Acoustic Analysis of the Effects of Alcohol on the Human Voice

By

Orla Cooney, BSc ,

Presented to the Office of Academic Affairs in fulfilment of the thesis requirements for the degree of Master of Science in the School of Physical Sciences, Dublin City University, Dublin 9

June, 1998

Under the supervision of

Professor Martin Henry,
School of Physical Sciences,
Dublin City University,
Dublin 9

Dr Kevin McGuigan,
Department of Physics,
Royal College of Surgeons in Ireland,
123 St Stephen's Green,
Dublin 2

(Single Volume)

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of MSc is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: ~~Orla Cooney~~ **ID No.:** 95971238

Date: 10th June 1998

For Mam and Dad

Acknowledgements

I would like to start by thanking Dr Kevin McGuigan for being my project supervisor and for all his help and guidance throughout the course of this research I would like to thank Professor Martin Henry for taking on the responsibility of being my external supervisor

Thanks to Ronan Conroy for all his help and advice with the statistics for this work

Many thanks to Peter Murphy for all his help and ideas regarding the present project

I would like to say a special thanks to the girls, Eimear O'Brien, Julie Duncan, Karen Pierce and Gillian McMahon, in the Chemistry & Physics Department for all the fun and great nights had over the last couple of years Special thanks also to Joe Jolley for the many laughs had in the lab I would like to thank Tony Brennan and Terry Murphy for being there when any help was required

Thanks to all the staff of the Chemistry & Physics Department, past and present

Finally, I would like to thank my parents for all their help and support throughout the course of my studies

Table of Contents

1 Theory of Speech Production

1 1 Speech Production 1

1 2 Effects of the Vocal Tract 2

1 3 Spectral Details of Speech 3

1 3 1 Fundamental Frequency and Harmonics 3

1 3 2 Formant Frequencies 4

1 4 Physical Effects of Alcohol on Speech 6

2 Materials and Methods

2 1 Hardware 10

2 2 Experimental Details 12

2 3 Software and Analysis 13

2 3 1 Fourier Transform and the Fast Fourier Transform 13

2 3 2 Windowing 16

2 3 3 Cepstral Analysis 20

2 3 4 Formant Frequency Analysis 23

2 3 5 Sentence Duration Analysis 25

3 Results

3 1 Sustained Vowel Phonation and Isolated Utterances 26

3 2 Utterance of vowel sounds within spoken sentences 34

3 3 Effects of alcohol level on sentence duration 41

4 Discussion and Conclusion 47

5 Simulations of Noise induced Hearing Loss

5 1 Introduction 53

5 2 Choice of Filter 54

References 62

Appendix A: Coding and graphical outputs for principal Matlab program files

Fundfr m 1

Durat m 7

Formant m 10

Appendix B: Coding and graphical outputs for functions called by principal programs

Extract m	1
Specform m	4
Hamwind m	6
Specenvel m	7
Trackvec m	9

Acoustic Analysis of the Effects of Alcohol on the Human Voice.

By: Orla Cooney

Abstract

An experiment has been designed and implemented to determine the physical effects of alcohol on the human voice Sustained phonation and prescribed text voice recordings were taken from twelve volunteers at different stages of alcohol consumption A computer based analysis has been carried out on these voice recordings to determine whether a significant change in voice frequencies occurs when alcohol has been consumed The effect of alcohol consumption on sentence duration has also been investigated The results obtained from the analysis show that there is no association between the fundamental frequency of a sustained phonation and the level of alcohol consumption (regression coefficient = 0 146, F = 0 078, df = 71, p= 0 781) Similar results are recorded for the formant frequencies For the utterance of vowel sounds within spoken sentences it was found that the fundamental frequency rose on average by 1 7Hz (std err = 0 57, F = 8 9, df = 237, p = 0 003) per ml alcohol per kilogram bodymass Prescribed-text, sentence duration is observed to increase by 6 4% (std err 0 5%, p < 0 0001) per ml of alcohol consumed per kilogram bodymass There is no evidence that the sex of the speaker is related to the effect of alcohol on sentence duration We conclude that there is no correlation observed between the level of alcohol consumed and variations in the characteristics of the human voice

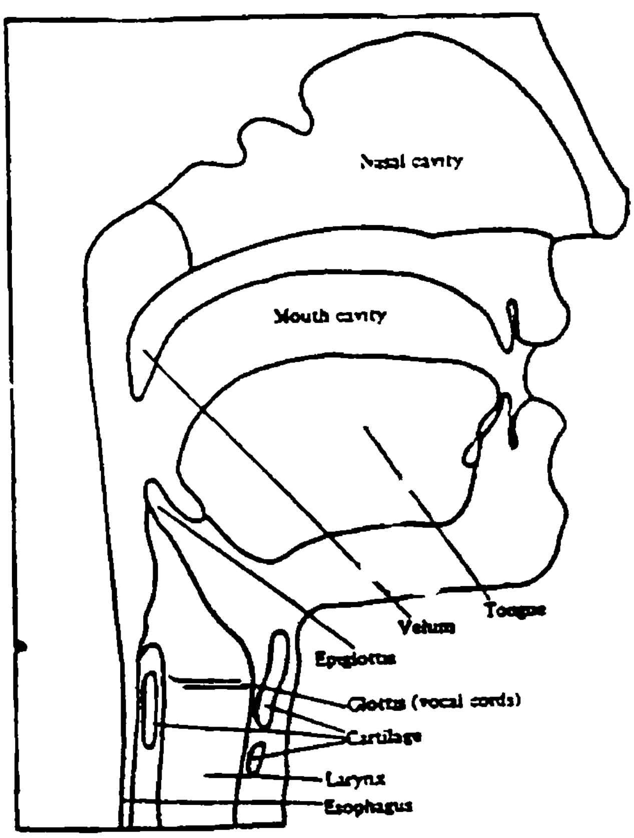
A system has been developed whereby speech samples are filtered as a means of simulating noise induced hearing loss A Chebychev low pass filter has been designed using the Matlab programming package and this filter has been applied to voice samples that are spoken in environments of different noise levels The resulting voice samples give an indication as to the level of hearing damage incurred by individuals suffereing from noise induced hearing loss

Chapter 1 :Theory of Speech Production

1.1 Speech Production

Human speech is produced by utilising the lungs, throat, mouth and nose The lungs force out a stream of air which is exhaled via the trachea through the rest of the vocal mechanism, known as the vocal tract Figure 1 1 [1] shows an illustration of the anatomy of the vocal tract

Fig.1 1. *Anatomy of the human vocal tract*



There are three ways in which the air flow, sustained by a steady pressure from the lungs, is converted into an acoustic signal having power components throughout the audio frequency range Firstly, the air stream vibrates the vocal chords situated in the larynx The vocal chords are folds of fleshy ligament stretched between the arytenoid cartilages and the thyroid cartilage of the larynx During normal breathing the arytenoid cartilages move apart to leave a large opening When swallowing the chords are held tightly together and the top of the trachea is shielded by the epiglottis to prevent food entering the lungs When the vocal chords are to be used for producing sound, the chords are held fairly close together and air is forced through the opening or glottis The air accelerates through the narrow glottis and then slows down as it enters the much wider pharynx above

The second main sound source is the air turbulence produced at a constriction somewhere in the vocal tract The type of sound produced is different from the vocal chord sound, its method of production is inherently random and it cannot have the repetitive structure which is typical of the glottal airflow

The third type of sound source results from the build up of pressure that occurs when the vocal tract is closed at some point A sudden release of pressure causes a transient excitation of the vocal tract which results in a sudden onset of sound.

1.2 Effects of the Vocal Tract

Sound from all the above mentioned sources enters the vocal tract The configuration of the tract, controlled by the positions of the tongue, jaw, velum and the lips, determines its acoustic properties and modifies the energy/frequency distribution of the sound source in conformity with the resonances and anti-resonances of the acoustic structure The frequencies of the anti-resonances of the complete system are also affected by the point at which the vocal tract is excited and the way in which sounds radiated from the nose and mouth are combined The resonances of the vocal tract cause concentrations of energy at certain frequencies which are known as formants of the speech

1.3 Spectral Details of Speech

1.3.1 Fundamental Frequency and Harmonics

The pitch or fundamental frequency of a speaker's voice is determined by the lowest frequency of vibration of the speaker's vocal chords The pitch of the voice of an adult male is usually lower than that of an adult female This is because the vocal chords are generally longer and heavier in the adult male than in the adult female Typical values of fundamental frequency for male, female and child would be 110Hz, 220Hz and 300Hz respectively [1]

A speech signal is composed of many different frequency values which are all integer multiples of the fundamental frequency The fundamental frequency is the first harmonic and all higher frequencies are higher harmonics

1.3.2 Formant Frequencies

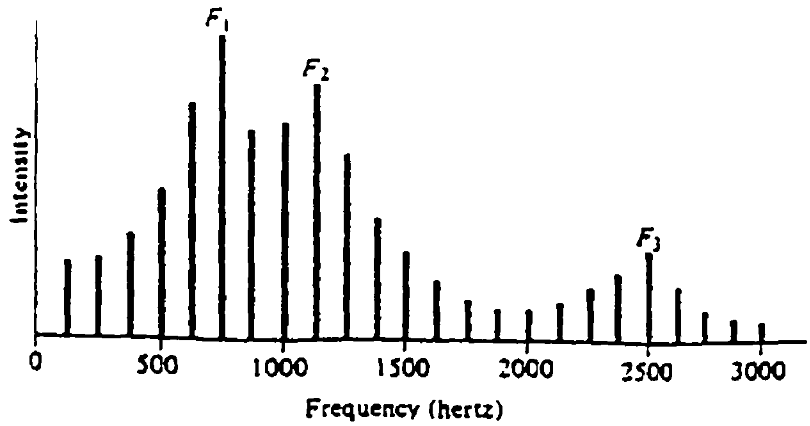
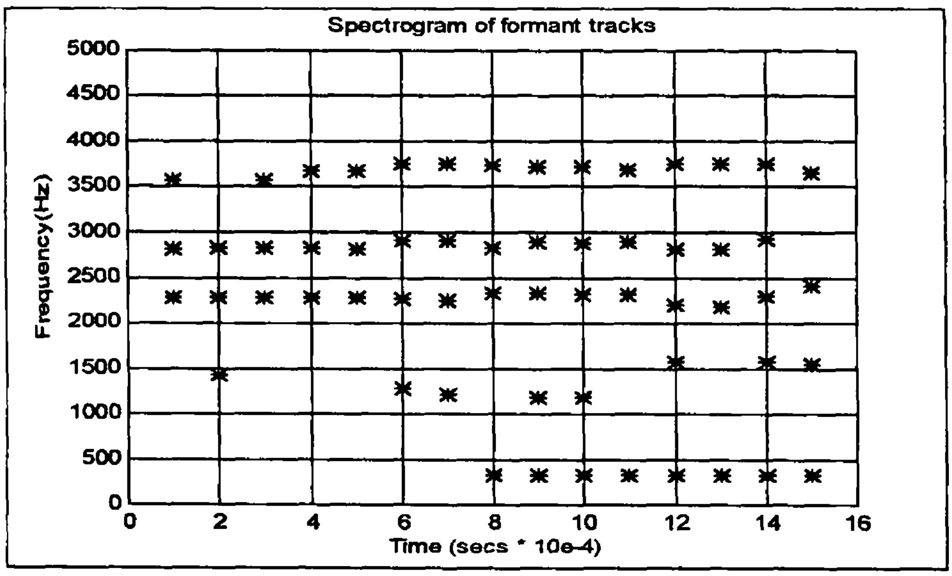


Fig 1.2. *Spectrum of a set of frequency data*

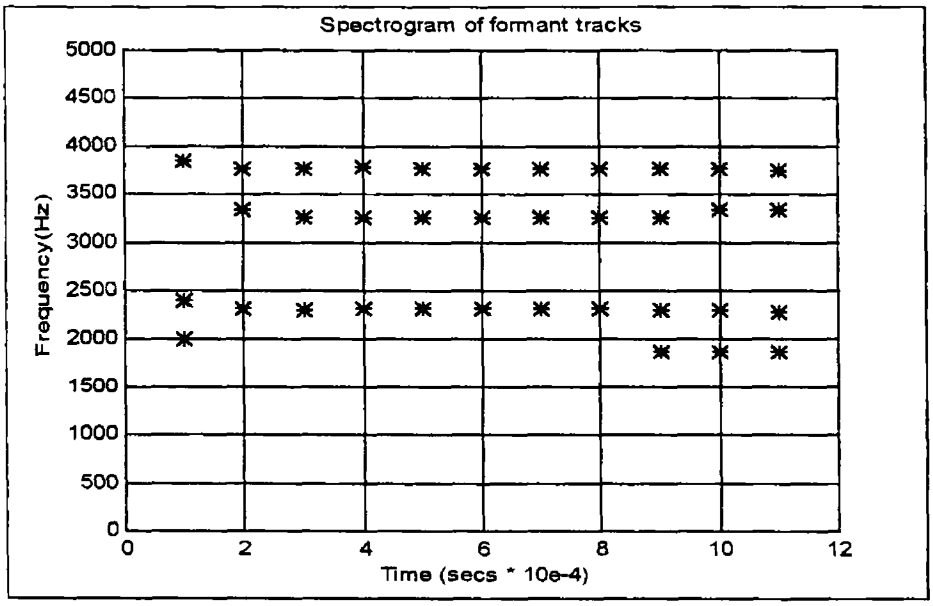
The pitch and intensity of speech sounds are determined predominantly by the vibrations of the vocal chords However, the spectrum of these sounds is strongly shaped by the resonances of the vocal tract Sounds are distinguished from each other according to the character of these resonances The vocal tract acts as a band pass filter for different regions of the spectrum These regions are seen as global peaks in the spectral envelope and are called formant peaks The frequency values at these peaks are the formant frequencies and they determine the vowel sounds The first two formants are of the highest amplitude and hence are of the greatest importance Figure 1 2 [1] shows an example of the spectrum of a set of frequency data The global peaks are the formant peaks Different vowel sounds have independent values of formant frequencies Rossing [1] quotes findings from Peterson and Barney (1952) which show the formant frequencies and amplitude of vowels averaged for seventy-six speakers made up of men, women and children The relative strengths of the three formants for each vowel are shown in this paper From this study it is observed that /a / (ah) has the strongest second formant, only 4 dB weaker than the first formant, whereas for /i/ (ee), the second formant is 20 dB below the first The actual formant frequency values will differ for each individual and they will also have different values depending on the particular vowel sound

Fig 1.3a. *Spectrogram of formant tracks for isolated word 'me'*



Figures 1 3a and 1 3b show the spectrograms of the formant tracks for the vowel sounds in the isolated words 'me' and 'you' for the same individual respectively These spectrograms were calculated using the Matlab program specform m The sample rate here is 10kHz

Fig. 1.3b. *Spectrogram of formant for isolated word 'you'*



1.4 Physical Effects of Alcohol on Speech

To date, there has been very little research undertaken to determine the physical effects of alcohol on the human voice Perceptually, the effects of alcohol on speech are quite apparent Past perceptual research has shown that a person listening to voice samples can reliably discriminate between sober and intoxicated speech Pisoni and Martin [2] found from their perceptual analyses that speech produced in the intoxicated condition is slower, lower in overall amplitude, more negatively judged in subjective perceptual tests, and more prone to errors at the sentence, word and phonological levels than speech produced in a sober condition This suggests that alcohol reduces the control and coordination of speech articulation, phonation and respiration

Pisoni and Martin also report that duration of speech was the most consistent difference observed between sober and intoxicated speech The average magnitude of sentence lengthening for individual talkers ranged from 75 to 158msec [2] Pitch level was found to be much more variable in the intoxicated condition Only very small differences were observed in the amplitude of speech segments in both the sober and intoxicated conditions

In their acoustic analyses, Klingholz et al looked at the effects alcohol had on fundamental frequency (F_0), signal to noise ratio (SNR), F_1/F_2 (the ratio of the first formant to the second formant), variation speed of the frequencies F_0, F_1, F_2 and the long term averaged spectrum (LTAS) These parameters were examined for their suitability in differentiating between sober and intoxicated conditions SNR and F_0 distributions as well as the LTAS discriminated with an error rate less than 5% F_1/F_2, describing the articulation, varied strongly among individuals [3] Findings from Dunker and Schlosshauer which indicate that vocal fold vibration after alcohol consumption (like vocal fold vibration for people with hoarse voices) is more variable and lower in pitch [4]

Sobell and Sobell in their analyses found no significant influence of intoxication on pitch [5] Trojan and Kryspin – Exner found that pitch effects varied from speaker to speaker but reported a general increase in fundamental frequency (F_0) as alcohol consumption increases [6]

The majority of studies conclude that pitch variability increases with alcohol consumption There are two possible reasons for this The first reason is that intoxicated speech may involve more extreme intonations than normal speech Secondly, intoxicated speech may involve increased period to period variation or vocal jitter Both the above effects will produce an increase in the variation of fundamental frequency

All previous studies have found that the rate of speech decreases (segmental durations increase) after alcohol consumption

The work reported in this thesis was carried out as a result of a garda inquiry The gardai had recordings of two different telephone conversations It was obvious from one of these recordings that the person speaking was intoxicated in some way The gardai wanted to know if it was the same person speaking in both recordings Before any comparisons could be made between speakers, the effects that alcohol has on the human voice had to be determined

The principal focus of this analysis was to determine what change the consumption of alcohol has on the fundamental frequency and subsequent formant frequencies of human speech The effects that alcohol consumption has on the duration of spoken sentences has also been investigated The application of this kind of analysis in forensic investigations as a means of determining whether or not an individual is sober or intoxicated is considered

Chapter 2: Materials and Methods

2.1 Hardware

The basic requirements for carrying out speech analysis research are a personal computer (PC) with an additional plug-in input/output module The I/O module used here is a variable sampling frequency, 14- bit resolution PCL-814 (Integrated Measurement Systems, Southampton)

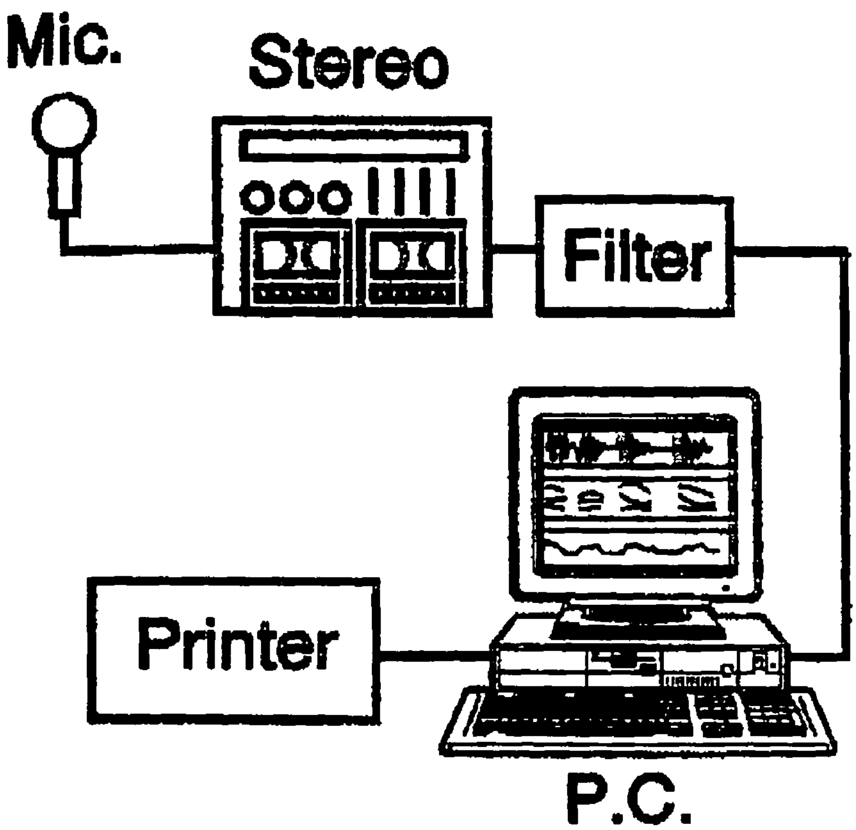


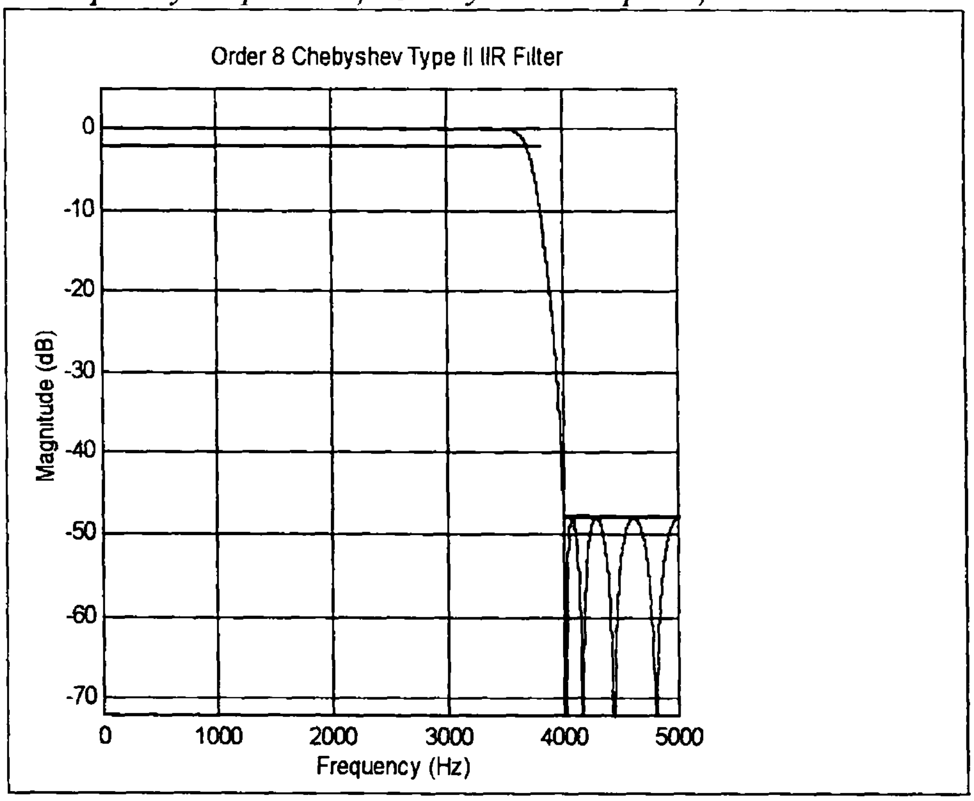
Fig 2.1. *Schematic of speech analysis system*

analogue to digital converter labcard in a 486-PC clone

Also, recording equipment that will give speech recordings of sufficient quality is required (see fig 2 1)

