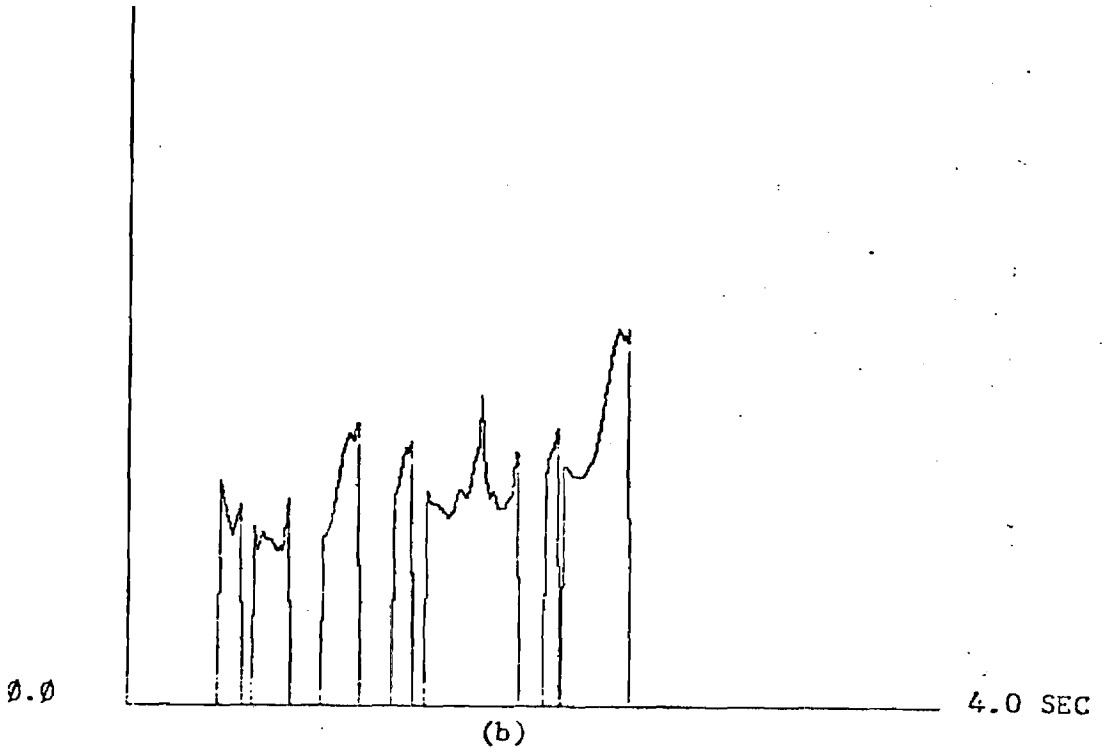


Figure 2. A Typical Pitch Contour

3.1.2 Evaluation of Synthetic Pitch Contours

The most common use of the pitch contours derived from PTGTC is a visual comparison using the handpainted pitch contour as a reference (using LOOK). This gives the experimenter a quick test for the location and size of his error. In addition, there is an error statistics program, called PCHECK, which does statistics on the comparison between two pitch contours. In particular, PCHECK outputs:

1. The number of errors
2. The average errors per sample in voiced regions
3. The number of gross errors (greater than a threshold)
4. The average gross errors
5. The number of subtle errors (less than a threshold)
6. The average subtle errors
7. The number of voicing errors
8. Standard deviations for the above averages

In addition, PCHECK outputs a histogram of the errors to the disk for further examination.

3.2 The Pitch Extractors

Several pitch extraction configurations were tested as part of this program. In each case, the algorithm was programmed using interactive graphics, run on specially prepared test utterances, and then evaluated using PCHECK.

3.2.1 Zero Crossing Detectors

The simplest pitch detectors investigated were the zero crossing detectors. The basic method used is to first low pass filter the speech with a cutoff of about 400 Hz, and then estimate the pitch from the zero crossings.

The operation of one of the zero crossing detectors used is illustrated in Figure 3. The speech signal is first analog-to-digital converted, and then low pass filtered using a digital filter. The filter used in this study was an eight pole Butterworth impulse invariant filter with a variable cutoff, usually set at 400 Hz. Depending on the pitch, this remaining signal contains the fundamental pitch frequency and possibly several harmonics.