A Twin Cassette Deck Tape Recorder (Pioneer Electronic Corporation, Japan) with a Linear Dynamic Microphone (Sony Corporation, China) was used to record all speech samples The speech samples were recorded on TDK chrome tape cassettes of 57dB signal to noise ratio The speech samples were then played back into the input of the A/D card described earlier Each digitized voice sample corresponded to seven seconds of speech digitized at a sampling rate of 10kHz An eighth order Chebychev low pass filter [7, 8] with a cut-off frequency at 3 8kHz and with a rate of roll-off of -48 dB per octave and 2 dB ripple across the pass band was implemented to overcome the alias frequencies Fig 2 2 shows the frequency response of the Chebychev low pass filter used

Fig 2.2. *Frequency response of Chebychev low pass filter*



The driver for the analogue to digital board was a "Turbo C ++" (Borland, Scotts Valley, California, U S A) program module The digitized speech samples were then saved to disk in binary file format for analysis The subsequent analysis was then carried out on a Pentium P C clone

2.2 Experimental Details

In this experiment twelve volunteer subjects were divided into three groups of four and were grouped according to gender and approximate body weight

A text was prepared which required the subject to phonate the vowel sound /a / ('aaah') for five seconds Each subject was then asked to read the following randomly chosen phonetically balanced sentences twice,

"Joe took father's shoe bench out"

"You and I have to go today"

"He gave me a card"

Each subject was then asked to utter the isolated words "me", "you" and "to", three times each

A recording was taken of each subject before any alcohol was consumed and these recordings were used as the "normals" for the experiment

Alcohol was supplied in the form of the spirits gin and vodka Each individual was required to consume at least one measure of alcohol per half hour (in the Republic of Ireland spirits are sold in ¼ gill measures which correspond to 35 5 ml) Voice recordings were taken at these half hourly intervals using the analysis equipment outlined in paragraph 2 1 At each recording the cumulative volume of alcohol consumed was recorded

2.3 Software and Analysis

All the numerical analysis carried out in this project was done in the Matlab (The Maths Works Inc , Natick, Mass , U S A) programming environment The Matlab Signal Processing Toolbox was incorporated into the system The digital signal processing functions of this toolbox enhanced the analysis undertaken

A range of numerical techniques, such as the Fourier Transform, were implemented in the analysis of the speech data files These shall now be described in greater detail

2.3.1 Fourier Transform and Fast Fourier Transform

The Fourier Transform can be described by the following general equation

$$H(f)=\int_{-\infty}^{\infty} h(t)\, e^{2\pi i f t}\, dt \qquad (2\ 1)$$

H(f)= transformed signal

h(t)= signal function

f= frequency

t= time

i= complex number

Fourier transformation is used to convert the amplitude versus time signal into the frequency domain By calculating the Fourier transform of a section of the speech waveform, spectrum analysis can be achieved Although the calculation of the transform is essentially a discrete process, giving the result corresponding to a particular time segment, it is possible to get the successive analysis of the signal as close in time as desired by suitable overlapping of the analysis segments A signal that is sampled will have a finite number of points and the transform of the data will be discrete and windowed A Discrete Fourier Transform (DFT) is used for this purpose However, due to the fact that the DFT is a computationally taxing method, the Fast Fourier Transform (FFT) is used for such calculations The MATLAB FFT() function is used for this purpose This function uses a quick FFT algorithm if the length of the array is an integer power of two If the length of the array is not an integer power of two a slower algorithm is used

If Y= fft(x), then the fft of the data x is given by,

$$Y(k)=\sum_{j=1}^{N} x(j)\,\omega_N^{(j-1)(k-1)} \qquad (2\ 2)$$

where

$$\omega_N=e^{-2\pi i/N}$$

is an N-th root of unity (9)

Execution time for the FFT function is dependent upon the sequence length The FFT of the sequence is computed quickly if the length of the sequence has many prime factors Execution is slower if the sequence length has few prime factors In this case the Matlab FFT() function uses the slow DFT algorithm Hence the reason that it is usually better to carry out the FFT on sequences that have a length that is a power of two

The FFT() function returns a complex data array of length N

The intensity of all frequency components (f_n) is determined by calculating the absolute magnitude of the data array The intensity of the frequency components (f_n) is given by the following equation,

$$f_n = n.f_s/N \qquad (2\ 3)$$

where

$n=0,1,2, \quad ,N-1$

N= number of points

f_s= sampling frequency

and the length of the array is N/2

So that we don't misinterpret frequencies when sampling a signal, the signal must be sampled at a frequency of at least twice the highest frequency of interest If the signal is not sampled fast enough the frequency components that are close together become indistinguishable and 'aliasing' occurs According to the Nyquist criterion, the range of frequency components detected in the signal is between zero and half the sampling frequency [10]

2.3.2 Windowing

The choice of 'window' applied to the signal is also important The application of a suitable windowing function reduces the effect of the discontinuities at the ends of the signal and hence the subsequent analysis will be sharper

When sampling a single frequency wave we only sample for a limited period of time The frequency analysis is then carried out on that set of data The FFT assumes that this set repeats itself infinitely and because of this there can be discontinuities in the data There is an alternative view which is that this is an infinite sine wave multiplied by a rectangular window, that is, a function that is zero outside the sample time and equal to one inside it

What we would see as a result of the FFT is the analysis of an infinite wave multiplied by the analysis of a rectangular window The effect of this is the broadening of the peaks of the analysis The signal's spectrum is contaminated by the frequency components of the rectangular pulse If sinusoids that were widely separated in the frequency domain were being analysed then this would not really present a problem However, if the sinusoids were close together, or if you wanted information about the bandwidths of the components, then it would be desirable to eliminate this effect In order to overcome this problem, care must be taken in choosing a suitable windowing function

The window used in this application is a Hamming window whose function is

$$w[n] = \begin{cases} 0\ 54 - 0\ 46\cos(2\pi n/m), & 0 \le n \le m, \\ 0, & \text{otherwise}, \end{cases} \qquad (2\ 4)$$

where

$n=1,2, \quad , m$

$m=$ number of points

Fig 2 3 shows a graph of the Hamming window function and the Fourier Transform of a sinusoidal signal with a frequency of 2kHz using a rectangular window function and also the Fourier Transform of the same signal using a Hamming window function It can be clearly seen from both of the Fourier Transform graphs that the dominant frequency is 2kHz There are other frequency components present but when the Hamming window is applied, these are nearly all below zero

There are several types of windowing functions that could have been used in this application, for example, a Hanning or Bessel window could have been used These, along with the Hamming window, are standard windowing functions However, both the Hamming and Hanning windows are much more computationally efficient than the slow Bessel window even though this type of window gives less transmission in the sidelobes All available windowing functions will give some transmission in the sidelobes (at least 9% approx) due to the Gibbs effect [11] A Hamming window was chosen for this application because it presents the 9% transmission in the sidelobes in the least obtrusive way of all similar windowing functions Sidelobes for this window are attenuated by 40 dB [12], which is sufficient for most speech processing applications Even though a Hanning window works as quickly as a Hamming window and has fewer sidelobes , these sidelobes have a

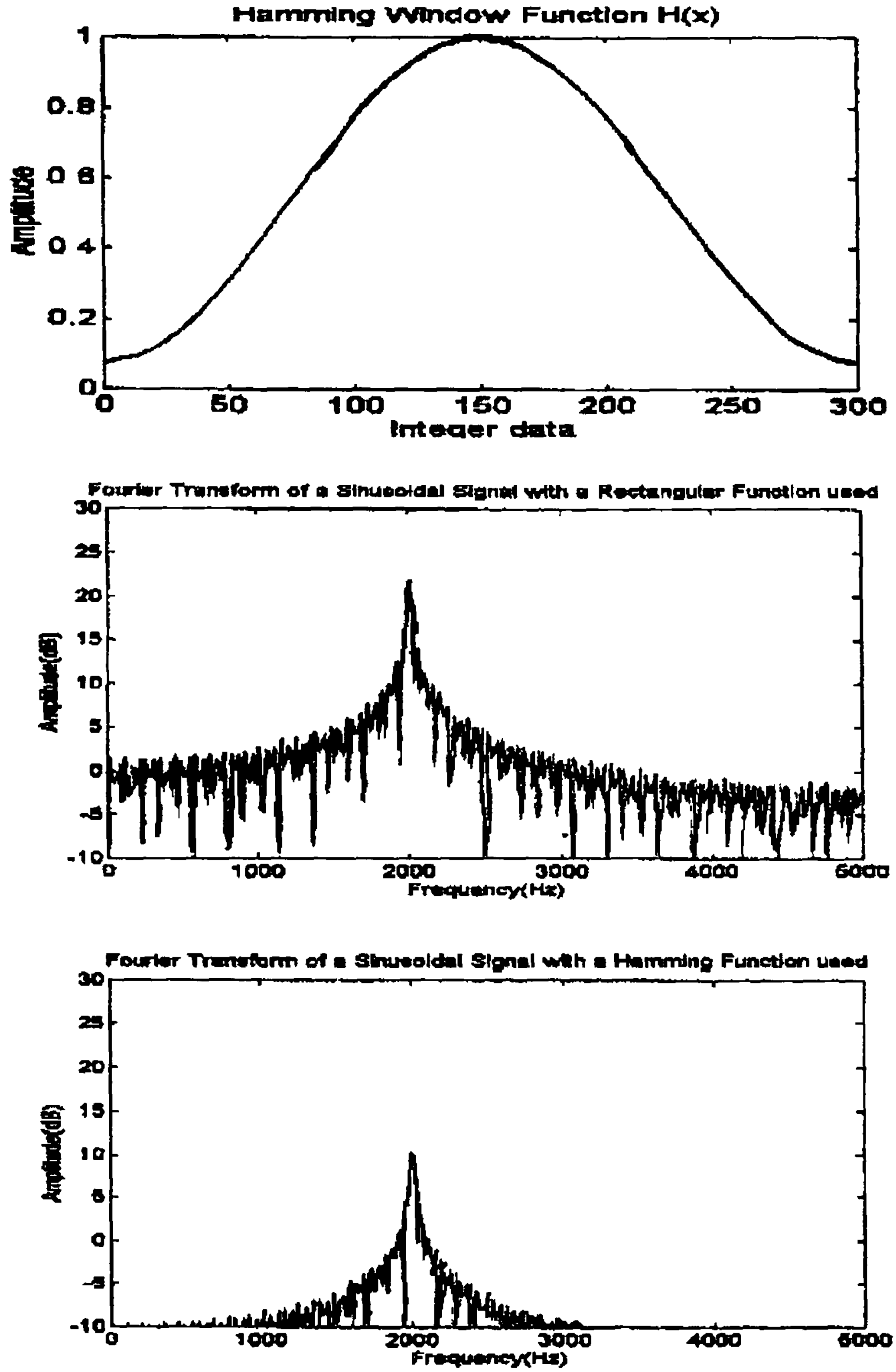


Fig. 2.3. *Hamming window function and F T 's of sinusoid using rectangular window and Hamming window respectively*

higher transmission than the sidelobes of a Hamming window and hence the frequency response is not as smooth as that for a Hamming window

2.3.3. Cepstral Analysis

Determination of the fundamental frequency was carried out by means of a cepstral analysis The cepstrum is defined as the Fourier transform of the logarithm of the power spectrum Fourier transformation of the log spectrum gives a new independent variable called quefrency whos dimension is time

A cepstral analysis can extract the fundamental frequency from a set of spectral data A spectrum consists of high frequency and low frequency data and can be represented in the frequency domain as the product of two functions,

$$S\,(e^{i\theta}) = H\,(e^{i\theta})\,E\,(e^{i\theta}) \qquad (2\ 5)$$

where

$H\,(e^{i\theta})$ = low frequency data

$E\,(e^{i\theta})$ = high frequency data

By taking the logarithm of the spectrum we obtain,

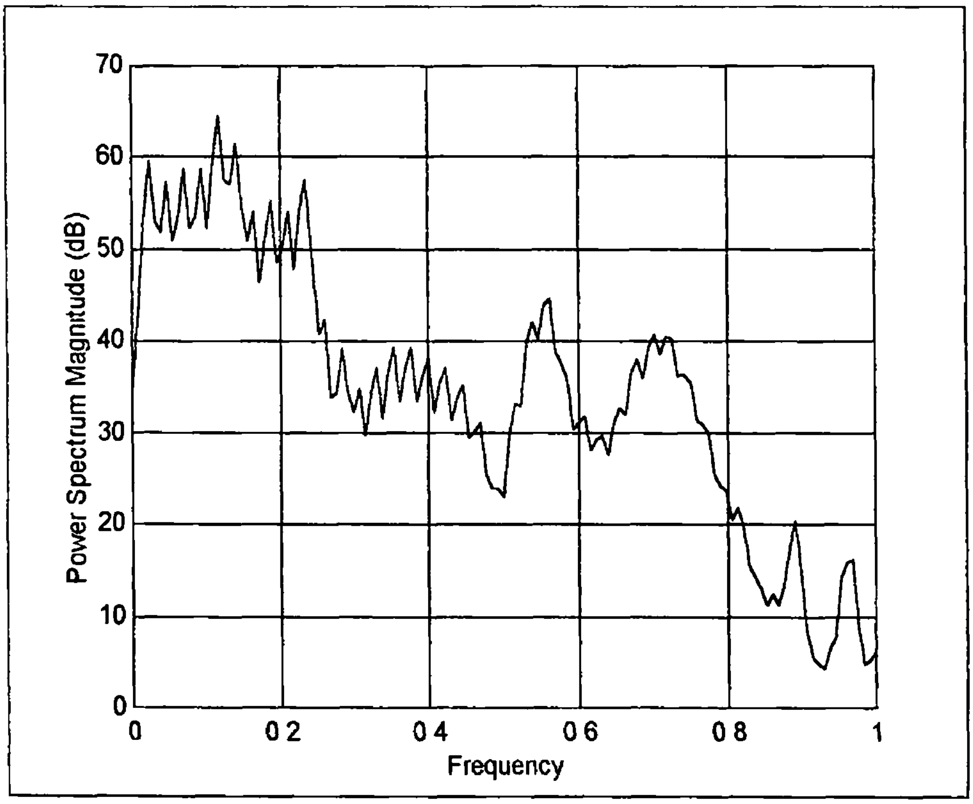
$$\log\,[S(e^{i\theta})] = \log\,[H(e^{i\theta}) * E(e^{i\theta})] \qquad (2\ 6)$$

which becomes,

$$\log\,[S(e^{i\theta})] = \log\,[H(e^{i\theta})] + \log\,[E(e^{i\theta})]. \qquad (2\ 7)$$

The low frequency and hıgh frequency components of the spectrum have now been separated ınto two addıtıve components An Inverse Fourıer Transform wıll separate these components

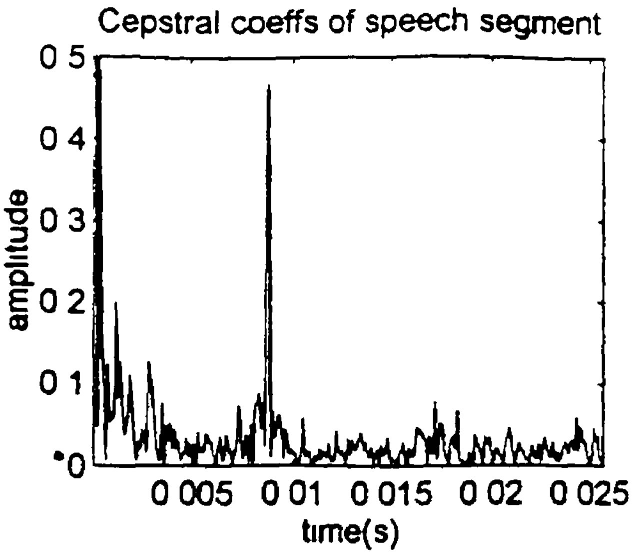
Fıg. 2.4. *FFT spectrum before cepstral anaysis*



It wıll return hıgh amplıtudes at places correspondıng to these frequency components From thıs the cepstral analysıs ıs obtaıned Fıgure 2 4 shows a graphıcal example of an FFT spectrum before the cepstral analysıs has been carrıed out

The vocal chord frequency ıs represented on the cepstrum by lınes at a quefrency equal to the glottal perıod and at frequencıes whıch are multıples of thıs (harmonıcs)

Fig. 2 5 *Cepstral plot of frequencies shown in Fig 2 4*



Therefore the fundamental frequency can be easily calculated by inverting the glottal period (Fig 2 5)

The sequence of commands used in Matlab in order to obtain the spectrum and real cepstrum of a signal x are,

y = real (ifft (log (abs(fft(x)))))

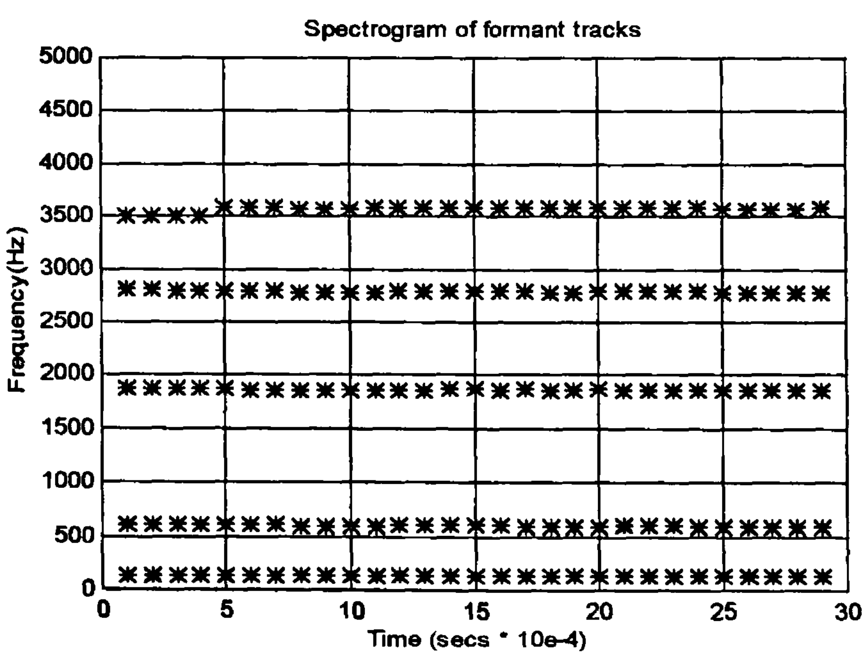
The program fundfr m was written for the purpose of extracting the fundamental frequency from a particular speech sample In running this program a sonogram of the speech sample of interest is plotted from which a section of the sample is then extracted and the cepstral analysis is carried out on this portion of speech This program was applied to segment lengths of 400 points, which corresponds to 40ms of speech, with an overlap of 30 points In order for the cepstrum calculation to work at least 300 points are required This data was then padded with zeros to a length of 512 ın order for the fft calculatıon to be more efficıent Matlab codıng and algorıthms for thıs program can be found ın Appendıx A

Fundamental frequency calculatıons were carrıed out on the sustaıned vowel phonatıon /a / for each subject when they were sober and at dıfferent levels of alcohol consumptıon The same ıs true for the ısolated word utterances "me", "you" and "to" and also the utterance of these words ın spoken sentences

2.3.4. Formant Frequency Analysıs

A Matlab program was wrıtten for the purpose of detectıng peaks ın a smoothed set of data The basıc prıncıple behınd thıs program ıs that a peak poınt ıs detected ıf there ıs a change ın slope from posıtıve to negatıve about that poınt

Fig 2.6. *Vısual dısplay of formant frequencıes plotted agaınst tıme*



Using this program, formant m, the digitised speech samples are segmented, windowed (Hamming) and Fourier transformed The program returns a visual display of the formant frequencies plotted against time Four frequency values are then picked from each of the first four formant tracks via mouse control (see fig 2 6) These are then stored as a sixteen element feature vector which is displayed on the screen Matlab coding for formant m can be found in Appendix A

Formant frequency analyses were carried out on the sustained vowel phonation /a,/ for each subject both when sober and at differing levels of alcohol consumption This analysis was also carried out on the isolated word utterances “me, “you” and “to” and the utterance of these words within a spoken sentences

2.3.5 Sentence Duration Analysis

The program dur m (Matlab coding available in Appendix A) was used to calculate the duration of the sentences outlined in section 2 2 The time taken for one element of digitised speech was calculated and hence the duration of a speech sample could be calculated by determining the number of elements in that particular sample of speech

Once again, this analysis was carried out on the sentences for each subject both when sober and at various levels of alcohol intoxication

Chapter 3: Results

3.1. Sustained Vowel Phonation and Isolated Utterances

The details of the subjects who participated in the experiment and the volume of alcohol consumed by each individual during the study are listed in Table 3 1 Figure 3 1 shows plots of f_0 for the sustained vowel utterance /a / versus volume of alcohol consumed This figure shows variations in f_0 do occur with alcohol consumption However, no consistent inter-speaker trends are observed Figures 3 2, 3 3 and 3 4 show similar plots of f_0 versus volume of alcohol consumed for the phonated vowel sounds of isolated utterances 'me', 'you' and 'to' respectively There are no consistent inter-speaker trends for these isolated utterances

First order regression analysis of the variation of f_0 with alcohol consumption for these isolated utterances and for the sustained vowel utterance /a / has been carried out and the results are listed in Table 3 2 Again no consistent inter-speaker trends are evident but considerable intra-speaker variation is observed Only three of the subjects display a uniform change in f_0 for all four utterances studied and two of these are increases (speakers 8 & 10) while the other is a decrease (speaker 2)

Table 3 1. *Details of the subjects who participated in the experiment*

Speaker No / Symbol	Group	Sex	Age (yr)	Weight (kg)	Alcohol Consumed (ml)	Group Average		
						Age (yr)	Weight (kg)	Alcohol Consumed (ml)
1 (c)	1	M	28	72 5	426	34 2 (SD = 15 5)	76 6 (SD = 3 0)	328 (SD = 79)
2 (s)	1	M	30	79 4	284			
3 (ut)	1	M	22	78 4	355			
4 (it)	1	M	57	76 2	248			
5 (c)	2	F	23	56 8	302	26 5 (SD = 6 3)	57 3 (SD = 3 2)	240 (SD = 42ml)
6 (s)	2	F	23	59 0	231			
7 (ut)	2	F	24	53 0	213			
8 (it)	2	F	36	60 3	213			
9 (c)	3	F	28	66 7	319	24 7 (SD = 2 9)	66 0 (SD = 3 3)	310 (SD = 73ml)
10 (s)	3	F	26	69 8	319			
11 (ut)	3	F	21	61 7	390			
12 (it)	3	F	24	66 0	213			
Average			28 5 (SD = 9 91)	66 6 (SD = 8 7)	294 (SD = 72ml)			

Symbols c= circle, s= square, ut= upright triangle, it= inverted triangle

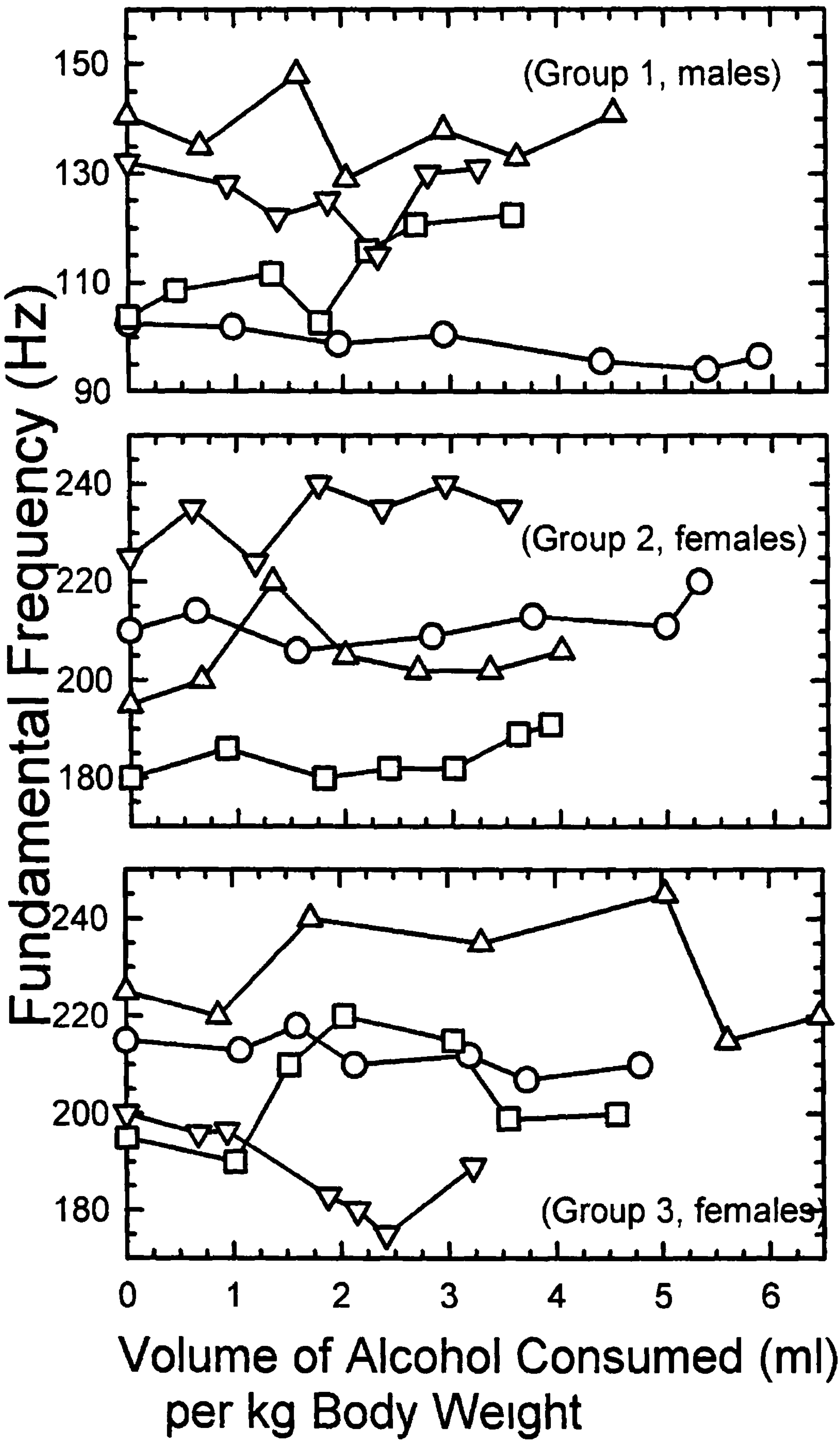


Fig 3.1. *F_o versus Alcohol Consumed for sustained vowel /a /*

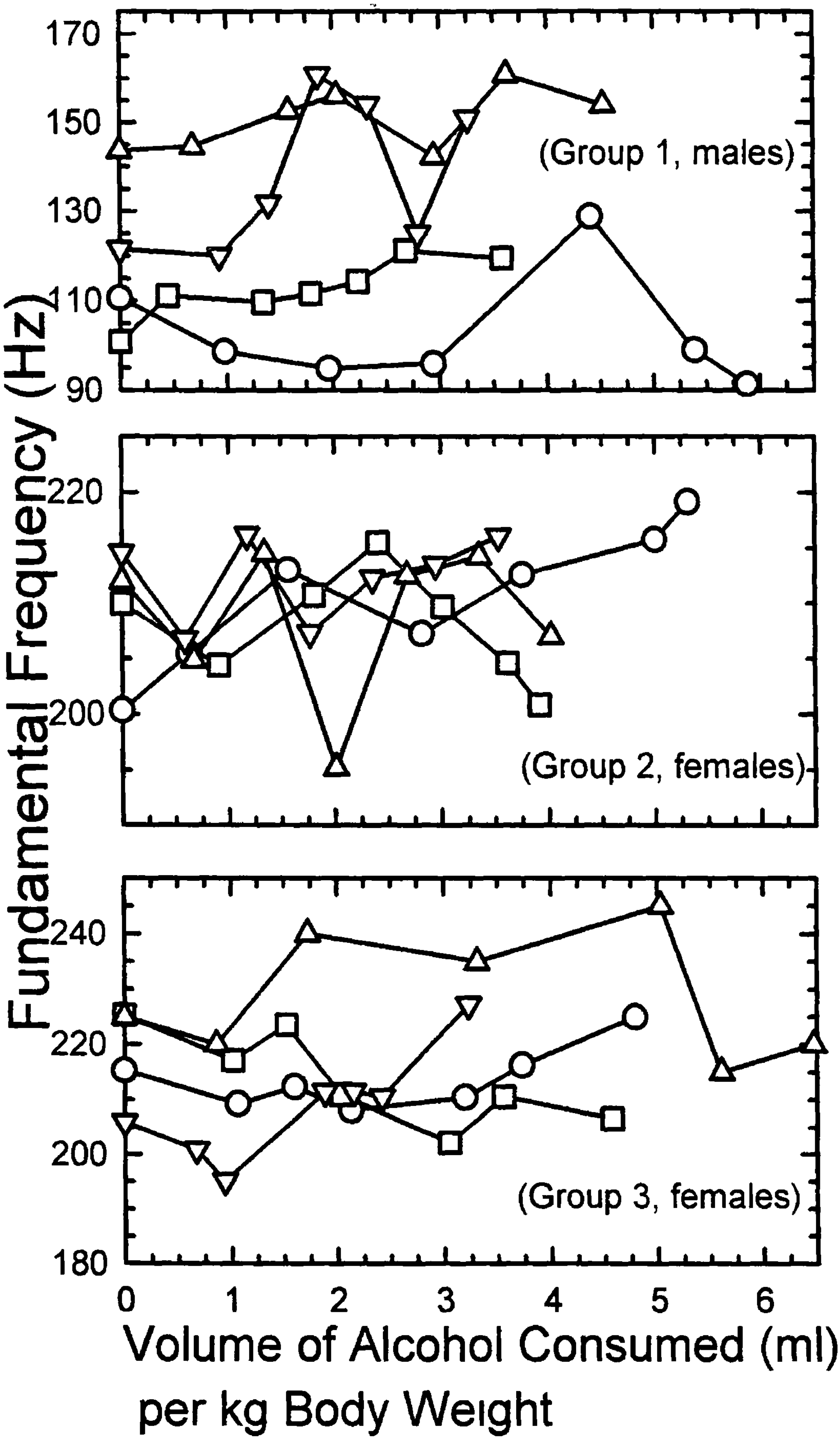


Fig 3.2. *F_o versus Alcohol consumed for isolated word 'me'*

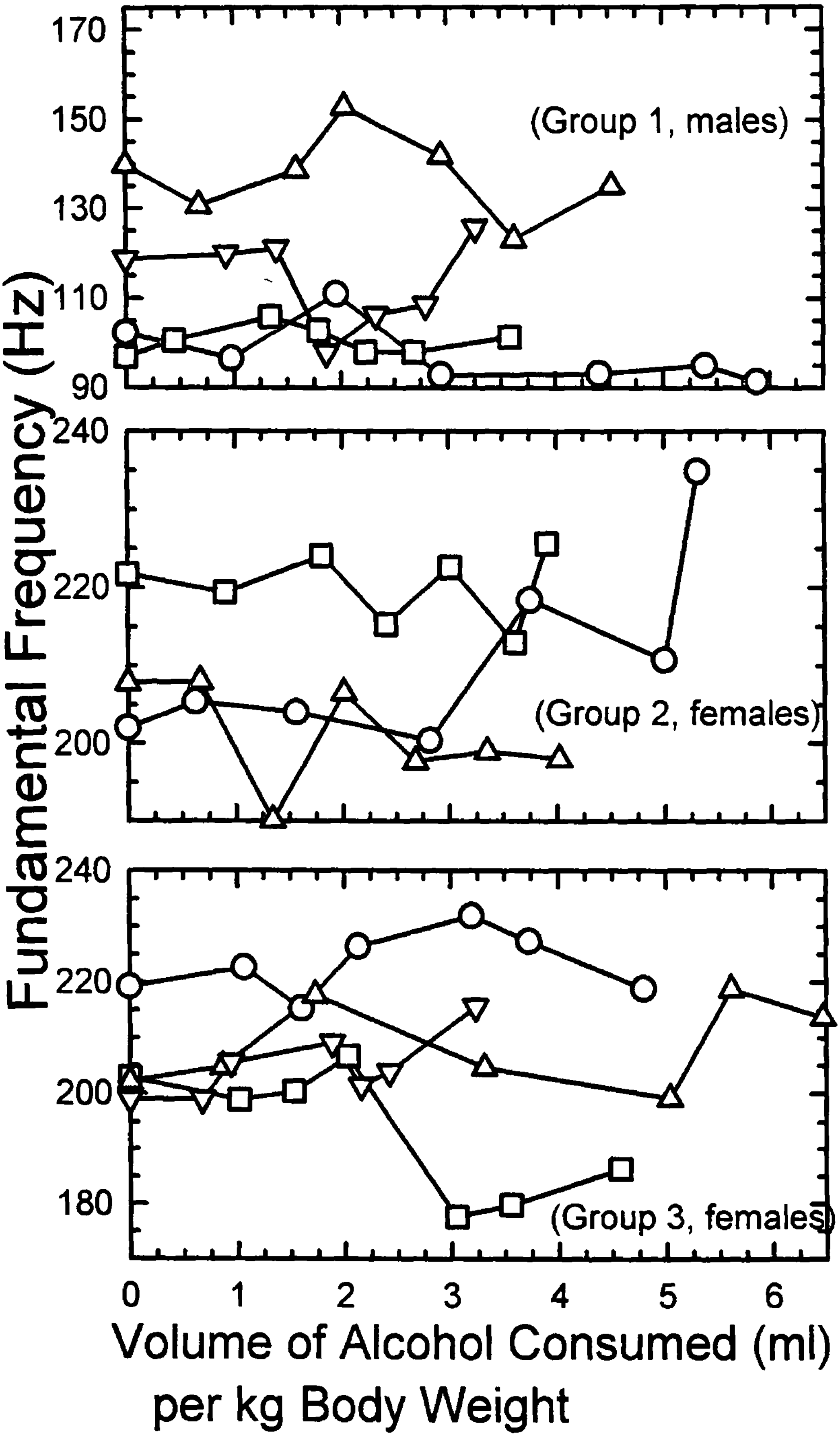


Fig 3 3. *F_o versus Alcohol consumed for isolated word 'you'*

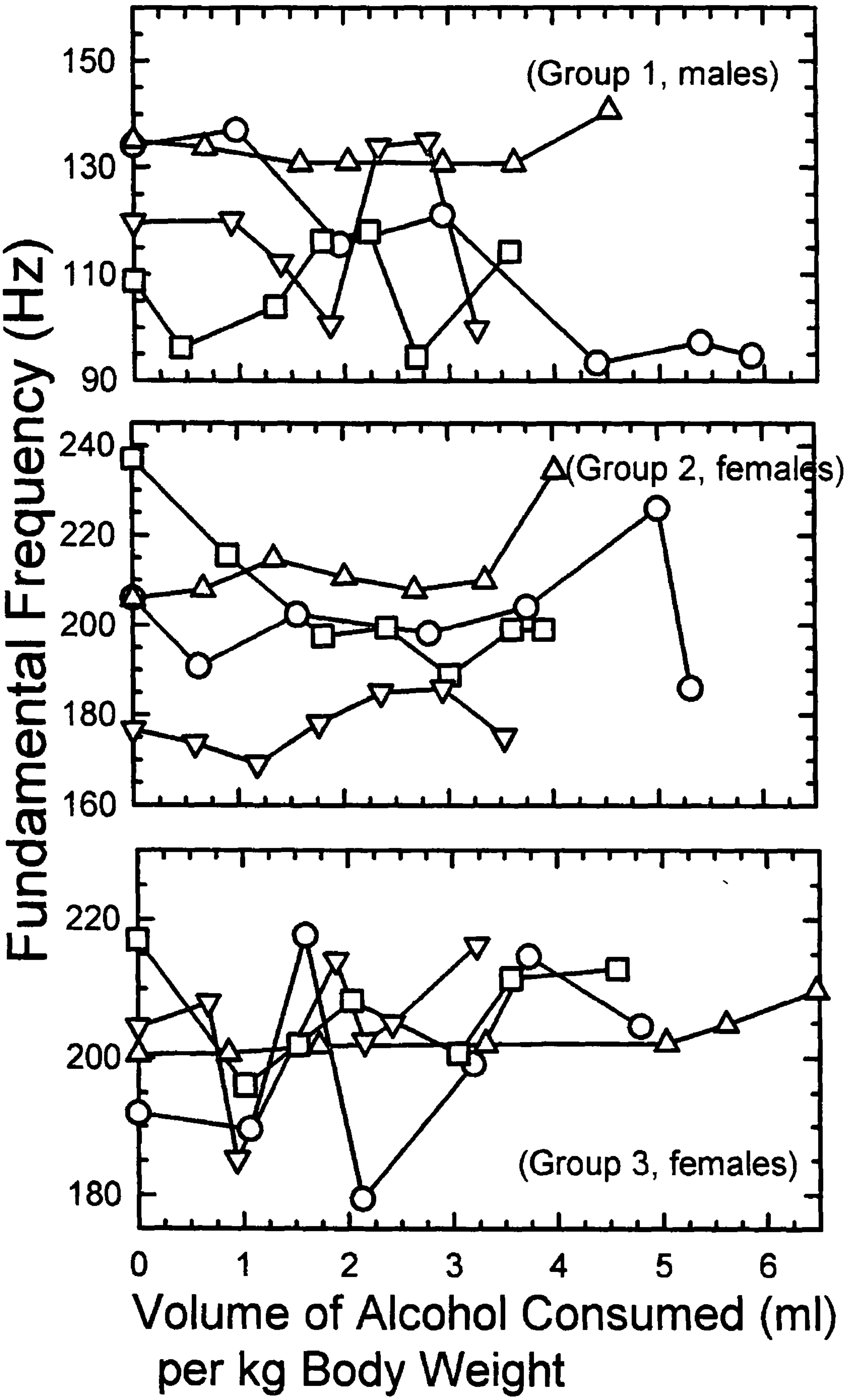


Fig 3.4. *F_o versus Alcohol consumed for isolated word 'to'*

F_o data were analysed using a random effects general linear model (sometimes referred to as a repeated measures by nesting design) [13] In this design each subject is treated as an experimental block drawn from a population of such blocks In addition to alcohol level, speaker ID and sex were included as factors

F_0 was normally distributed within each sex (correlation with the expected normal distribution was 0 978 for men and 0 991 for women) As would be expected both sex and speaker ID were strongly associated with f_0 There was no association between f_0 and alcohol level The regression coefficient was 0 146(F = 0 078, df = 71, p= 0 781) The value of 'p' must be less than 0 005 for a result to be of any statistical significance The parameter 'F' reflects the complexity of the prediction factor with df, the number of degrees of freedom, reflecting the sample size and the total amount of variation in the data For example, a value of F=2 0 on 1 and 320 df means that the results obtained from the statistical analysis are twice as accurate as they would have been had no additional experimental factors been taken into consideration or if only random effects had been considered The value of df indicates the amount of extra factors that are considered Model residuals were examined and correlated highly with the expected normal distribution (r = 0 995)

The effect of alcohol level on f_1 values was modelled correcting for sex and phonation as fixed effects and subject ID as a random effect There was no evidence of an interaction between alcohol and either sex or phonation, nor was there evidence of a relationship between alcohol level and f_1 value ($F = 0\ 6$ on 1 and 320 df, $p = 0\ 435$)

The effect of alcohol level on f_2 was modelled in the same way Once again there was no evidence of a significant interaction between alcohol and either sex or phonation in determining the value of f_2 Alcohol level was not associated with the value of f_2 ($F = 0\ 8$ on 1 and 320 df, $p = 0\ 374$)

Table 3.2. *Results of 1st order regression analysis of the variation of f_o with alcohol consumption*

Speaker No	Sex	Change in f_o (Hz) per ml alcohol consumed per kg body weight			
		"aah" phonation	"Me"	"You"	"To"
1	M	- 0 8	+ 4 3	+ 0 2	+ 5 3
2	M	- 0 45	- 0 45	- 2 0	- 8 5
3	M	+ 5 08	+ 5 6	- 1 33	- 0 66
4	M	- 1 29	+ 3 86	- 1 29	- 1 93
5	F	+ 1 12	+ 2 72	- 3 48	+ 4 59
6	F	+ 1 0	- 1 16	+ 1 0	+ 2 12
7	F	+ 1 79	+ 0 3	- 2 09	-8 36
8	F	+ 3 06	+ 0 85	+ 4 59	+ 1 02
9	F	- 0 38	+ 3 38	+ 3 19	+ 3 38
10	F	+ 1 57	+ 3 15	+ 4 33	+ 4 33
11	F	- 5 74	- 4	+ 1 04	+ 0 52
12	F	- 1 49	+ 1 86	- 5 21	+ 1 12

The relationship $f_2 - f_1$, the difference between the first and second formants, varied significantly by speaker, but not by sex or phonation Alcohol was unrelated to the $f_2 - f_1$ difference ($F = 2\ 0$ on 1 and 320 df, $p = 0\ 159$)

The ratio of the first formant to the second formant f_1/f_2 was unrelated to any of the experimental factors, including alcohol ($F = 2\ 3$ on 1 and 320 df, $p = 0\ 127$)

3.2. Utterance of vowel sounds within spoken sentences

The data from the utterances of vowel sounds within spoken sentences was analysed using a general linear model in which sex was entered as a fixed factor, speaker identity and utterance as random factors and alcohol as a continuous predictor

Figures 3 5, 3 6 and 3 7 show plots of f_0 versus volume of alcohol consumed for the utterance of the words 'me', 'you' and 'to' within the spoken sentences 'He gave me a card' and 'You and I have to go today' respectively F_0 has a satisfactory approximation to the normal distribution within each sex (r with expected normal distribution 0 992 in men and 0 998 in women) The linear model showed no significant interaction between sex and alcohol level in determining f_0 ($F = 0\ 19$, df = 1, 236, $p = 0\ 662$), indicating that the effect of alcohol level on f_0 did