This signal was examined just before sample K*C, where C = 1,2,... and K is a fixed interval. Moving back from K*C, the first zero crossing is found. The distance between this zero crossing and the next is then measured. This process is repeated, in pairs, until a time limit has been exceeded. Thus pairs of zero crossing intervals, ξ_i and ξ'_i, form the input to the detection algorithm. The detection algorithm itself is capable of removing the effects of one harmonic. It does this by comparing $|\xi_i - \xi'_i|$ to a threshold. If it is greater than this threshold, then the interval $\xi_i + \xi'_i$ is taken to be only half a pitch period. The voicing decision is based solely on the variation of the ξ_i's and ξ'_i's.

As is typical of zero crossing pitch detectors, this detector worked well for some subjects and utterances and not well for others.

3.2.2 The Autocorrelation and Hard Limited Autocorrelation Pitch Detectors

The autocorrelation pitch detector uses the short time autocorrelation function, defined as

$$R_{K,\ell} = \sum_{i=1}^{N} S_{K-i-1} \, S_{K+\ell-i-1} \qquad (3.1)$$

where S_i is the $i^{\underline{th}}$ speech sample, ℓ is the lag, and N is the window size. The hard limited autocorrelation function is defined by

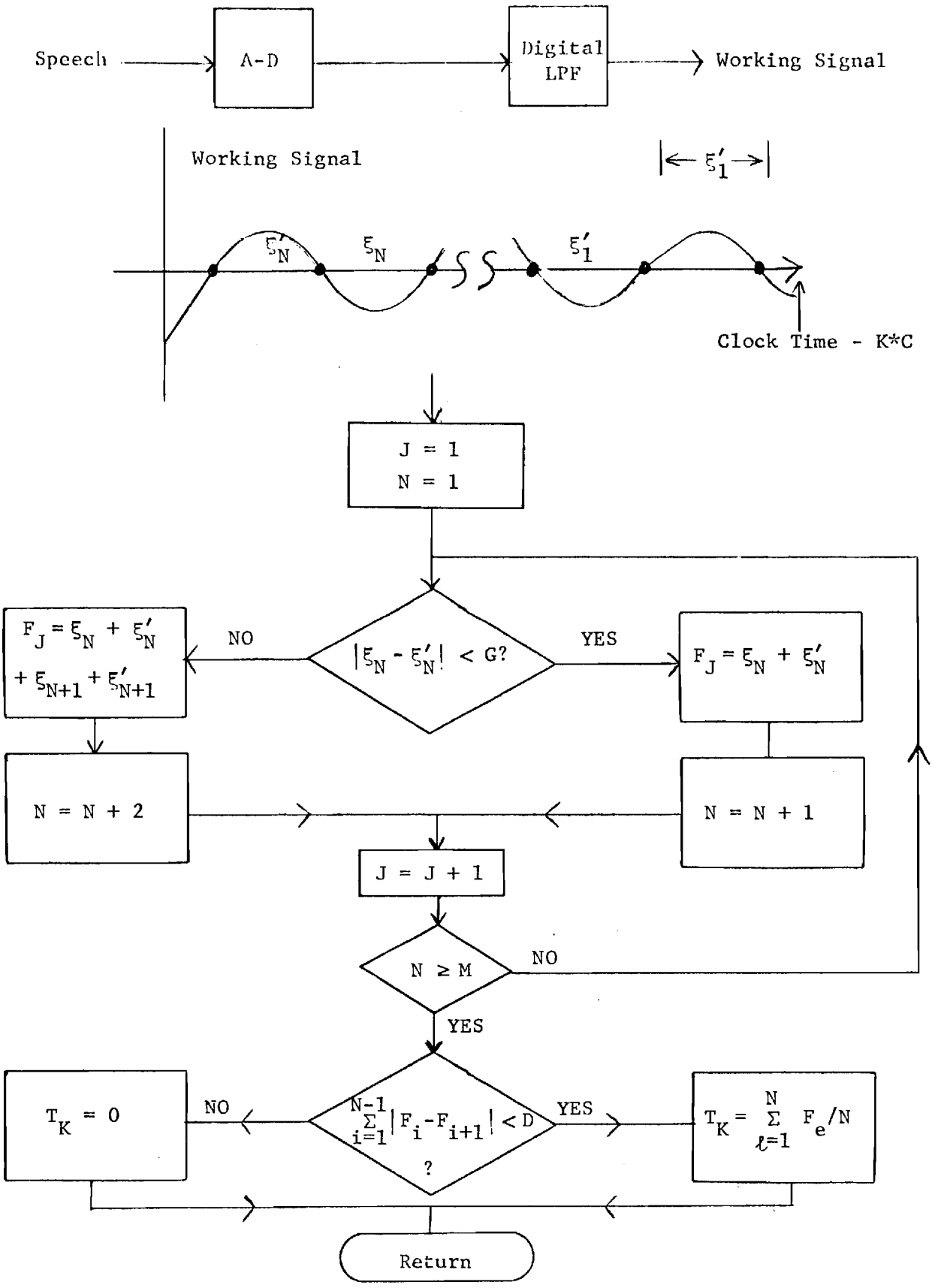


Figure 3. Zero Crossing Pitch Detector

$$R_{K,\ell} = \sum_{i=1}^{N} SGN(S_{K+i-1}) \; SGN(S_{K+\ell-1-i}) \qquad (3.2)$$

This function is much easier to calculate since it involves only "count ups" and "count downs" rather than multiplies and adds.

The basic operation of these pitch detection algorithms is illustrated in Figure 4.

3.2.3 The Minimum Difference Detector

The minimum difference pitch detector is very similar in concept to the autocorrelation detector. Instead of finding peaks in the autocorrelation function, it finds minimums in the minimum difference function, defined by

$$MDF_{K,\ell} = \sum_{i=1}^{N} |S_{K+1-i} - S_{K+\ell-1-i}| \qquad (3.3)$$

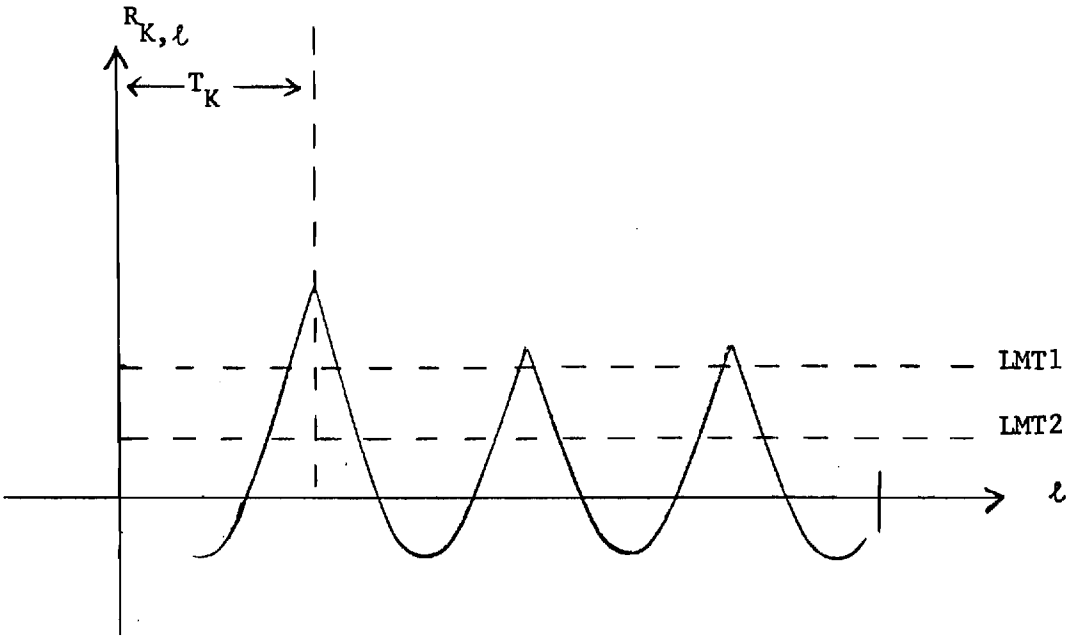
Figure 5 illustrates the operation of the basic minimum difference algorithm.

3.2.4 The Cepstrum Detector

The cepstrum detector tested is illustrated in Figure 6. It was originally thought that this might be the best detector. However, testing showed it to run excessively long, and due to its windowing constraints, to be comparatively inaccurate.