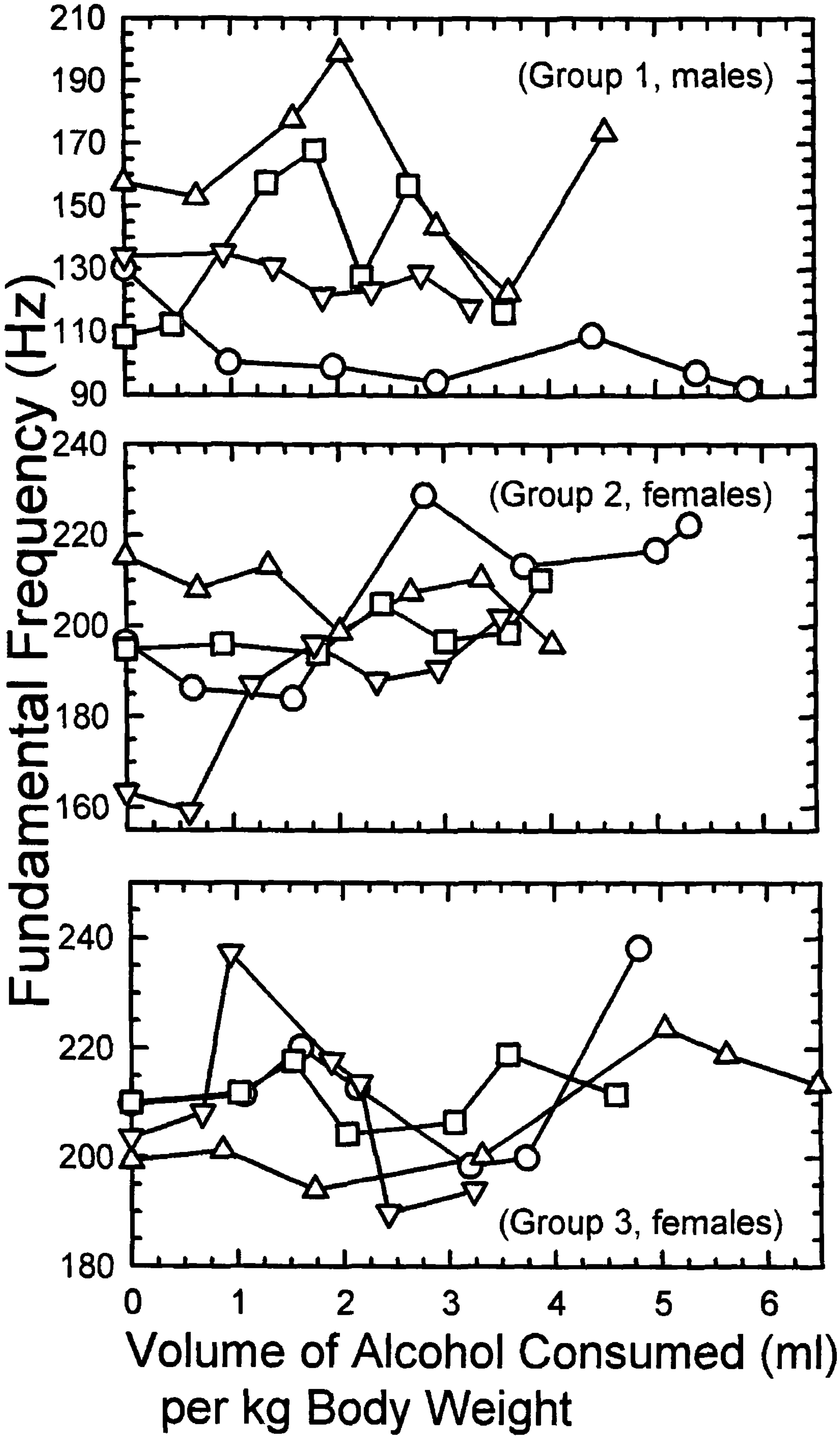


Fig 3.5. *F_o versus Alcohol consumed for 'me' spoken in sentence*

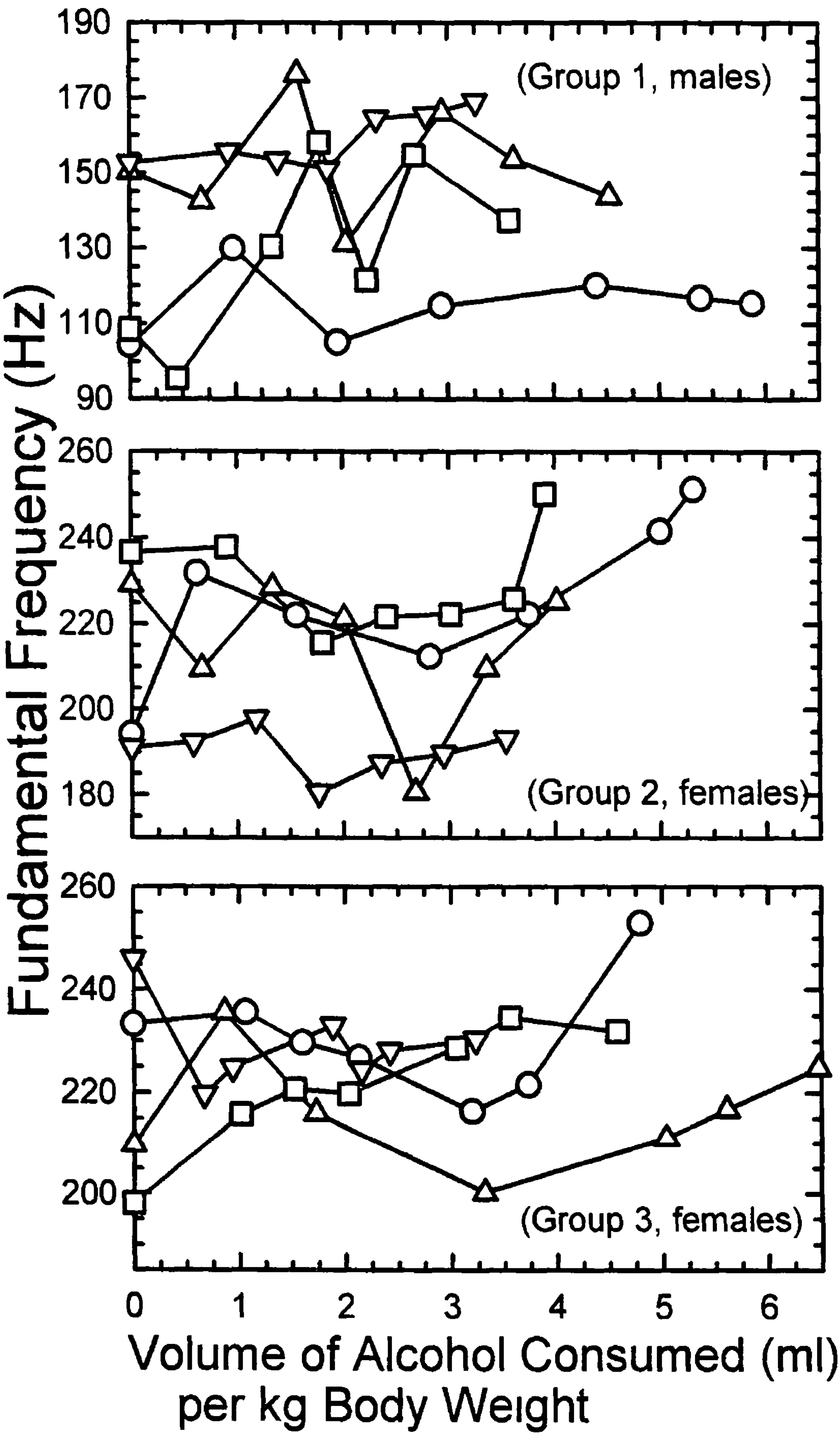


Fig 3.6. *F_o versus Alcohol consumed for 'you' spoken in sentence*

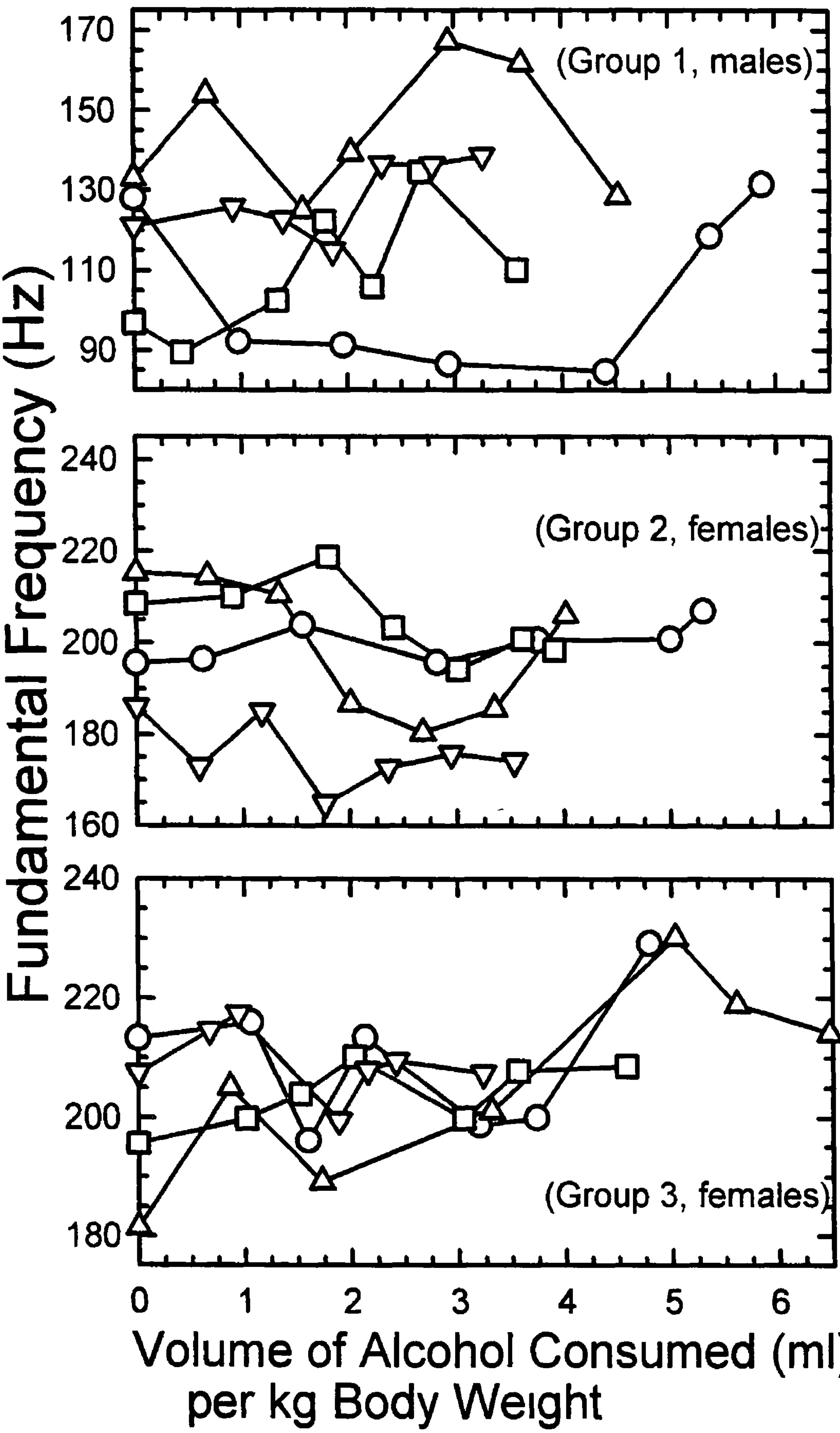


Fig 3.7. *F_o versus Alcohol consumed for 'to' spoken in sentence*

not differ between the sexes Variation between speakers and between utterances was highly significant (both $p < 0.0001$) The slope of the relationship between alcohol and f_o was also significant ($F = 8.9$ on 1 and 237 df, $p = 0.003$) The coefficient was 1 7 (SE 0 570) indicating that f_o rose by 1 7Hz for each unit increase in alcohol

The correlation between f_1, the first formant, and the expected normal distribution was 0 964 in men and 0 984 in women Again, there was no significant interaction between alcohol and sex ($F = 0.002$, df = 1, 236, $p = 0.969$) indicating that the slope of the relationship between alcohol and f_1 did not differ between the sexes The slope was essentially zero (-0 25, SE 5 1) ($F = 0.002$, df = 1, 237, $p = 0.961$)

F_2, the second formant, correlated reasonably with the expected normal distribution in men ($r = 0.964$) but poorly in women (0 924) Examination of a normal probability plot revealed that f_2 followed a truncated, but symmetrical distribution in women The interaction between sex and alcohol level was not significant ($F = 0.10$ on 1 and236 df, $p = 0.747$) It was found that there was no change in f_2 with alcohol level (slope 0 37, SE 6 9, $F = 0.002$, df = 1,237, $p = 0.958$)

$F_2 - F_1$, the difference between the first and second formants, varied significantly from speaker to speaker, but not by sex or phonation There was no significant relationship found between alcohol level and the difference between the first and second formants, $f_2 - f_1$ The slope

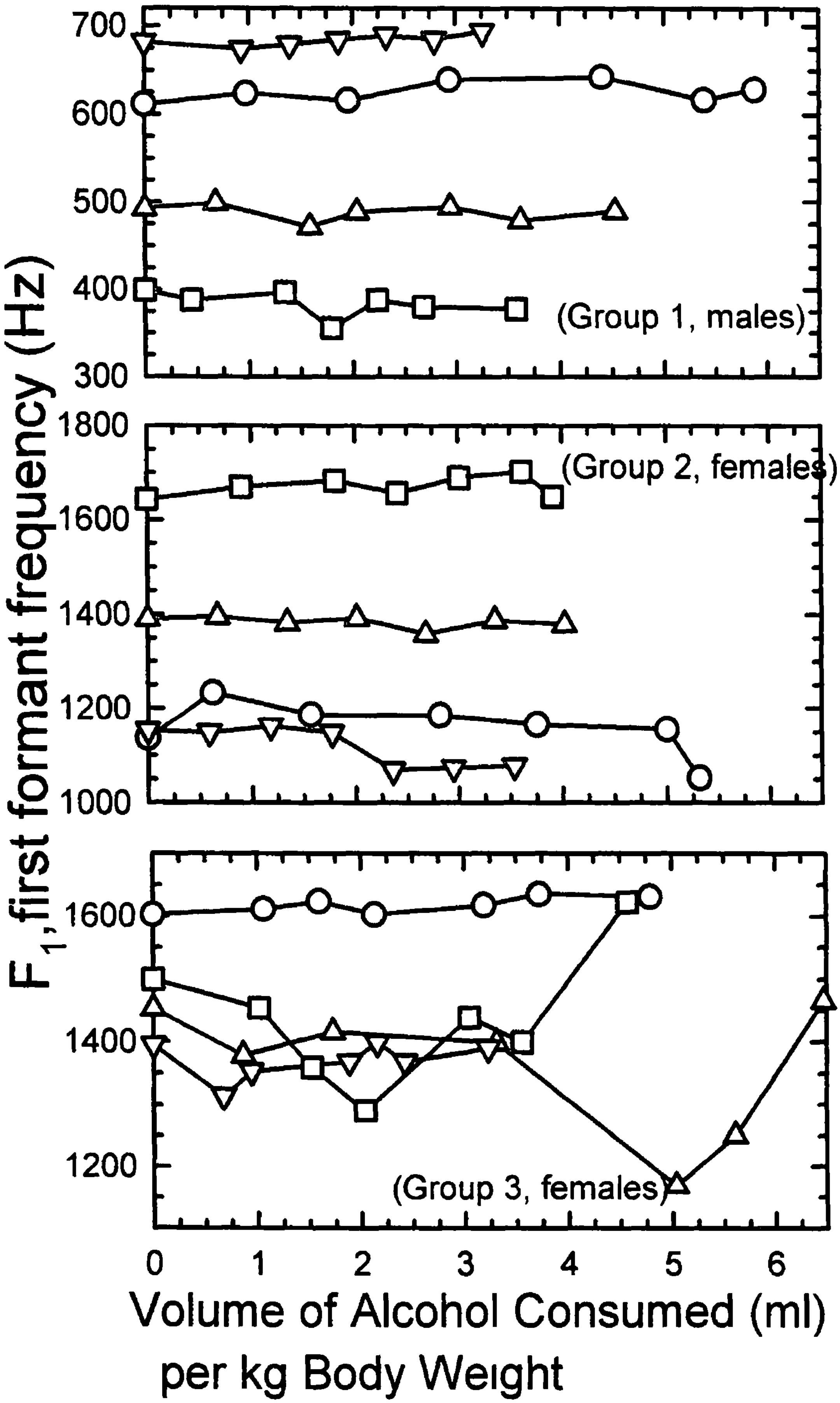


Fig 3.8. *F_1 versus Alcohol consumed for sustained vowel /a /*

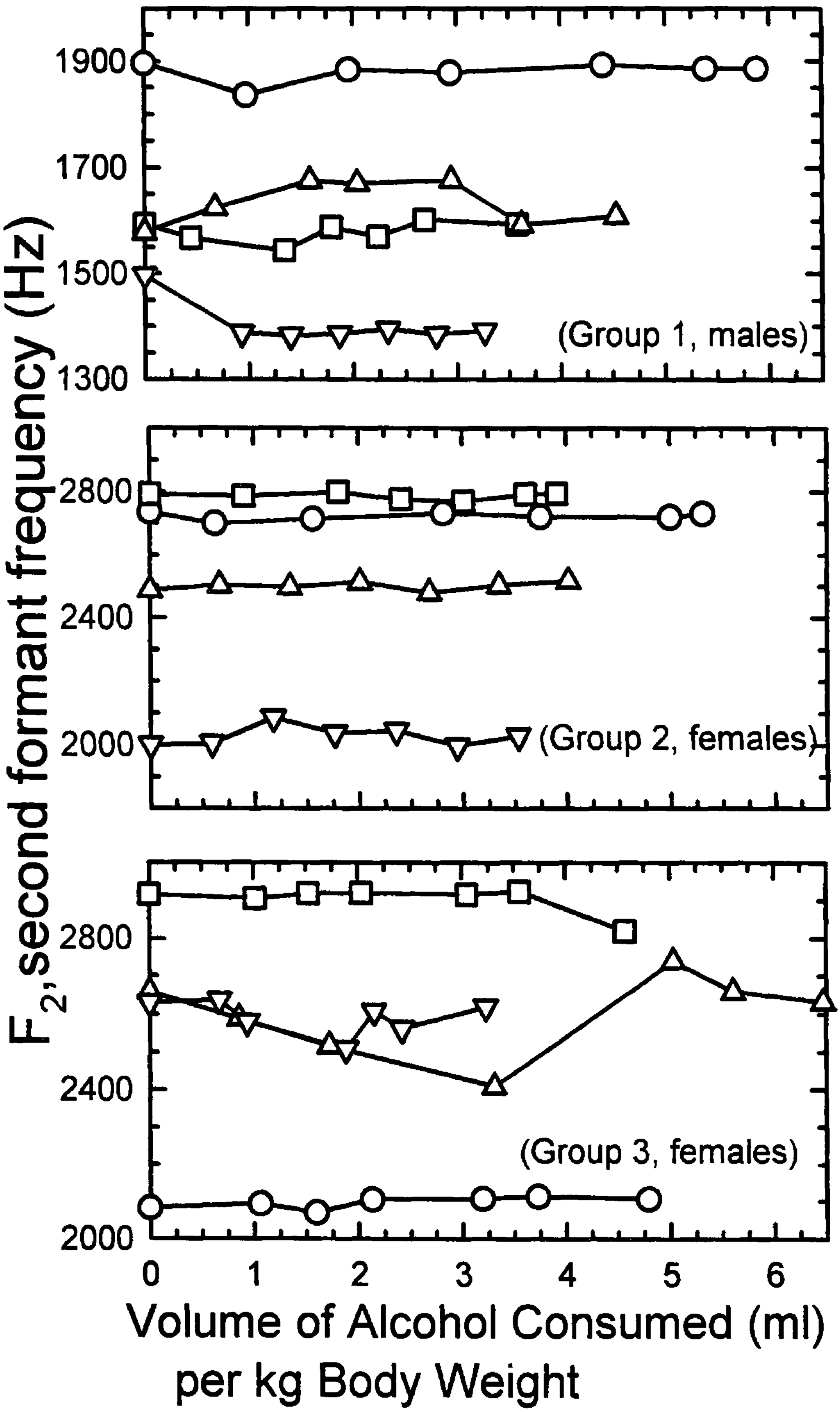


Fig 3.9. *F_2 versus alcohol consumed for sustained vowel /a /*

remains essentially zero (0 52, SE 6 2, $F = 0\ 007$ on 1 and 236 df, $p = 0\ 934$)

Alcohol level and the ratio of the first formant to the second formant, f_1/f_2, were found to be unrelated (slope 0 55, SE 5 9, $F = 0\ 008$ on 1 and 236 df, $p = 0\ 937$)

Figures 3 8 and 3 9 show plots of f_1 versus alcohol level and f_2 versus alcohol level respectively for the sustained vowel utterance /a / A similar relationship between the first and second formants and alcohol level has been found for the utterance of isolated words and the utterance of these words within spoken sentences, that is, there is no significant relationship Due to this, only the plots for the sustained vowel phonation have been included (data for the isolated word utterances and the utterance of these words in spoken sentences are available on request)

3.3. Effects of alcohol level on sentence duration

Figures 3 10, 3 11 and 3 12 show plots of duration versus volume of alcohol consumed for the sentences, 'Joe took father's shoe bench out', 'You and I have to go today', and 'He gave me a card' respectively A general increase in duration was observed as the volume of alcohol increase Figure 3 13 shows the percentage increase in duration per ml of alcohol consumed per kg bodymass for all the sentences for each individual

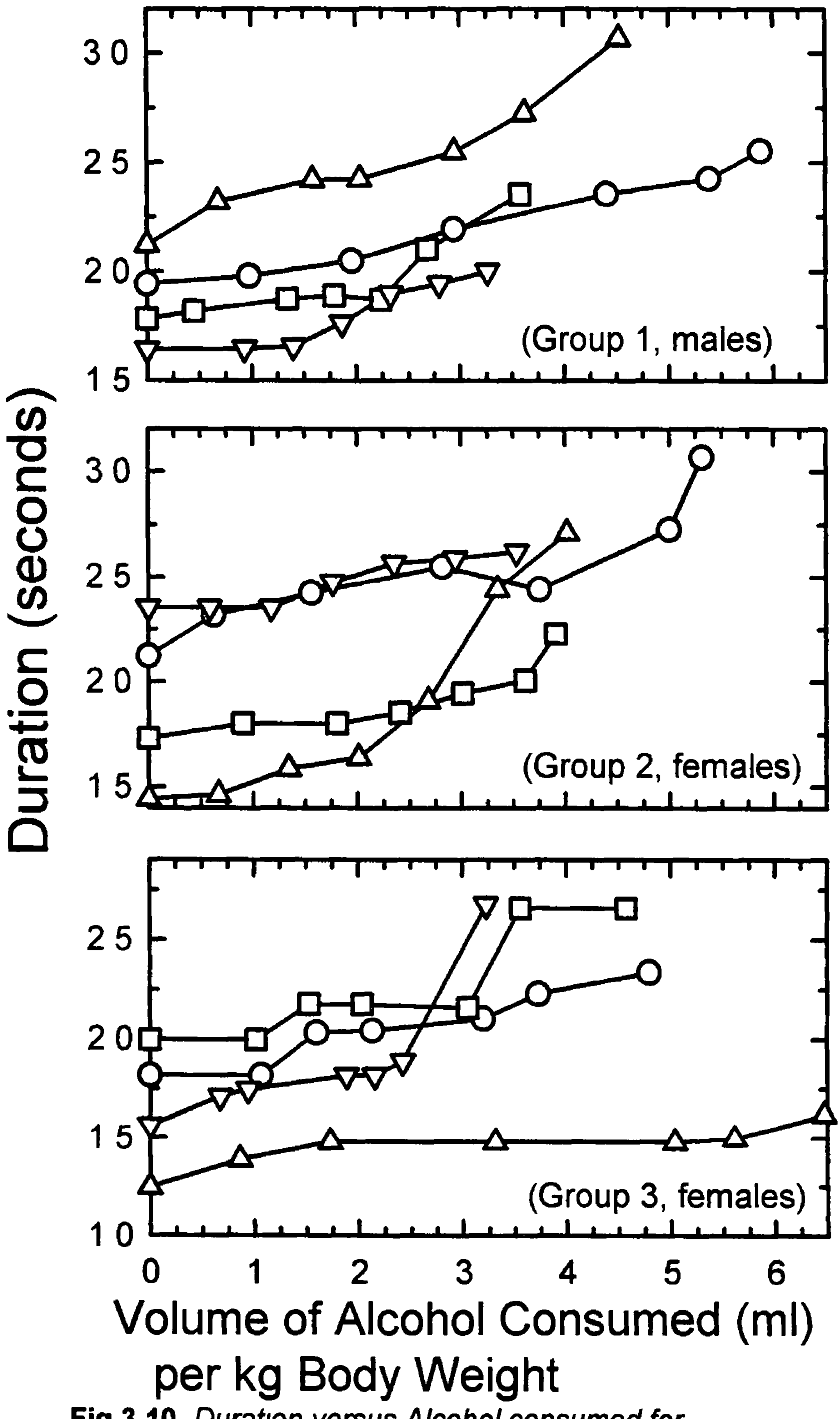


Fig 3.10. *Duration versus Alcohol consumed for* 'Joe took father's shoe bench out'

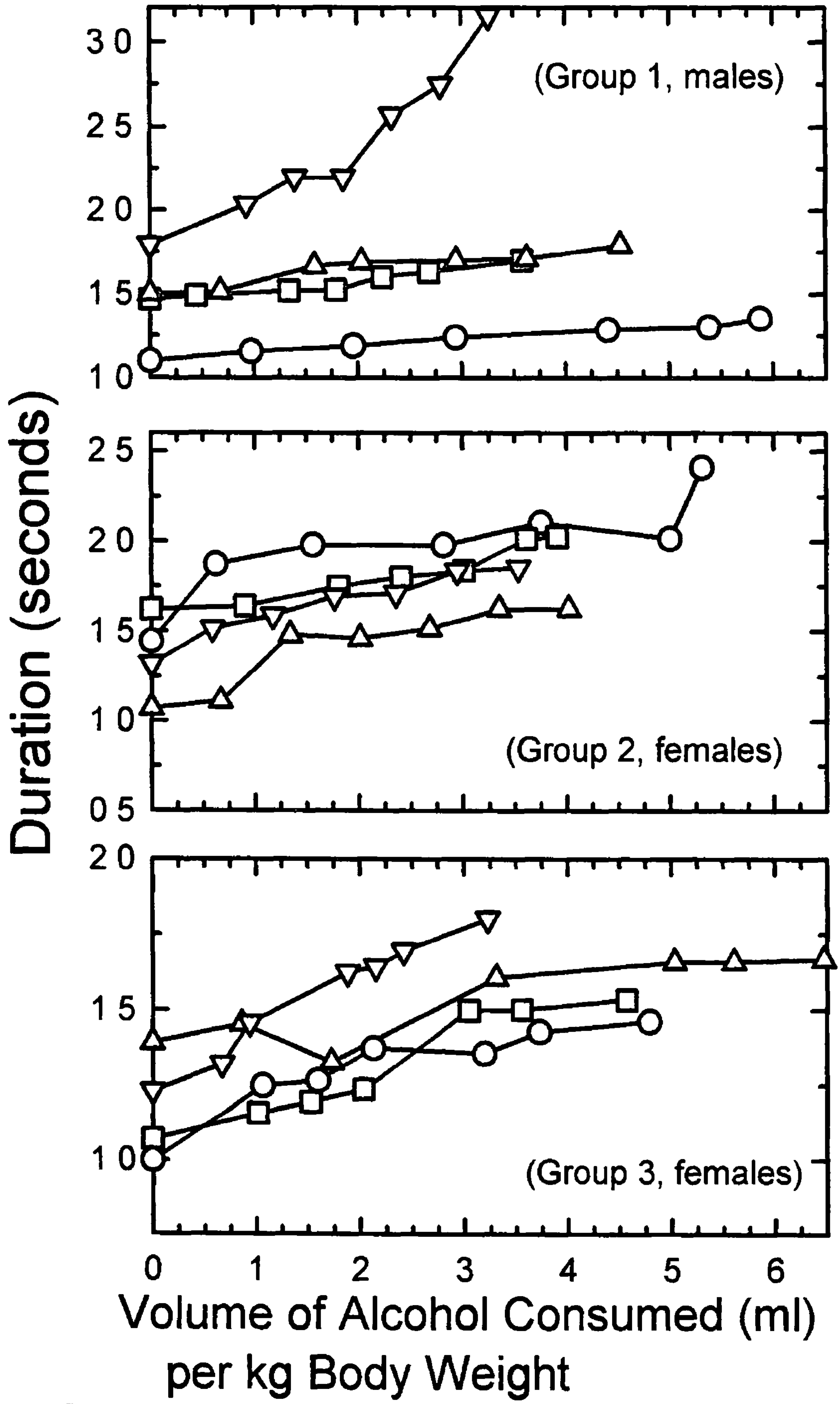


Fig 3.11 *Duration versus Alcohol consumed for 'You and I have to go today'*

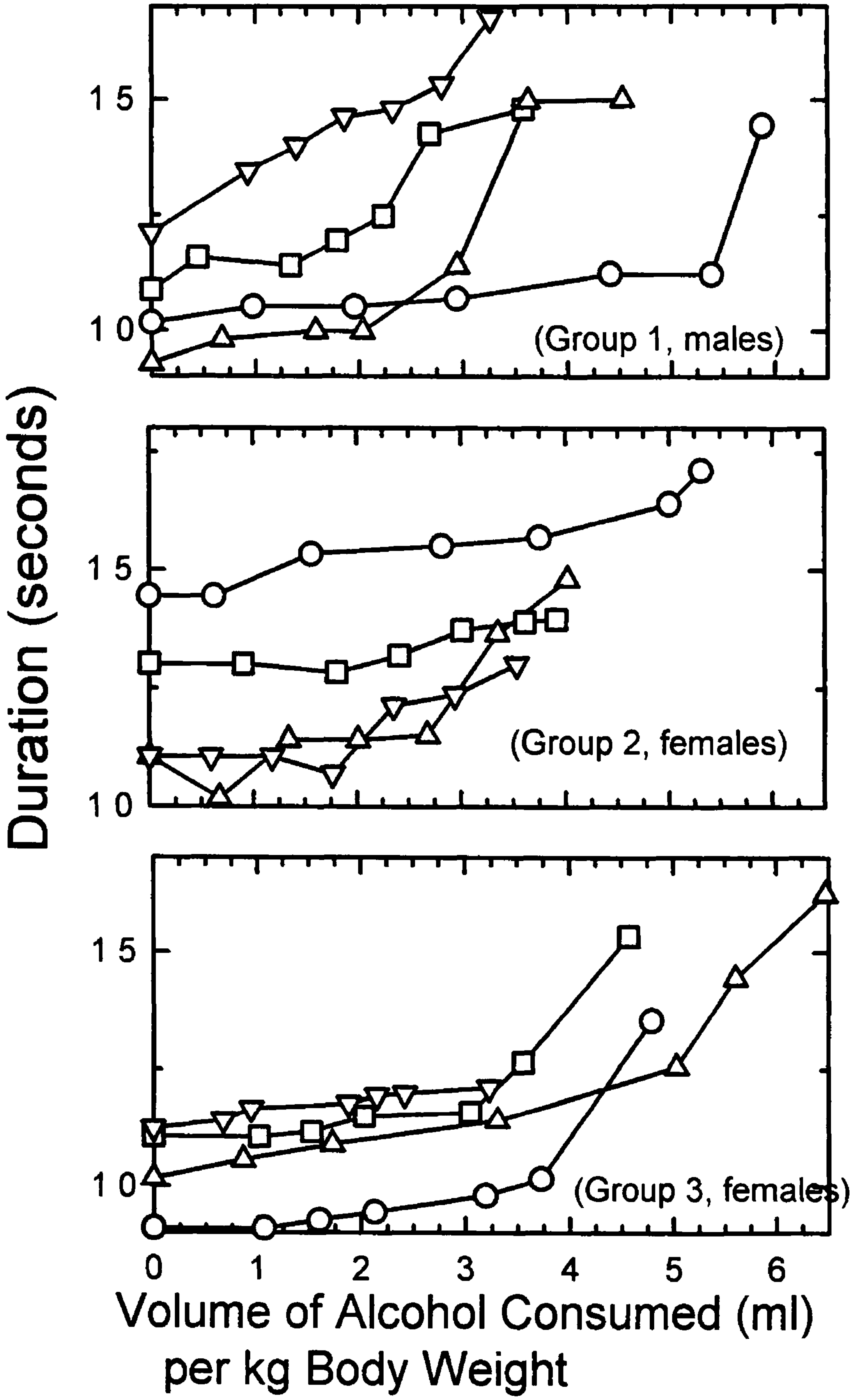


Fig 3.12. *Duration versus Alcohol consumed for*
' He gave me a card'

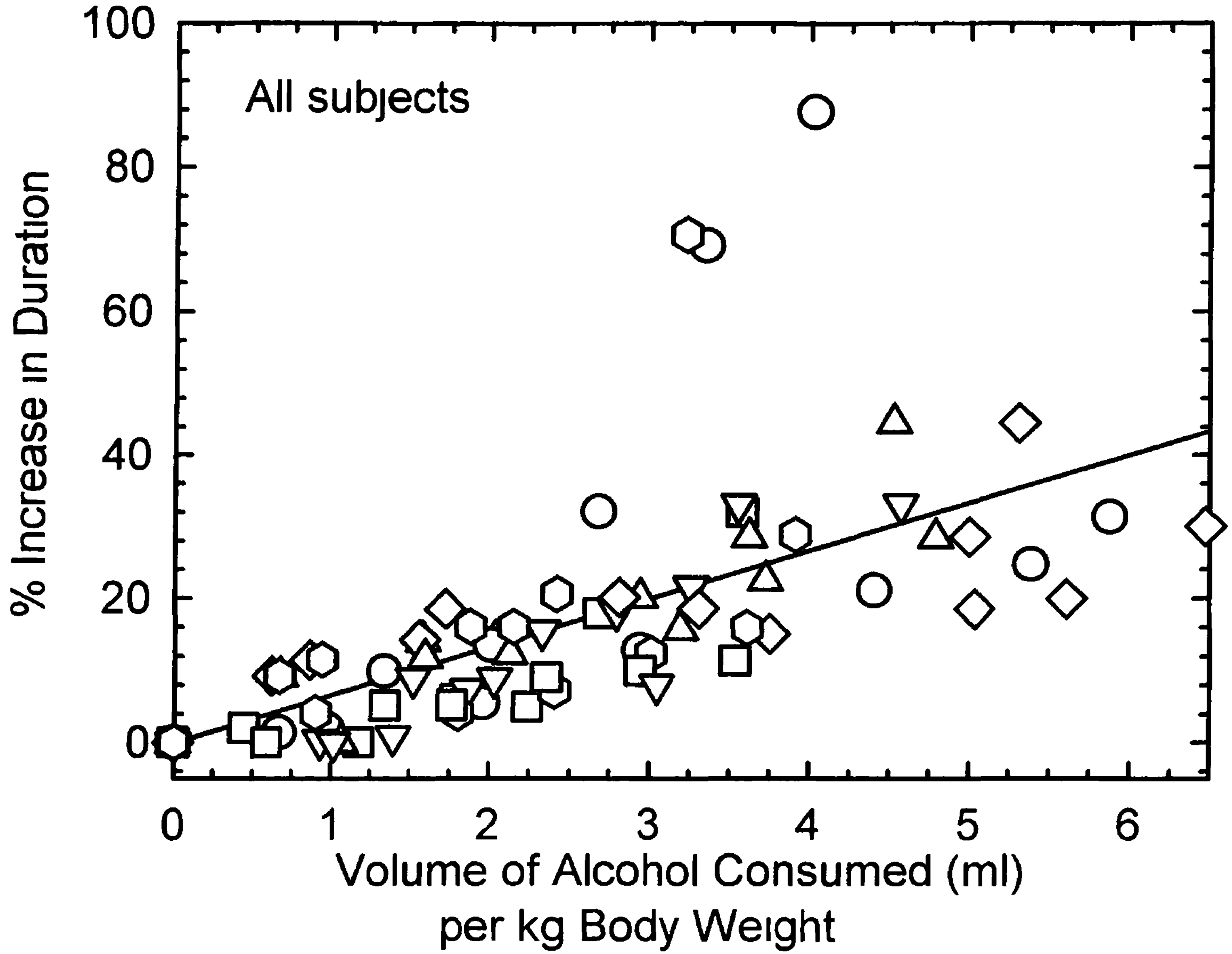


Figure 3.13. *% Increase in sentence duration vs Alcohol Consumed*

Sentence duration was examined for distribution and, on the basis of a concave probability plot, transformed by taking its logarithm This resulted in a close approximation to normal distribution (Pearson r with expected normal distribution = 0 993) Again the data were analysed using a random effects general linear model and in this case alcohol level (ml per kg bodyweight), sentence ID (to reflect the use of three sentences), speaker ID and sex were included as factors Analysis showed there was no evidence that sex was related to the effect of alcohol on sentence duration The p-value for the interaction term was 0 628 and ommitting this the main effect for sex was not significant (p = 0 523)

The remaining factors subject ID, sentence ID and alcohol level were all significantly associated with sentence duration The results for alcohol level showed that log duration increases by 0 027 (std err = 2 348e-3, t ratio = 11 51, p< 0 0001) for a 1-unit (ml/kg) increase in alcohol level This corresponds to a 6 4% ± 0 5% increase in sentence duration per ml of alcohol consumed per kg bodyweight Examination of residuals from the model showed that they were normally distributed (r = 0 997 with expected distribution) and examination of residuals showed no pattern consistent with a poorly fitted model

Chapter 4: Discussion and Conclusion

The results from the fundamental frequency analysis for the sustained vowel utterance /a / and the isolated word utterances 'me', 'you' and 'to', show that some variations do occur as the level of alcohol increases There is an increase in pitch variability for most subjects (see Table 3 2) However, there are no consistent trends observed There was no association found between f_1 and f_2 The difference between the first and second formant, f_2-f_1, was found to be unrelated to alcohol consumption as was the ratio of the first formant to the second formant, f_1/f_2

For the utterance of vowel sounds within spoken sentences it was found that the fundamental frequency rose on average by 1 7Hz (std err = 0 57) per ml alcohol per kilogram bodymass This is a statistically significant result Fig 4 1 shows the change in fundamental frequency versus alcohol consumed for all subjects The large spread of data about the mean indicates that the average increase of 1 7Hz per ml of alcohol per kg bodymass is not of any real practical significance However, one would speculate that if this experiment were to be carried out using non fixed text, then different results may be observed

The results that have been obtained from this analysis are in keeping with research that has previously been carried out Trojan and Kryspin-Exner [6] found that changes in pitch as a result of alcohol consumption

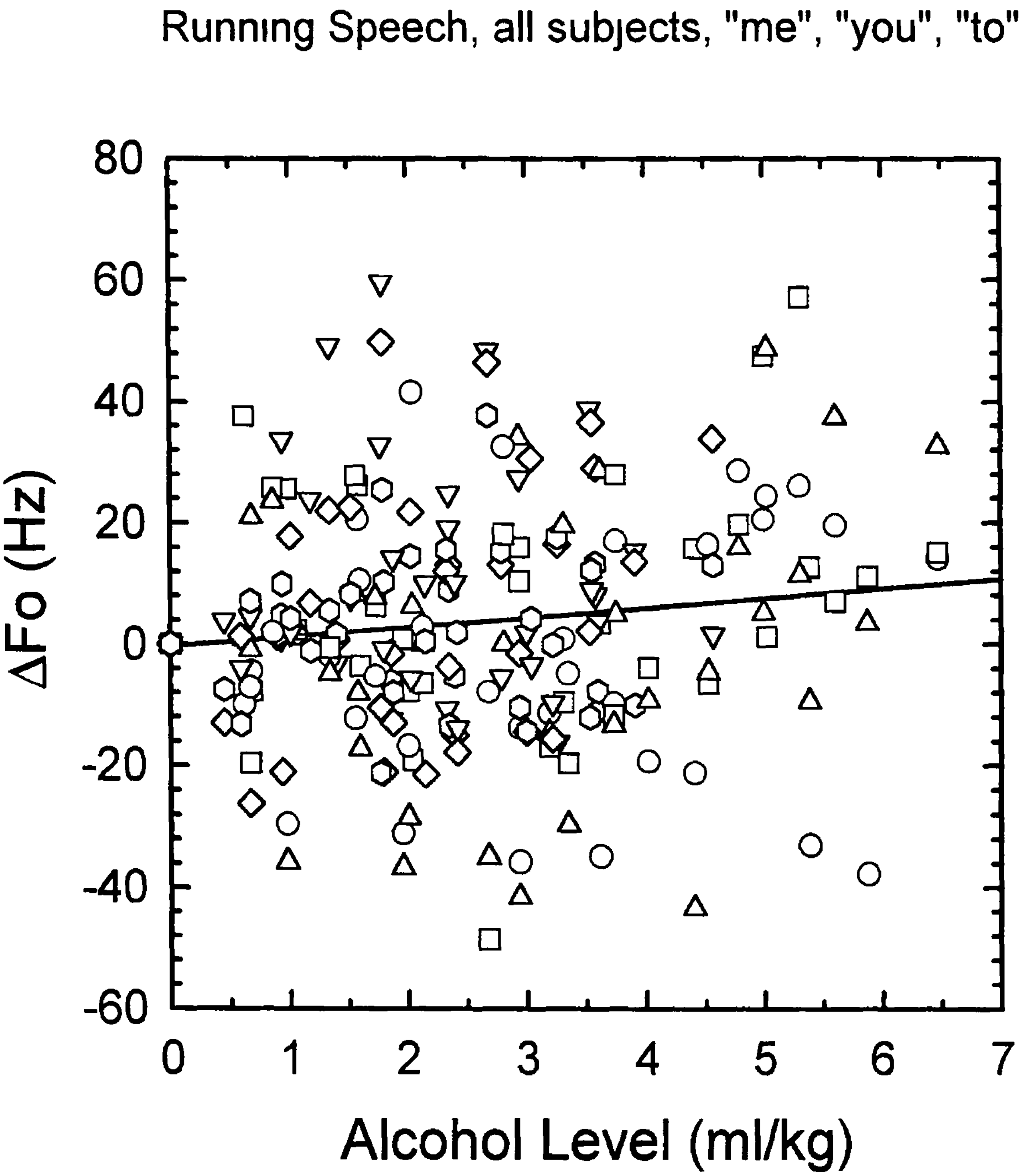


Figure 4.1. *Change in F_o versus alcohol consumed for all subjects*

vary from speaker to speaker but reported a general increase in fundamental frequency, whereas Pisoni and Martin [2] found pitch level to be much more variable in the intoxicated condition and that f_0 decreased for some but not all subjects Klingholz et al [3] from their research found a tendency for decreased fundamental frequency as the level of alcohol consumption increases They also found that the f_1/f_2 parameter, which describes the articulation varied from speaker to speaker but was not significantly related to alcohol level From all the research that has been carried out in this area, no one trend concerning the change in fundamental frequency as alcohol level increases has been established

Of all the acoustic analyses that were carried out in this project, the most consistent difference observed between speech samples in the sober and intoxicated states was an overall increase in duration In this analysis the change in duration varied significantly from person to person An average increase of 6 4% ± 0 5% in sentence duration per ml of alcohol consumed per kg bodymass was observed This is also in keeping with all previous studies that have been carried out However, in all the research that has been carried out in this area prior to this study, there has been no actual quantified increase reported Pisoni and Martin [2] found a change in the duration of spoken sentences to be the most consistent difference to be observed between sober and intoxicated speech with an average magnitude of sentence lengthening for individual talkers of from 75 to 158msec.

One of the major limitations of this experiment is the fact that the speakers had to read from a prescribed text. It might have been more useful to carry out the analysis on spontaneous voice recordings in which the individual would be much less conscious of the fact that their speech was being examined and hence more realistic resu lts could be obtained from the analysis. However, at the time that the experiment was conducted, a successful way of recording spontaneous speech samples for comparison had not been established.

On examination of the results it is found that the only result, in terms of change in f_0 as a function of alcohol consumption, that is of any significance is for the utterance of vowel sounds within running speech. A possible reason for this is that the words are being spoken more naturally when said in this manner rather than when they are being uttered in isolation. A way of overcoming this problem may be to have the subjects recite nursery rhymes rather than utter words in isolation and carry out the analysis. Alternatively, the analysis could be carried out on voice recordings of spontaneous speech where certain words and phrases that are generally repeated in speech could be picked out for analysis.

Another shortcoming of this research was in the measurement of blood alcohol level. Each subject was required to breathe into the Lion

Alcolmeter, an electronic breathalyser (Lion Laboratories Ltd , Barry, U K), before each speech recording was taken This instrument gives a reading of blood alcohol level Unfortunately, it was later found that the readings obtained from this device were abnormally high This was due to the fact that subjects would give a breath sample within thirty seconds of drinking alcohol and may still have a certain amount of the alcohol in their mouths For this reason the readings from the alcolmeter had to be disregarded Alcohol levels then had to be expressed in the form of volume of alcohol consumed (ml) per kilogram bodymass This does not give as accurate a reflection as to the level of intoxication of the individual as a properly functioning blood alcohol level meter does This is due to fact that there are other factors apart from bodymass which influence the level of alcohol intoxication of an individual For instance, the individuals metabolic rate or whether or not food had been consumed before the intake of alcohol would play a role in determining the level of intoxication of a particular person The most accurate way to determine the blood alcohol level of an individual would be to take a blood sample and get a direct reading of alcohol level from that

An additional improvement that could be made to this research would be to carry out the experiment on a better sampling of the population There were only four male subjects in this experiment A suggestion would be that the experiment be carried out on a larger number of subjects with equal numbers of male and female participants with a good spread of ages and bodymasses

A drawback in using this type of system as a means of discriminating between a sober and intoxicated voice is that for any particular individual there would have to be a sample of sober speech available in order that a comparison could be made between the sober speech sample and the suspected intoxicated one Although it has been found that for all subjects the duration of speech increases, this in itself would not be enough to definitely determine whether or not the individual is intoxicated

Chapter 5:Simulations of Noise induced Hearing Loss

5.1. Introduction

When individuals are continuously subjected to very loud noises over a period of time they often experience a deterioration in their hearing level This often happens to people who work in noisy environments such as construction sites where equipment such as kango hammers and high powered drills are in operation If sufficient hearing protection is not worn the individual is prone to some form of hearing loss Another example of where this kind of hearing loss may occur would be for people who spend a lot of time in, or work in, discos or night clubs where the volume of the music being played is often at a very high level

Recently, there have been a number of claims made by army personnel that hearing loss has been experienced by many individuals in the army as a result of wearing no hearing protection when being exposed to such loud noises as gunfire etc on a contmuous basis Many of these claims have been taken to court as the army personnel involved want compensation for the hearing damage they have incurred They claim that their hearing has been affected in several ways as a result of this exposure People have to speak at a higher volume than they usually would in order for the affected individual to hear at the same level they have heard at prior to being exposed to extremely loud environments Also in situations where there is background noise (pubs, restaurants etc) present a lot of the conversation sounds very muffled and it is very difficult to differentiate between different conversations High frequency sounds are virtually impossible for them to hear The problem with claims such as these is that it is very hard to determine the level of hearing loss that a particular individual has suffered and how much it actually affects the person Further to a request from the Department of Defence a system has been developed whereby samples of speech have been modified to simulate what a person suffering from noise induced hearing loss would hear In a court of law this will give a judge an indication as to the extent of the hearing loss that an individual has suffered

5.2 Choice of Filter

In order for speech samples to be made simulating noise induced hearing loss, a filter had to be designed which would block out certain frequencies hence simulating a deterioration in hearing level Speech samples were required that simulated a 55 dB drop off in hearing level for sounds made at frequencies above 2kHz and 3kHz in different environments These filters were designed using the Matlab programming package Several different types of filters were applied to the normal speech samples in order to determine which one gave the best response for the application in question From the different filters that were designed using Matlab, a Chebychev filter, order 9, sampling frequency of 44100Hz, and final attenuation to –55dB with 0 5 dB ripple in the passband was chosen Figure 5 1 shows the frequency response of this filter with a cutoff frequency of 2000Hz while figure 5 2 shows the frequency of the same filter with a cutoff frequency of 3000Hz Both these filters were applied separately to the different samples of speech chosen for the application