3.2.5 The Multiband Pitch Detector

A block diagram for the multiband pitch detector is shown in Figure 7. In that diagram, the pitch estimates are zero crossing detectors. Note that the filter always isolates the fundamental in the lowest band. The decision algorithm receives information from all four detectors to help decide on the proper pitch.



$$R_{K,\ell} = \sum_{i=1}^{N} SGN(S_{K+i-1})\ SGN(S_{K+\ell-1-i})$$

$$P_{\ell MAX} = \sum_{\ell=\ell_{MAX}-P}^{\ell_{MAX}+P} SGN(R_{K,\ell} - R_{K,\ell-1})\ SGN(R_{K,\ell+1} - R_{K,\ell})$$

LMT1 = A*N , LMT2 = C*N

T_K = Pitch Period

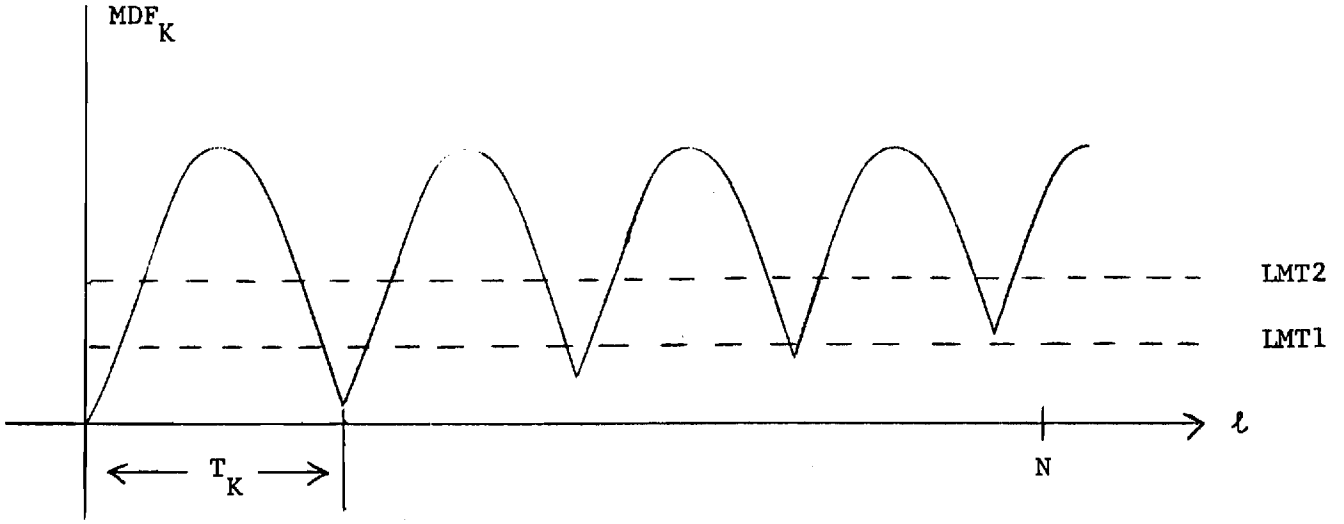
A,B,C,D,P,N,Z = Algorithm Parameters

Pitch Period = First local maximum for which R_{MAX} > LMT1 and $P_{\ell MAX}$ > B. If no such maximum is found Pitch Period = R_{MAX} over N interval if R_{MAX} > LMT2 and number of zero-crossings < D*N. Otherwise, Pitch Period = 0.

Typical Values:

$A = 0.7_8$	$P = 20_8$
$B = 0.2_8$	$N = 200_8$
$C = 0.4_8$	$Z = 2000_8$
$D = 0.4_8$	

Figure 4. Illustration of Operation of the Basic Algorithm for the Hard-Limited Autocorrelation Pitch Detector.



$$MDF_{K,\ell} = \sum_{i=1}^{N} |S_{K+1-i} - S_{K+\ell-1-i}|$$

$$P_{\ell MIN} = \sum_{i=\ell_{MIN}-P}^{\ell_{MIN}+P} SGN(MDF_{K,i} - MDF_{K,i-1}) \; SGN(MDF_{K,i} - MDF_{K,i+1})$$

LMT1 = A*BL , LMT2 = B*BL

T_K = Pitch Period

A,B,C,D,P,N = Algorithm Parameters

Pitch Period = First local minimum for which $MDF_{MIN} < LMT1$ and $P_{\ell MIN} > B$. If no such minimum is found, pitch period = MDF_{MIN} over N interval if $MDF_{MIN} < LTM2$ and the number of BL-crossings < D*N. Otherwise, Pitch Period = 0.