Fig.5.1. *Frequency response of Chebychev filter with 2kHz cutoff*

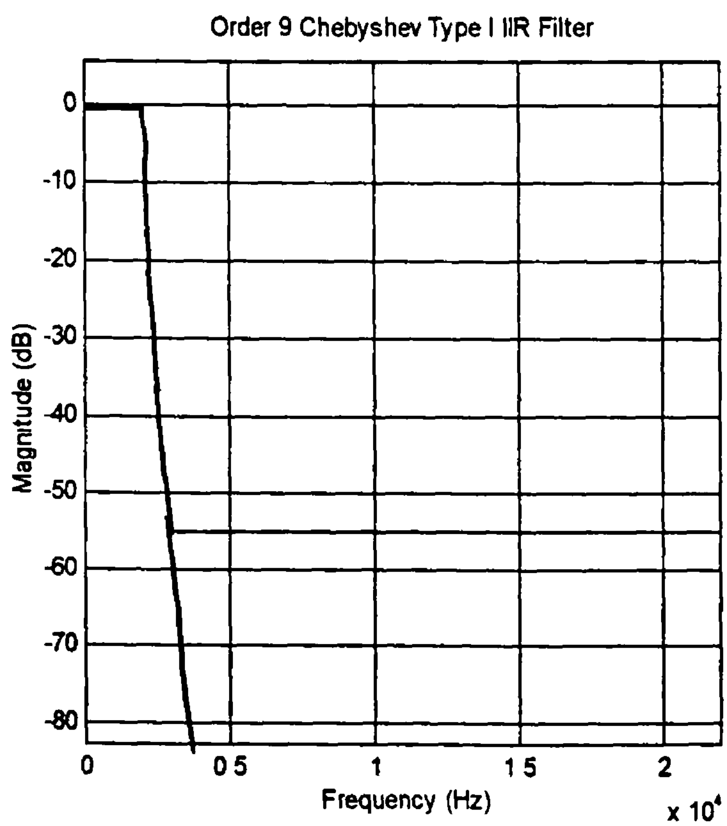
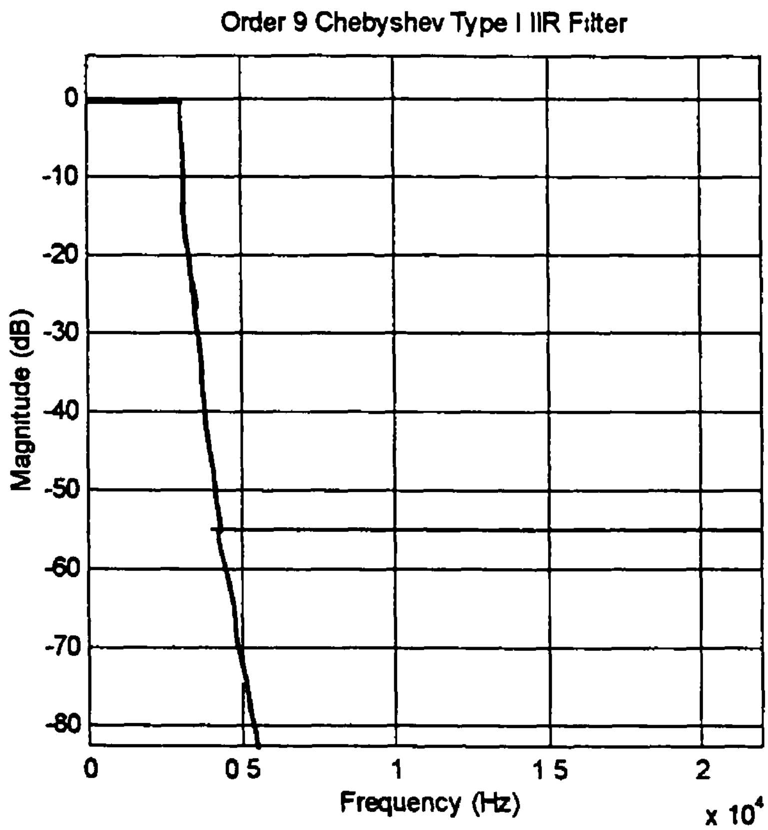


Fig 5.2. *Frequency response of Chebychev filter with 3kHz cutoff*



The sequence of commands used in Matlab in order to apply these filters to a particular segment of speech y are,

[b,a] = cheby1 (n, Rp, Rs, Wn, 'ftype');,

followed by,

x = filter (b, a, y);,

where

n = the order of the filter

Rp = the amount of ripple in the passband

Rs = the amount of attenuation

Wn = the cutoff frequency of the filter

Ftype = the filter type i e low pass

These filters were applied to different segments of speech spoken in different environments All the speech samples used in this application are stereo speech lifted from a compact disc Speech samples of a single individual have been added to restaurant and party sound effects by adding Matlab files Cool Edit '98 [Syntrillium Software Corporation, Phoenix, USA] has been used to replay the resulting speech samples and produce spectrograms for them

Figure 5 3 shows the spectrogram for a segment of speech in which a woman is reciting a piece of poetry with no background noise This spectrogram shows the frequencies present in this speech sample when no filter has been applied to it On listening to this sample of speech, it is what a person with normal hearing would hear The speech is clear and there is no loss of sound quality Figures 5 4 and 5 5 show the spectrograms for the same segment of speech with filters applied with cutoff frequencies of 2kHz and 3kHz respectively and a hearing attenuation of 55dB On listening to these filtered speech samples, the loss in quality is quite apparent The sample of speech is muffled and much more difficult to hear

Fig.5.3 *Spectrogram of speech sample with no background noise and no filter applied*

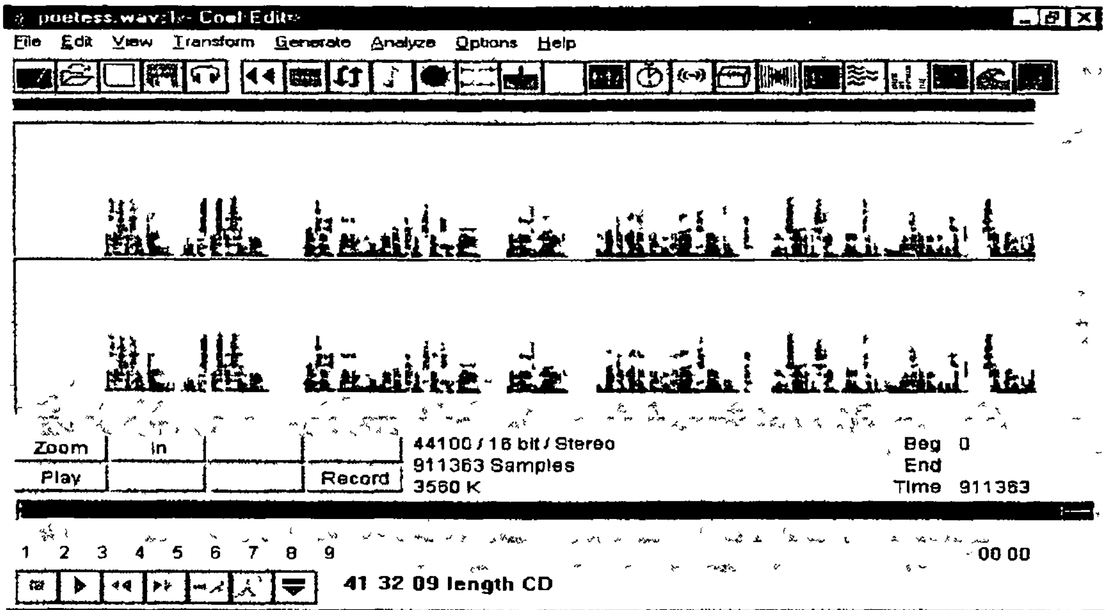
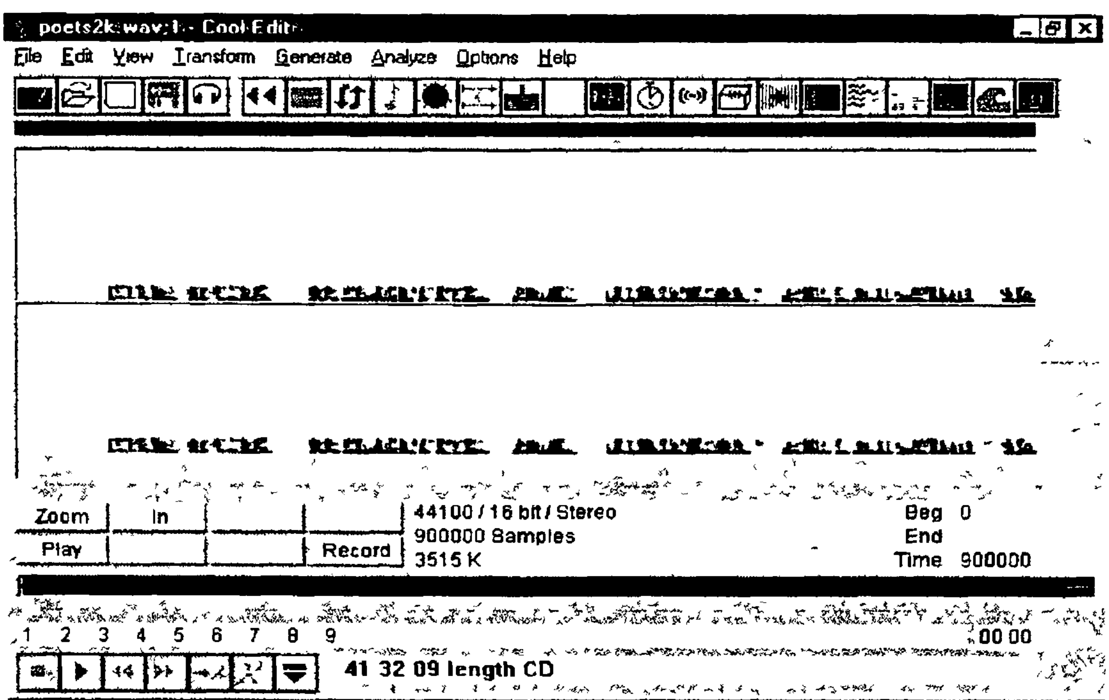


Fig. 5.4 *Spectrogram of speech sample with no background noise with 2kHz filter applied*



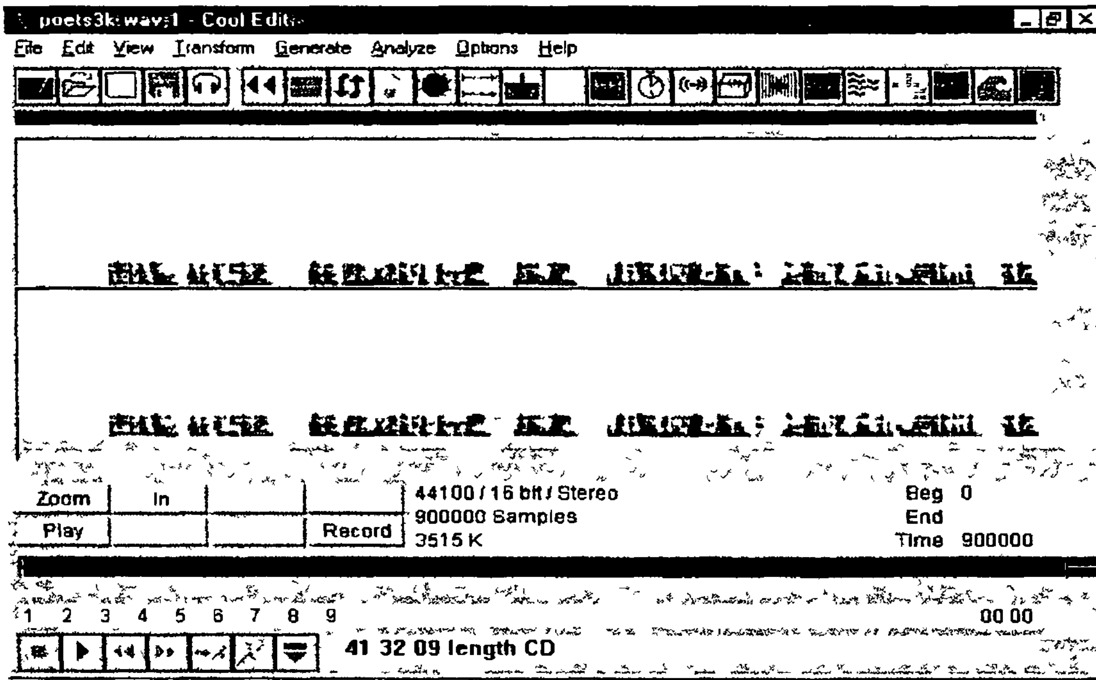


Fig. 5.5 *Spectrogram of speech sample with no background noise with 3kHz filter applied*

Figure 5 6 shows the spectrogram for the same sample of speech spoken in a pub background Here, there is no filter applied and hence no attenuation On listening to this sample of speech, a person with normal hearing can still quite clearly hear the woman reciting the poetry and is able to differentiate between her voice and the background voices from the pub Figures 5 7 and 5 8 show the spectrograms of what an individual suffering a hearing loss of attenuation 55dB and cutoff frequencies of 2 and 3kHz respectively, would hear On listening to these voice samples there is a substantial loss in quality of sound It is much harder to differentiate the background noise from the voice of the woman reciting the poetry All the high frequency sounds are omitted from these voice recordings and the poetry recital is very muffled All the sound files for these spectrograms are available in WAV file format on request

We can conclude that the design of this filter using Matlab and its application to speech samples in different environments has successfully simulated what a person suffering from noise induced hearing loss would experience

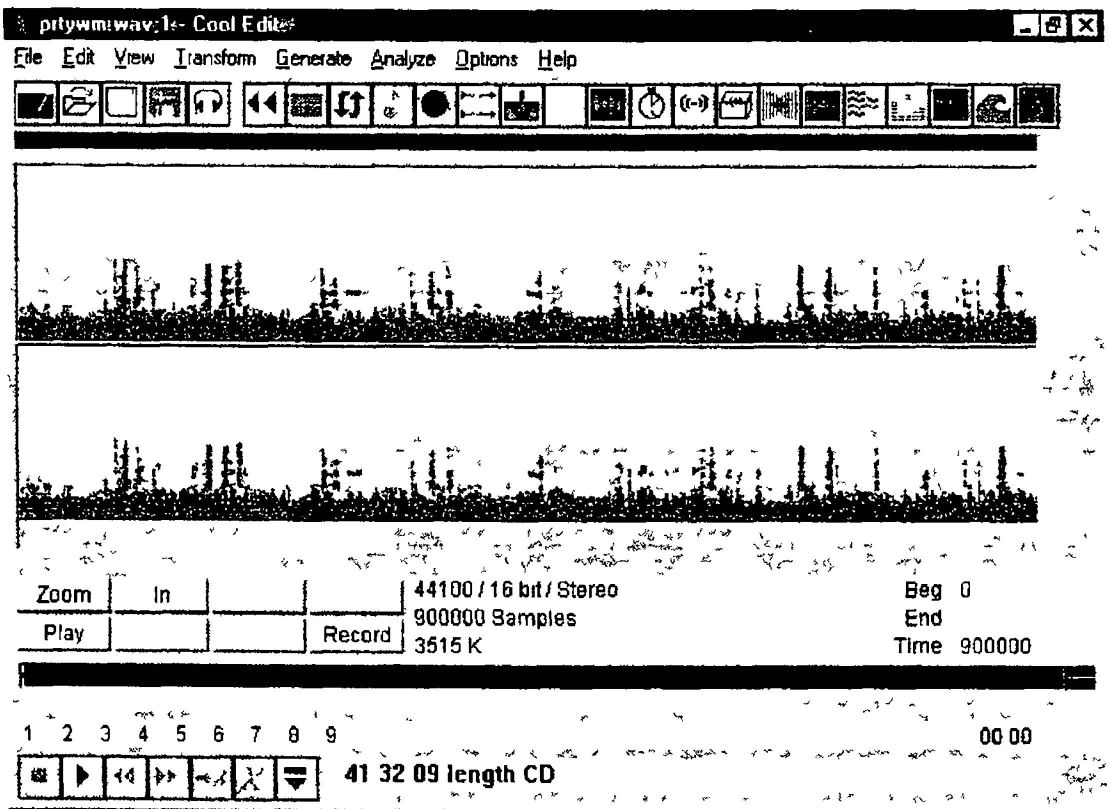


Fig. 5.6 *Spectrogram of speech sample with background noise with no filter applied*

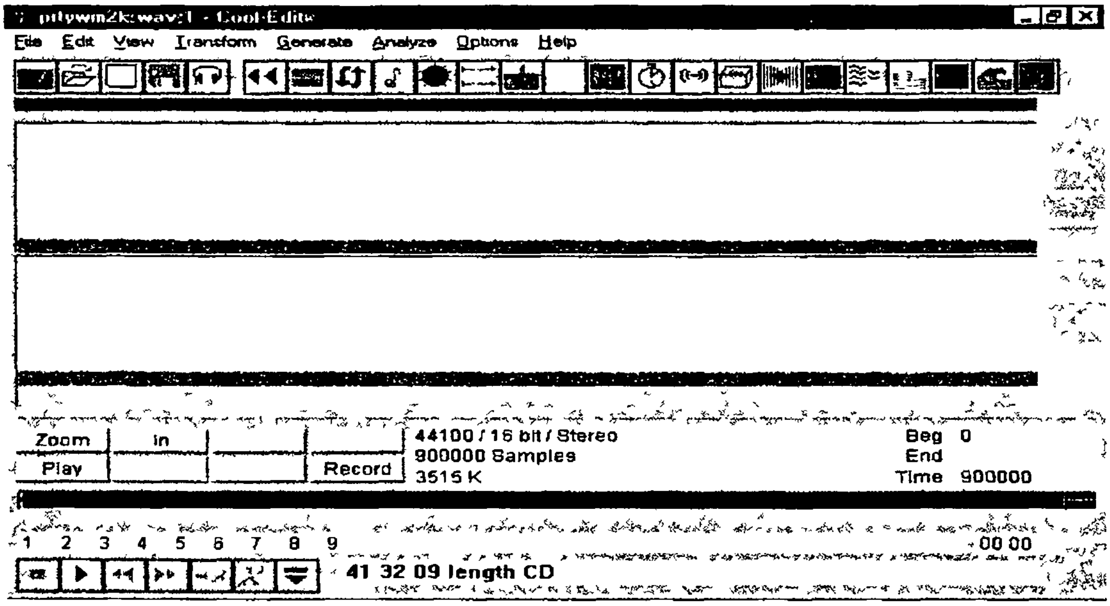


Fig. 5.7 *Spectrogram of speech sample with background noise with 2kHz filter applied*

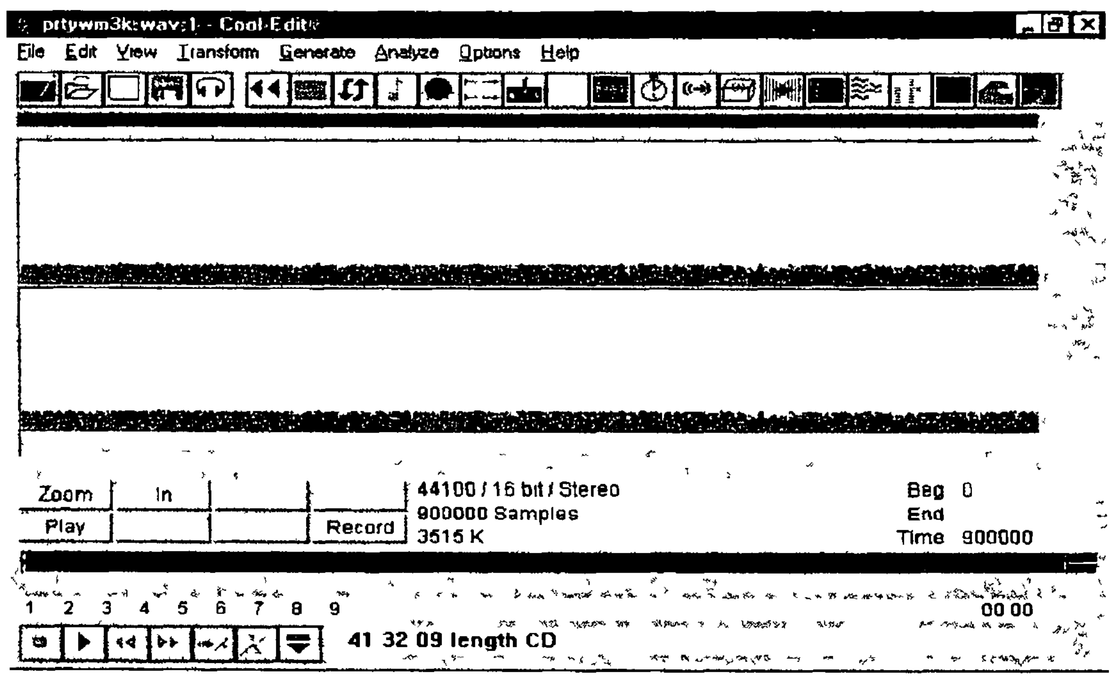


Fig. 5.8 *Spectrogram of speech sample with background noise with 3kHz filter applied*

References

1 Rossing, Thomas D The Science of Sound, 2nd edition Addison Wesley 15727, 1990

2 Pisoni, David B, and Martin, S effects of Alcohol on the Acoustic Phonetic properties of Speech Perceptual and Acoustic Analyses Alcoholism Clinical and Experimental Research Vol 13 No 4 July/August 1989

3 Klingholz, F et al Recognition of low - level alcohol intoxication from speech signal J Acoust Soc Am 84 (3), September 1988

4 Dunker, E and Schlosshauer, B Irregularities of the laryngeal vibratory pattern in healthy and hoarse persons Research potentials in Voice Physiology, edited by D W Brewer (International Conference at Syracuse 1961, Syracuse, NY), pp 151 - 184

5 Sobell, L C , and Sobell, M B Effects of Alcohol on the Speech of Alcoholics J Speech Hear Re 15, 861 - 868, 1972

6 Trojan, F , and Kryspin-Exner, K The decay of articulation under the influence of alcohol and paraldehyde Folia Phoniatr 20, 217-238, 1968

7 Lancaster, D "Low-pass and high- pass filter responses", in The Active Filter Cookbook Indianapolis Howard W Sams 7 Co , 1987

8 Horowitz, P and Hill,W The Art of Electronics Cambridge Mass Cambridge University Press 1986

9 The Matlab Reference Guide The Math Works Inc 1992

10 Rabiner, L and Schafer, R Digital Processing of Speech Signals, Englewood Cliffs, N J Prentice Hall , 1978

11 Oppenheim, A V and Schafer, R W Discrete- time signal processing Prentice Hall, 1989

12 Harris, F J On the use of windows for harmonic analysis with the discrete Fourier transform Proc of the IEEE, 1978, 66 51 142