Figure 5. Illustration of Operation of the Basic Algorithm for the Minimum Difference Pitch Detector.

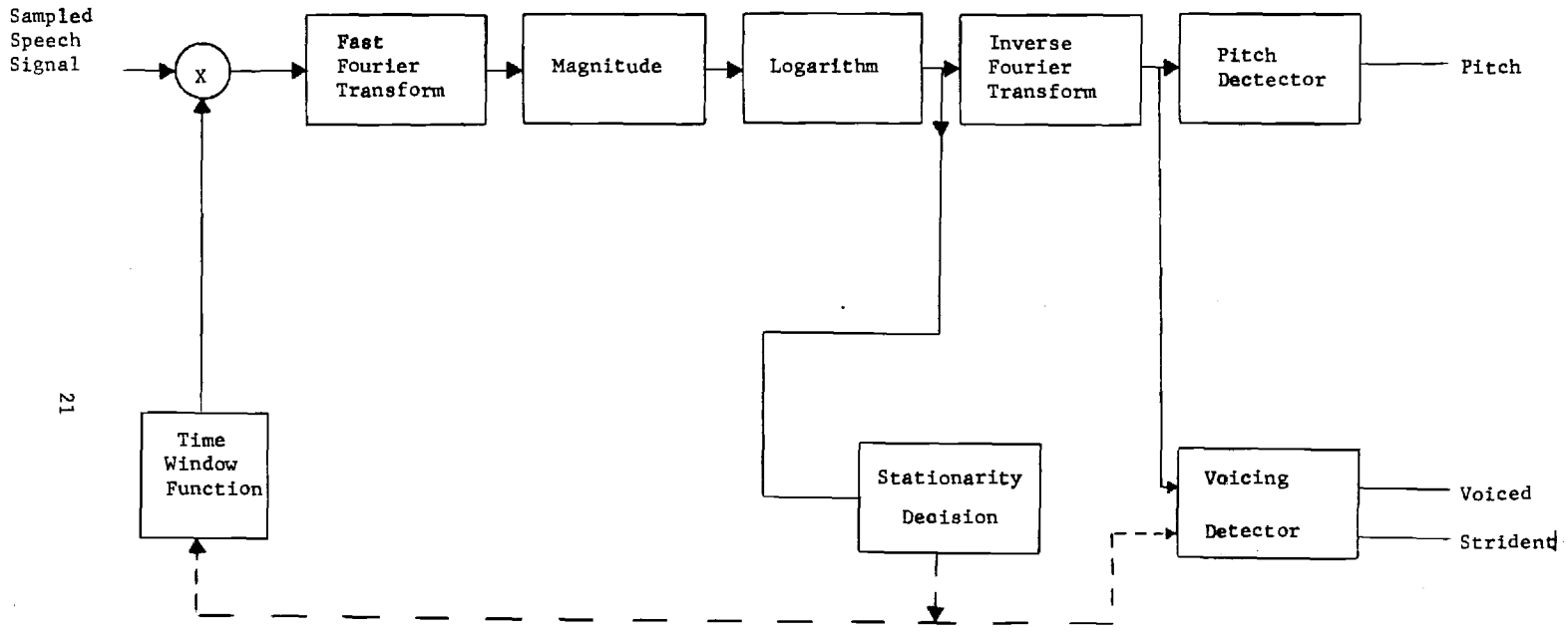


Figure 6. A Block Diagram of the Proposed Adaptive Cepstrum Pitch Detector

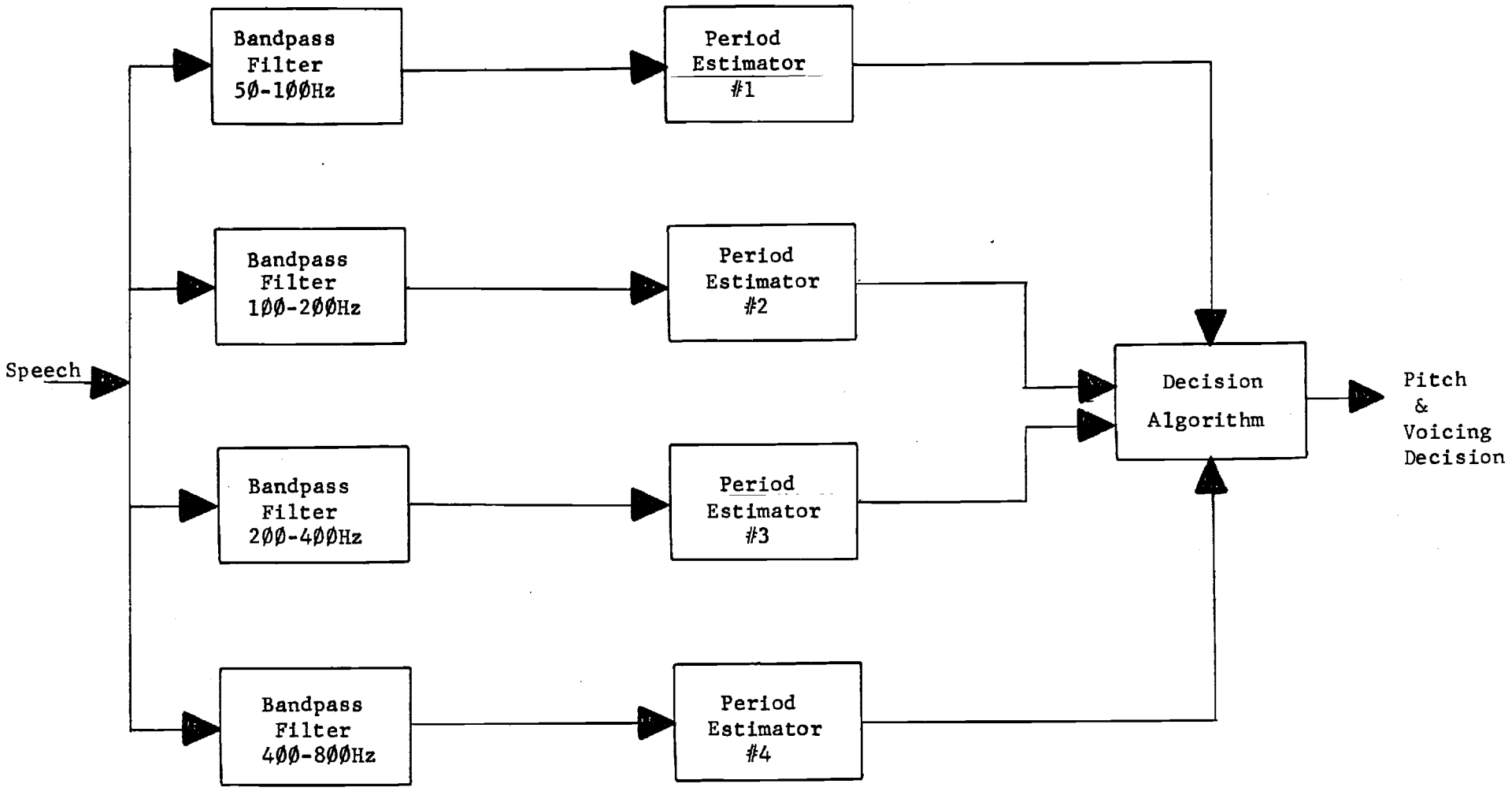


Figure 7. Multi-Band Pitch Detector

3.3 General Pitch Detector Results

It was originally thought that the cepstrum pitch detector would be the detector of choice. However, initial testing showed that the cepstrum detector ran very long, and gave results which had numerous small errors. The autocorrelation detector was very accurate but tended to make gross errors near voiced-unvoiced junctures. The additional logic needed to correct these errors plus the time involved in the calculation of the autocorrelation functions made this detector somewhat undesirable.

The two autocorrelation "speed up" detectors, the hard limited autocorrelation detector and the minimum difference detector, behaved similarly to the autocorrelation detector. Both were very accurate, but required longer windows to obtain this accuracy. This partially offsetted the gains in time due to the simpler basic functions. Like the autocorrelation detector, considerable extra logic was necessary to correct local gross errors.

The various zero crossing detectors studied were by far the fastest algorithms. Also, during many regions, they were quite accurate. However, they made numerous local gross errors, which made them particularly unusable for vocoder applications.

The multiband detector, was, on the whole, the most successful detector. It was quite fast when compared with all but the zero crossing detector, and very accurate. It seems very suited for vocoder applications.