13 Weimer, Richard C Statistics 2nd edition, Wm C Brown Publishers, 1989

Appendix A

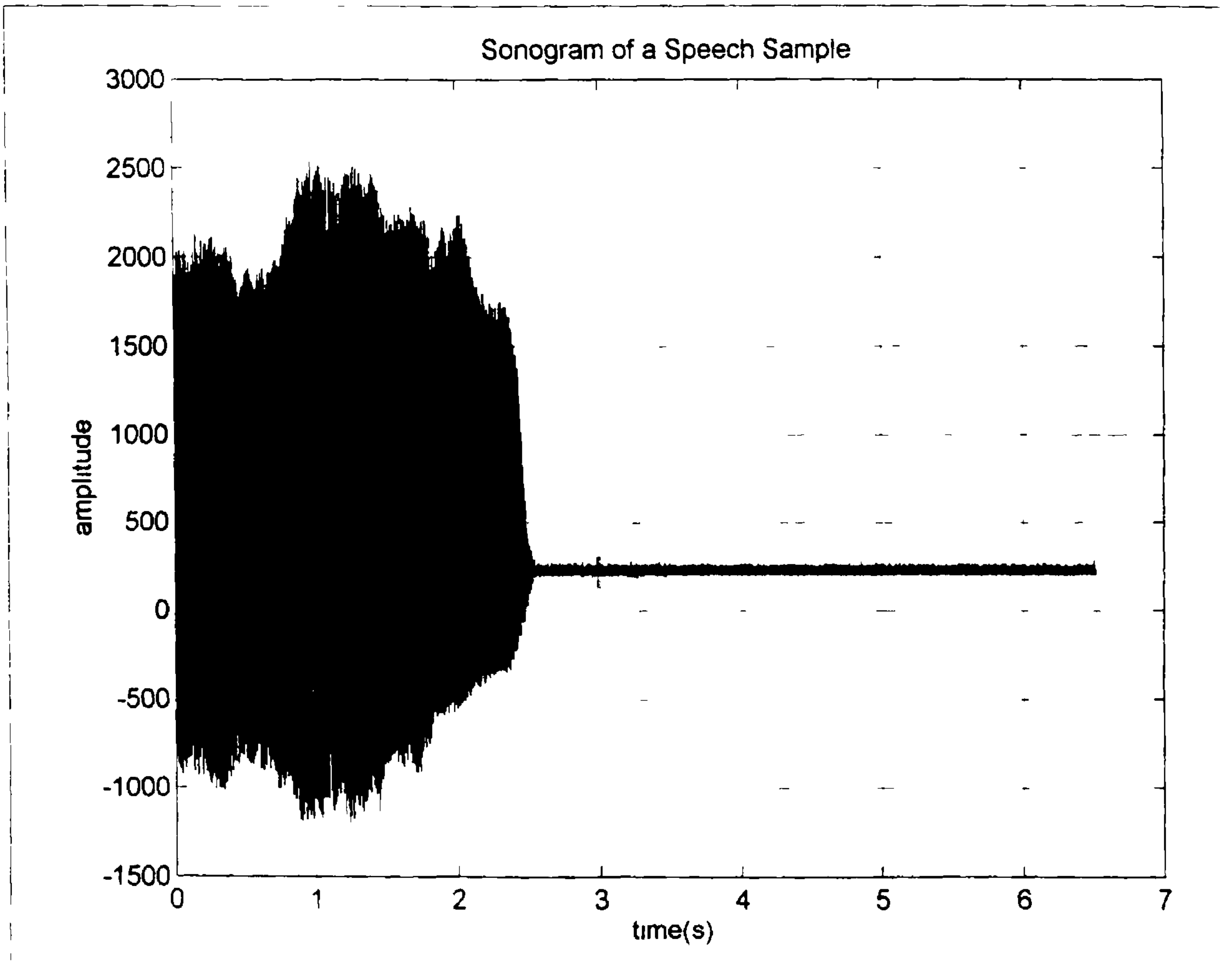
Lıstıng and graphıcal outputs for the programs

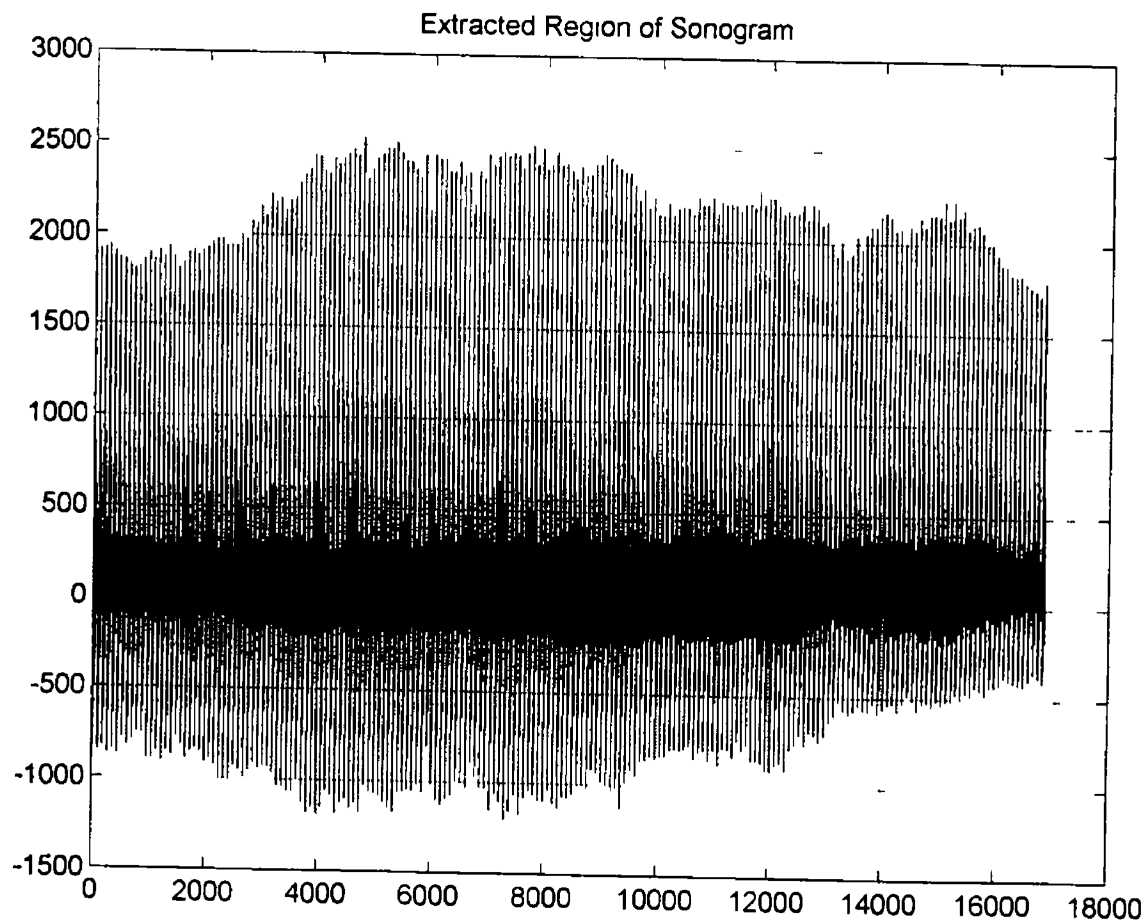
fundfr m, durat m, formant m

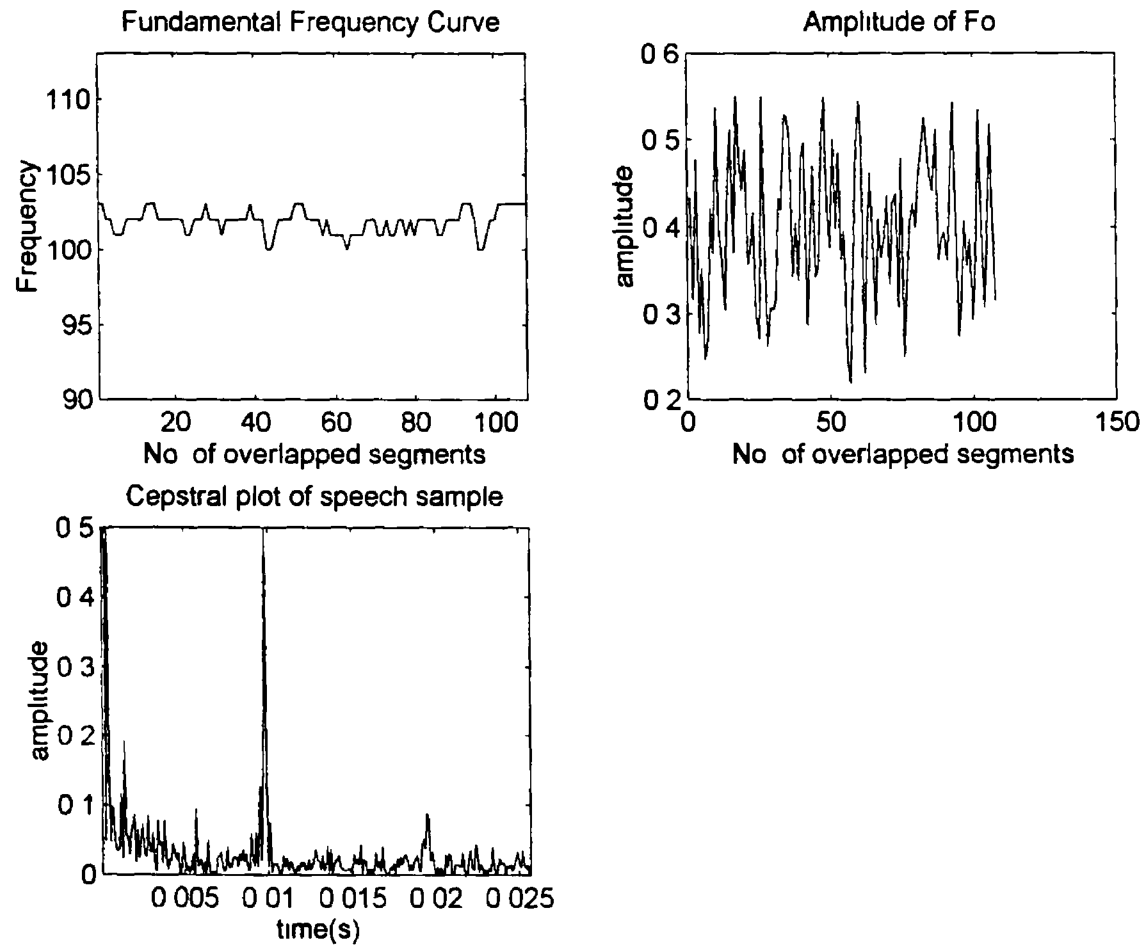
```
% A Matlab program to determine the fundamental frequency of a
particular
% speech sample
%
% Name  Fundfr m
%
%
%
% By  Orla Cooney
%
% Function  fundfr(w,n,pad)
% Variables  w= speech sample obtained from data file
%           n= length of data to be sampled
%           pad=pad out with zeros-chosen by user
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%

function fundfr(w,n,pad)
fsam=10000,

olap=100,
clf,
subplot(2,1,1),
plot(w),
title('Sonogram of a Speech Sample'),
xlabel('time(secs)')
ylabel('Amplitude')

disp('Pick points of interest with mouse')
grid
[x,y]=ginput(2),
sp1=floor(x(1)),
sp2=floor(x(2)),
sp3=sp2-sp1,  %Length of segment to be extracted

subplot(2,1,2),
S=(w(sp1 sp2)),
plot(S)
title('Extracted Region of Sonogram')
grid
```

```
%Fundamental Frequency calculation using Cepstrum Property

for i=1 ((sp3-n)/olap)
div=S((1+(i-1)*olap) (n+(i-1)*olap)),
lafdiv=log(abs(fft(div *hamming(n),pad))),

cepstrum=abs(real(ifft(lafdiv,pad))),
amp(i)=max(cepstrum(30 250)),
for k=30 250
if cepstrum(k)==amp(i)


fund(i)=1/((1/fsam)*k),
end
end

end

disp('Fundamental Frequency(Hz)'),
a=fund'
disp('Mean value of fundamental frequency'),
mean(fund)

disp('Press Spacebar to Continue'),
pause,
subplot(2,2,1),
plot(fund, 'm-'),
axis([1 length(fund) min(fund)-5 max(fund)+5]),
title('Fundamental Frequency Curve')
xlabel('No  of overlapped segments')
ylabel('Frequency(Hz)')

subplot(2,2,2),
plot(amp,'r-'),
title('Amplitude of Fo')
xlabel('No  of overlapped divisions')
ylabel('Amplitude')


for i=1 5 ((sp3-n)/olap)
disp('Press Return to Obtain Cepstrum')
pause,
```

```
subplot(2,1,2),
plot((1 (pad/2+1))*1/fsam,cepstrum(1 (pad/2+1)),'w-')
title('Cepstral plot of Speech Sample')
xlabel('time(secs)/Quefrency')
ylabel('Amplitude')
end
```

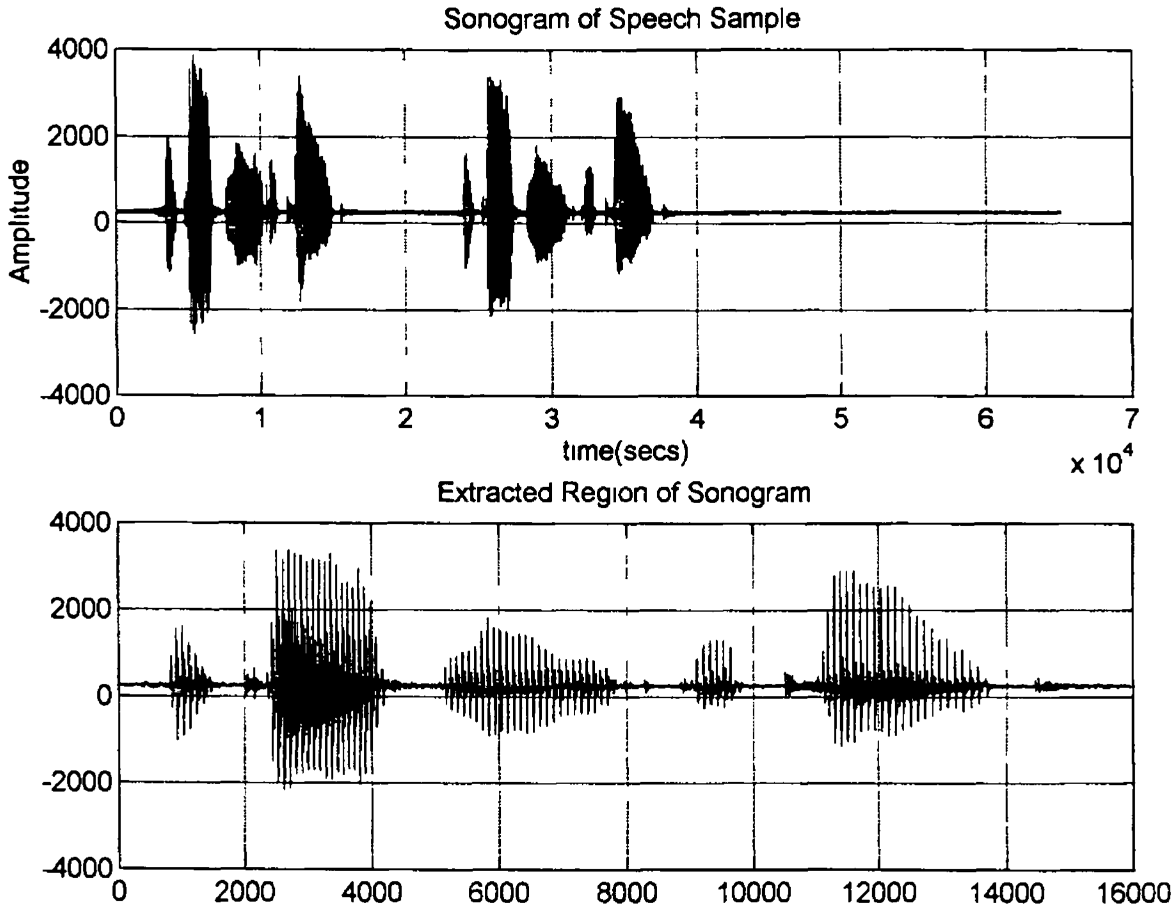

Sonogram of a Speech Sample
amplitude
time(s)
3000
2500
2000
1500
1000
500
0
-500
-1000
-1500
0
1
2
3
4
5
6
7


Extracted Region of Sonogram
3000
2500
2000
1500
1000
500
0
-500
-1000
-1500
0
2000
4000
6000
8000
10000
12000
14000
16000
18000


Fundamental Frequency Curve
Frequency
110
105
100
95
90
20 40 60 80 100
No of overlapped segments
Amplitude of Fo
amplitude
0 6
0 5
0 4
0 3
0 2
0 50 100 150
No of overlapped segments
Cepstral plot of speech sample
amplitude
0 5
0 4
0 3
0 2
0 1
0
0 005 0 01 0 015 0 02 0 025
time(s)

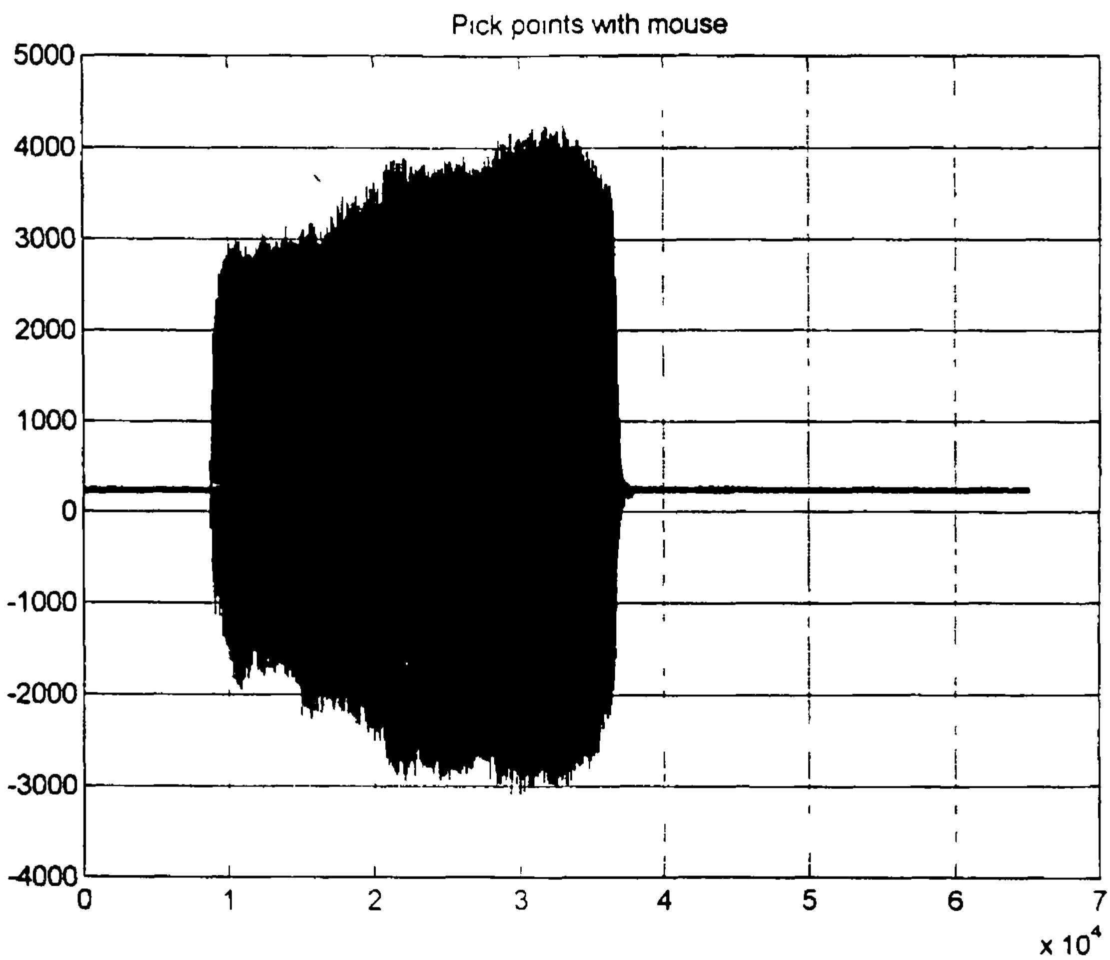
```
% A Matlab program to calculate the duratıon of a partıcular segment of
speech
%
% Name  Durat m
%
%
%
% By  Orla Cooney
%
%
% Functıon  Durat()
%
% Varıables
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

functıon durat(w,fsam)

fsam=10000,
clf,
subplot(2,1,1),
plot(w),
tıtle('Sonogram of Speech Sample'),
xlabel('tıme(secs)')
ylabel('Amplıtude')

dısp('Choose sectıon to be extracted')
grıd
[x,y]=gınput(2),
sp1=floor(x(1)),
sp2=floor(x(2)),
sp3=sp2-sp1,  %Length of segment to be extracted

subplot(2,1,2),
S=w((sp1 sp2)),
plot(S)
tıtle('Extracted Regıon of Sonogram')
grıd

G=length(sp1 sp2)
```

```
dısp('Duratıon of Speech Sample')
Z=G*(9 83e-5)
```