4. The Pitch Information Study

In the F_o study, the structures of interest were limited to three word noun phrases of the three most common types (Table 6). The words and the noun phrases were limited to single syllable CVC (consonant-

Table 6

Structure of the Noun Phrases used in the Pitch Experiment

	STRUCTURE	STRESS	EXAMPLE
1.	Adjective - Adjective - Noun		
	$NP[_A[X]_A\ NP[_A[Y]_A\ N[Z]_N]_{NP}]_{NP}$	231	Sad Fat Man
2.	Adjective - Compound		
	$NP[_A[X]_A\ N[_{NVA}[Y]_{NVA}\ NVA[Z]_{NVA}]_N]_{NP}$	213	Big Blackboard
3.	Three Word Component		
	$N[_N[_{NVA}[X]_{NVA}\ NVA[Y]_{NVA}]_{NVA}[Z]_{NVA}]_N$	132	Blackboard Eraser

N = NOUN

NP = NOUN PHRASE

NVA = NOUN, VERB, OR ADJECTIVE

A = ADJECTIVE

vowel-consonant) words, and the intonational contours studied included only the low fall (statement) and high rise (question) intonations. For the purposes of the study the phonemes were divided into seven classes: vowels; glides and liquids; voiced stop consonants; unvoiced stop consonants; nasals; voiced fricatives; and unvoiced fricatives.

A total of six vowels, /i/, /ɪ/, /æ/, /ʌ/, /ʊ/, and /ə/, were studied For each vowel a group of five phrases was constructed for each of the three structural groups in Table 6. The phrases were chosen so as to represent a reasonable population of all the six classes of consonants under study. Between groups phrases were chosen, whenever possible, so that the same phonemes or phoneme classes resulted (e.g. "sad sadsack" and "sad sad sack"). In all, 120 phrases were involved.

Each of the phrases was recorded by two subjects, one male and one female, using a high quality dynamic microphone and a Crown 822 tape recorder. The phrases were then low pass filtered to 3.2 kHz and sampled at 8 kHz with 10 bits linear resolution. Pitch contours for each of the phrases were then constructed using either the multiband pitch detector or a specially modified version of the zero crossing pitch detector.

Analysis of the pitch data in this study has not yet been completed. However, several important facts have come out of this preliminary study:

1. Similar structures result in pitch contours with similar features.
2. Pitch excursions are about the same in any particular structural location independent of the length of the vowel.
3. The fine detail of the pitch contours is definitely and systematically related to segmental contents.

PROJECT PERSONNEL

1. T. P. Barnwell, Principal Investigator

As principal investigator, Dr. Barnwell conducted all the project experiments, oversaw the hardware development, designed and implemented the software, and prepared this report.

2. J. H. Schlag

Dr. Schlag, Associate Professor of Electrical Engineering, provided invaluable aid in the design and construction of the new hardware. His efforts were contributed out of his "own" time.

3. Richard Patisaul, Graduate Research Assistant

Dr. Patisaul (Ph.D. in September 1974) was of great help in the design and testing of the multiband pitch detector. The Homomorphic vocoder used was largely a result of his thesis work.

4. D. L. Smith, Graduate Assistant

Mr. Smith was in charge of general laboratory maintenance, and helped with hardware construction and design.

PAPERS ACCEPTED OR IN PREPARATION

1. "The Multiband Pitch Detector," submitted to IEEE Transactions on Acoustics, Speech and Signal Processing.

2. "A Minicomputer Based Digital Signal Processing Facility," EASCON '74 RECORD, October 1974.

3. "Gapped Analysis for Speech Digitation," Proceedings of the National Electronics Conference, October 1974.

BIBLIOGRAPHY

1. Barnwell, T. P., "An Algorithm for Segment Durations in a Reading Machine Context," Ph.D. Thesis, M.I.T., 1970.

2. Voiers, W. D., "The Present State of Digital Vocoding Technique: A Diagnostic Evaluation," IEEE Transactions on Audio and Electroacoustics, vol. AU-16, pp. 275-279, 1968.

3. Bolinger, D. L., "A Theory of Pitch Accent in English," Word, vol. 14, no. 2-3, 1958.

4. Bolinger, D. L., "Ambiguities in Pitch Accent," Word, vol. 17, no. 3, 1961.

5. Denes, P., "A Preliminary Investigation of Certain Aspects of Intonation," Language and Speech, vol. 2, 1959.

6. Denes, P. and Milton-Williams, J., "Further Studies in Intonation," Language and Speech, 1962.

7. Fry, D. B., "Experiments in the Perception of Stress," Language and Speech, vol. 1, 1958.

8. Hadding-Koch and Studdert-Kennedy, M., "An Experimental Study of Some Intonational Contours," Phonetica, vol. II, 1961.

9. Lieberman, P., Intonation, Perception, and Stress, M.I.T. Press, 1967.

10. Lieberman, P., "Some Acoustic Correlates of Vowel Intonation," Q.P.R., R.L.E., M.I.T., no. 2, 1958.

11. Morton, J. and Jassem, W., "Acoustic Correlates of Stress," Language and Speech, vol. 8, 1966.

12. Lee, F. F., "A Study of Grapheme to Phoneme Conversion of English," Ph.D. Thesis, M.I.T., 1965.

13. Allen, J., "A Study of the Specification of Prosodic Features of Speech from a Grammatical Analysis of Printed Text," Ph.D. Thesis, M.I.T., 1968.

14. Mattingly, I. J., "Prosodic Features for Synthesized Speech," 1965.

15. Gold, B. and Rader, C., Digital Processing of Signals, McGraw-Hill Book Co., Inc., New York, 1969.

16. Noll, A. M., "Cepstrum Pitch Determination," Journal of the Acoustical Society of America, vol. 41, pp. 293-309, 1967.

17. Gold, B., "Computer Program for Pitch Extraction," The Journal of the Acoustical Society of America, vol. 34, no. 7, pp. 916-921, July 1962.

18. Gold, B., "Note on Buzz-Hiss Detection," The Journal of the Acoustical Society of America, vol. 36, no. 9, pp. 1659-1661, September 1964.

19. Gold, B., and Rabiner, L., "Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time-Domain," The Journal of the Acoustical Society of America, vol. 46, no. 2, pp. 442-448, February 1969.

20. Miller, R. L., "Performance Characteristics of an Experimental Harmonic Idenfitication Pitch Extraction (HIPEX) System," The Journal of the Acoustical Society of America, vol. 47, no. 6, pp. 1593-1601, June 1970.

21. Schroeder, M. R., "Period Histogram and Product Spectrum: New Methods for Fundamental-Frequency Measurement," The Journal of the Acoustical Society of America, vol. 43, no. 4, pp. 829-834, April 1968.

22. Harris, C. M., and Weiss, M. R., "Pitch Extraction by Computer Processing of High-Resolution Fourier Analysis Data," The Journal of the Acoustical Society of America, vol. 35, no. 3, pp. 339-343, March 1963.

23. Weiss, M. R. and Vogel, R. P., "Implementation of a Pitch Extractor of the Double-Spectrum-Analysis Type," The Journal of the Acoustical Society of America, vol. 50, no. 2, pp. 637-665, August 1971.

24. Atal, B. S., and Hanauer, S. L., "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," The Journal of the Acoustical Society of America, vol. 50, no. 2, pp. 637-665, August 1971.

25. Markel, J. D., "Application of a Digital Inverse Filter for Automatic Formant and F_o Analysis," IEEE Transaction on Audio and Electroacoustics, vol. AU-21, no. 3, pp. 154-160, June 1973.

26. Maksym, J. N., "Real Time Pitch Extraction by Adaptive Prediction of the Speech Waveform," IEEE Transactions on Audio and Electroacoustics, vol. AU-21, no. 3, pp. 149-154., June 1973.

27. Itakura, F. and Saito, S., "On the Optimum Quantization of Feature Parameters in the PARCOR Speech Synthesizer," Conference Record, IEEE 1972 Conference on Speech Communication and Processing, IEEE Catalog No. 72-CHO-596-7-AE, pp. 434-437, April 1972.

28. Boll, S. F., "A Priori Digital Speech Analysis," Advanced Research Projects Agency Report, AD762-029, and Ph.D. Thesis, University of Utah, March 1973.

29. Makhoul, J. I. and Wolf, J. J., "Linear Prediction and the Spectral Analysis of Speech," Bolt, Beranek and Newman, Inc,, Report BBN-2304, August 1972.