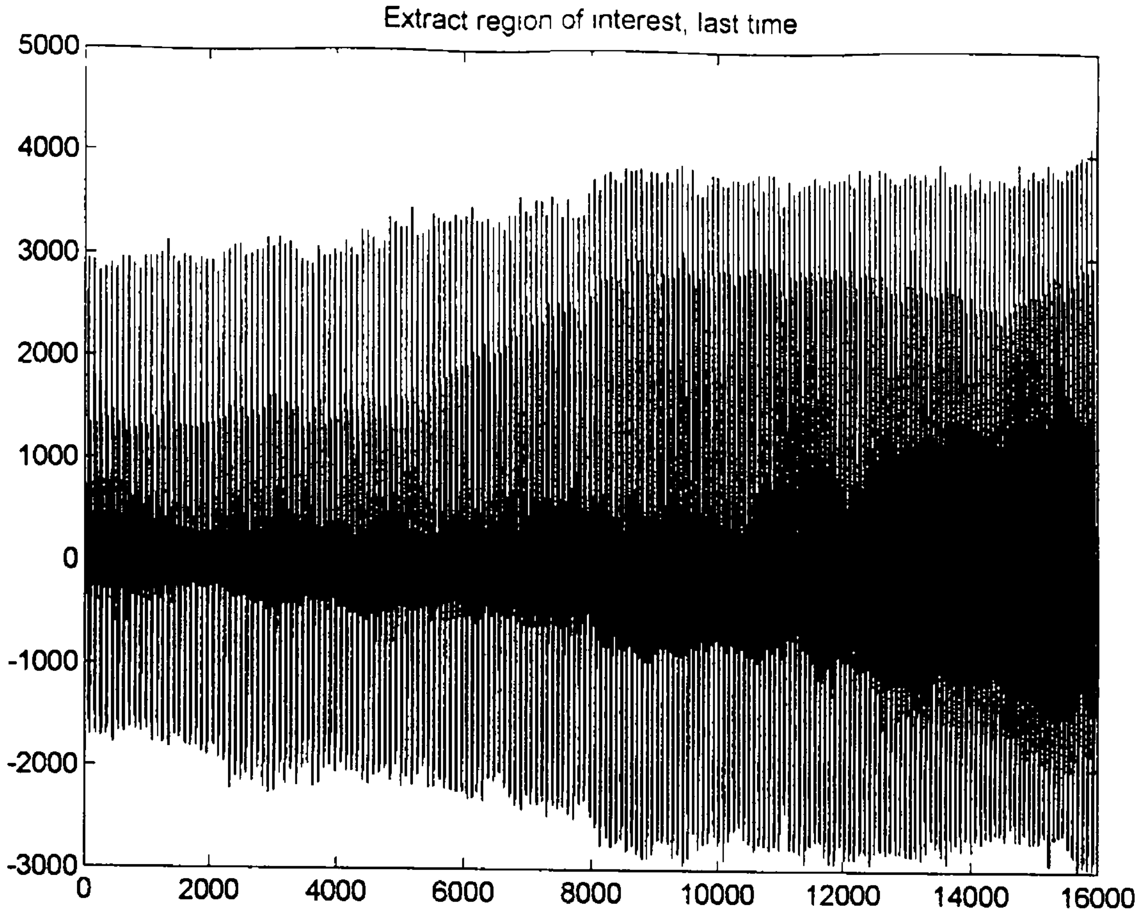

Sonogram of Speech Sample
4000
2000
0
-2000
-4000
Amplitude
0
1
2
3
4
5
6
7
time(secs)
x 10^4
Extracted Region of Sonogram
4000
2000
0
-2000
-4000
0
2000
4000
6000
8000
10000
12000
14000
16000

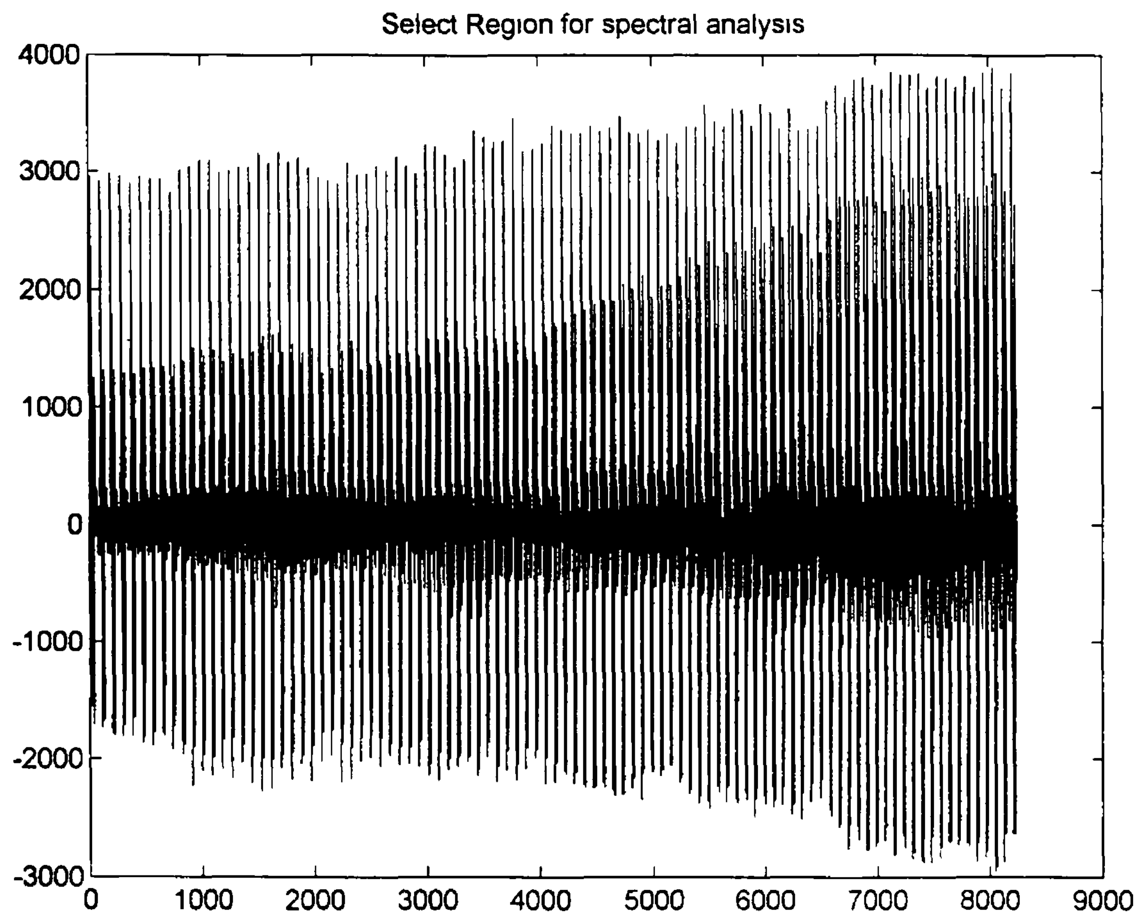
```
% A Matlab program to detect a word, extract a segment from that
% word, carry out spectral calculations and plot a spectrogram of
% the formant tracks versus time
%
% By  Orla Cooney
%
%
% Variables  seg = segment size
% pad = pad out with zeros
% olap = segment overlap
% fsam = sampling frequency
%
%
% function formant(w)
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function formant (w)

w1 = extract(w),
plot(w1),

title ('Select region for spectral analysis'),

[x,y] = ginput(2),

sp1 = floor(x(1)),
sp2 = floor(x(2)),
sp3 = sp2 - sp1,
w = w1(sp1 sp1 + sp3),
close,

% fft calculation

segno = floor ((length(w)/olap) - (seg/olap)),

f1 = fsam/pad*(0  (pad/2) - 1),

hs = hamwind (seg),

for k= 0 segno - 1,
w2 = w (1 + k*olap  seg + k*olap),
```

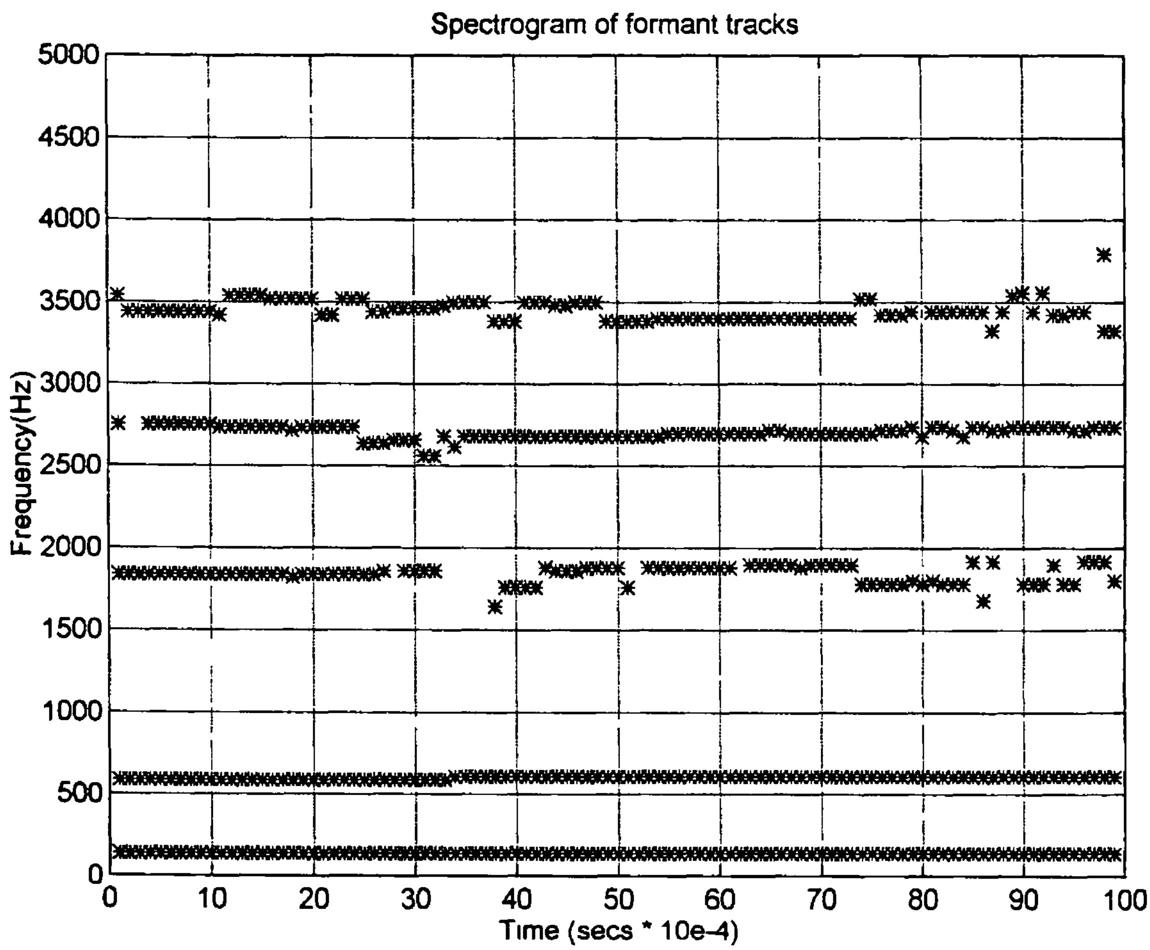
```
ffdat = (abs(fft((w2*hs),pad))),

dbcal = 10 * log 10 (ffdat),

end

w3 = specenvel (fsam, dbcal (1   (pad/2))),

pk_points = max (w3),

freqs = (fsam/512 *pk_points),

specform(w3,freqs),

[trackdata] = trackvec,

% close

disp ('Program Finished'),
```



Pick points with mouse
5000
4000
3000
2000
1000
0
-1000
-2000
-3000
-4000
0
1
2
3
4
5
6
7
x 10^4


Extract region of interest, last time
5000
4000
3000
2000
1000
0
-1000
-2000
-3000
0
2000
4000
6000
8000
10000
12000
14000
16000

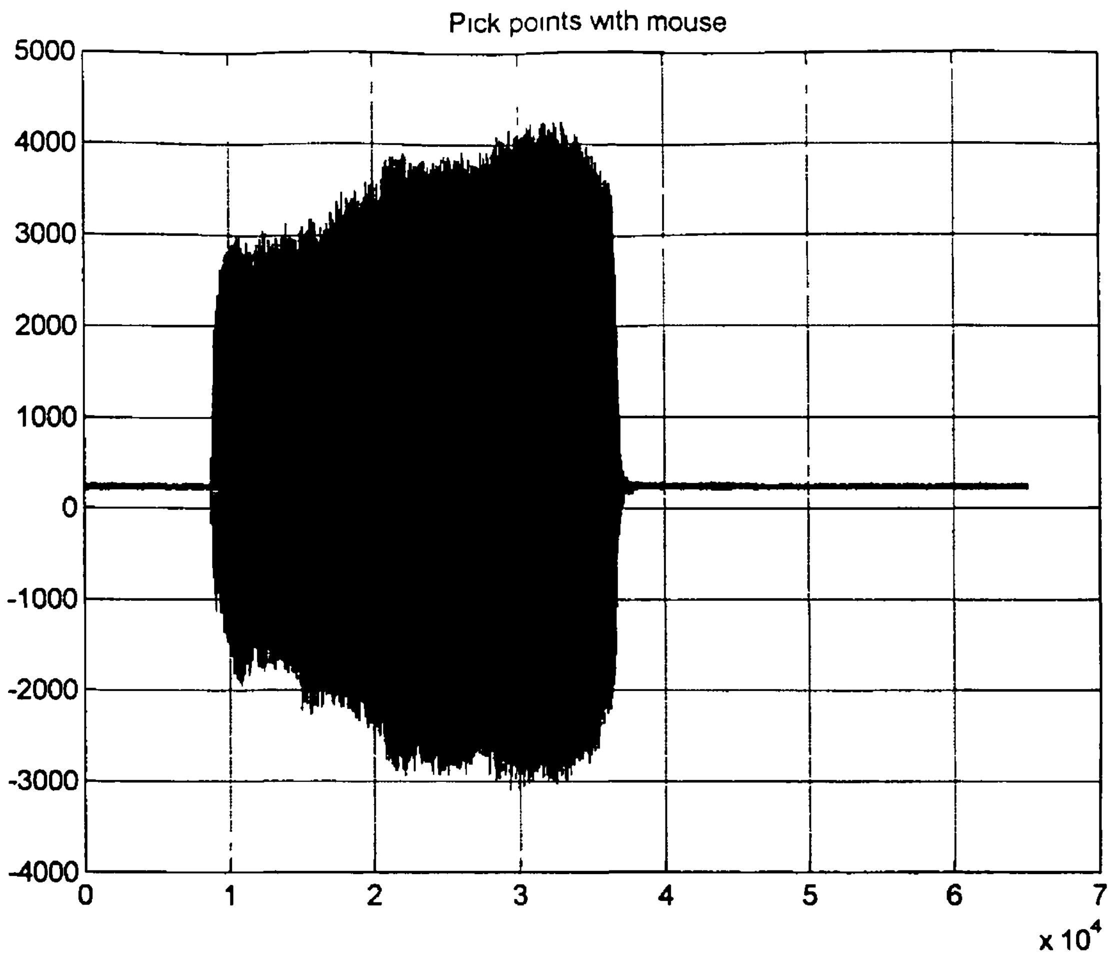

Select Region for spectral analysis
4000
3000
2000
1000
0
-1000
-2000
-3000
0
1000
2000
3000
4000
5000
6000
7000
8000
9000


Spectrogram of formant tracks
Frequency(Hz)
Time (secs * 10e-4)
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
10
20
30
40
50
60
70
80
90
100

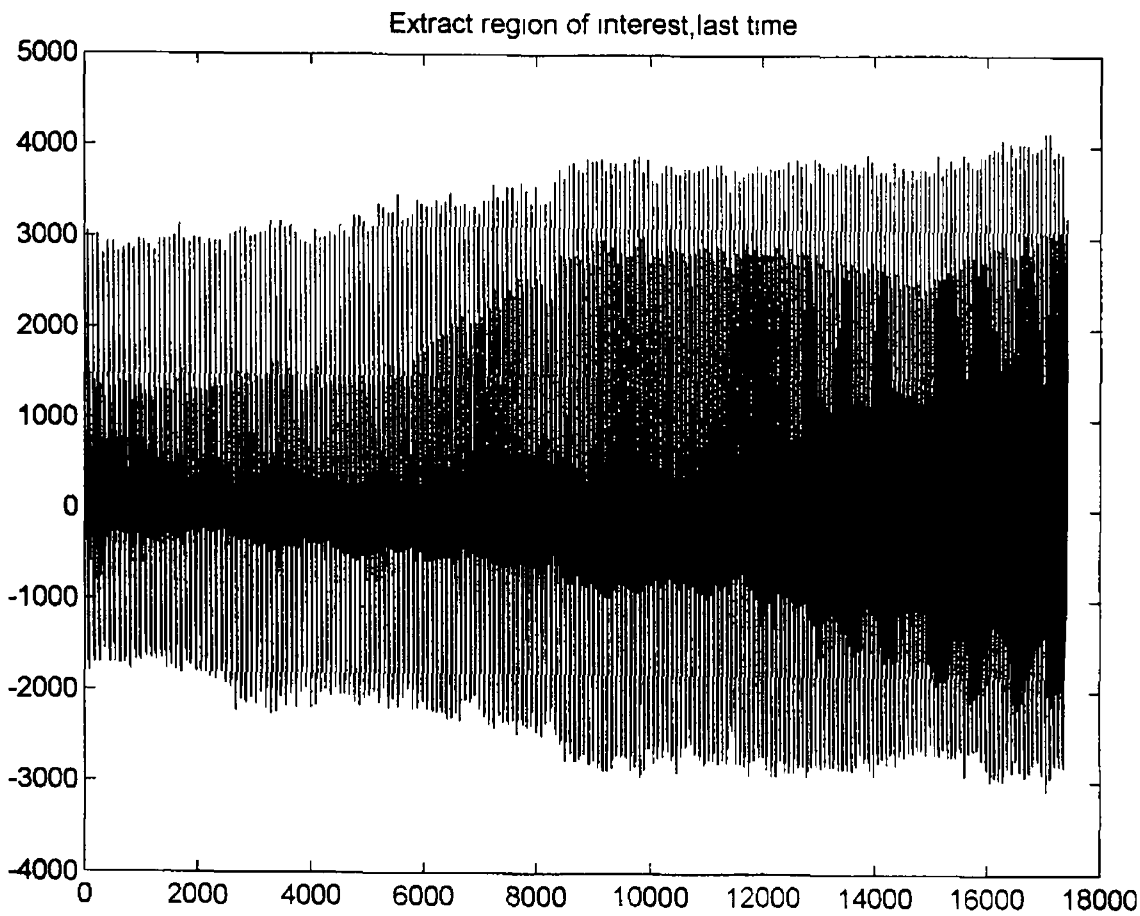
Appendix B

Listing and graphical outputs for the programs called as functions for the programs fundfr m, durat m, and formant m

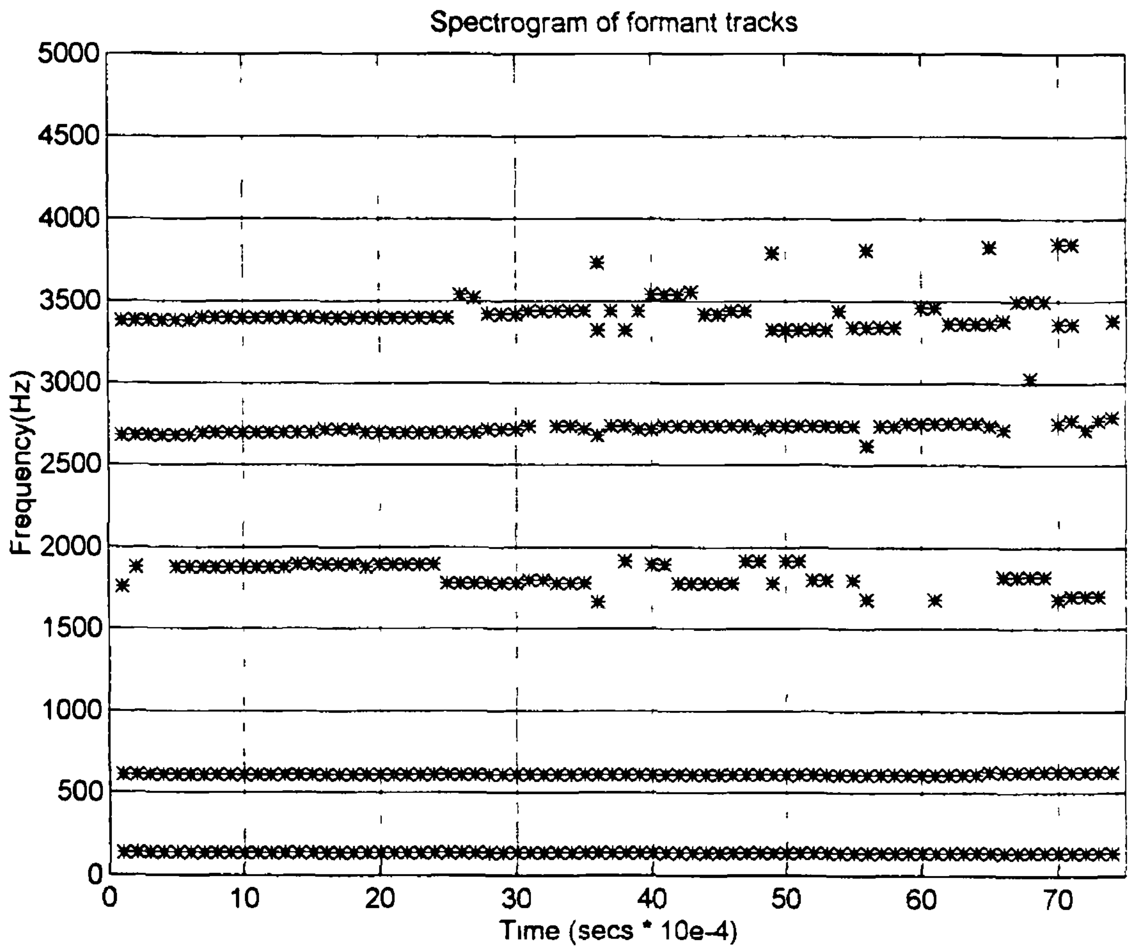
```
% Matlab program to detect a word, plot a sonogram of that word, and
% also a sonogram of a specified region of interest
%
%
% Filename  extract m
% By  Orla Cooney
% Date  08/11/1995
%
% Function  extract(w)
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function extract(w)

clf,
plot(w),
title('Pick points with mouse'),
grid
[x,y] = ginput(2),
sp1 = floor(x(1)),
sp2 = floor(x(2)),
sp3 = sp2-sp1,
clf,
plot(w(sp1 sp1+sp3)),
title('Extract region of interest,last time'),
[x,y]=ginput(2),
sp4=floor(x(1)),
sp5=floor(x(2)),
sp6=sp5-sp4,
regint=w(sp1+sp4 sp1+sp5),
clf,
grid
```



Pick points with mouse
5000
4000
3000
2000
1000
0
-1000
-2000
-3000
-4000
0
1
2
3
4
5
6
7
x 10^4


Extract region of interest,last time
5000
4000
3000
2000
1000
0
-1000
-2000
-3000
-4000
0
2000
4000
6000
8000
10000
12000
14000
16000
18000

```
% A Matlab program to plot the spectrogram of formant track data
%
% There are two vector arrays of data required to run this program,
% segment (a) and frequency (b)
%
%
% Filename  specform m
%
% By  Orla Cooney
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function specform (a,b)

lena = length (a),

p1 = [mm (a),max(a)],

p2 = [min (b), max (b)],

plot (p1,p2,'k '),
axis ([min (a)-1, max(a) + 1, 0, 5000])
hold,
for k = 1 lena,
plot (a(k), b(k), 'k '),
end
hold,
```



Spectrogram of formant tracks
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Frequency(Hz)
0
10
20
30
40
50
60
70
Time (secs * 10e-4)

```
% A Matlab program to create and return a Hammıng wındow of
% length m
%
% Fılename  Hamwınd (m)
%
% By  Orla Cooney
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

functıon [x] = hamwınd (m)

for k = 1 m,

x (k,1) = (0 54 – 0 46 * (cos (2 * pı * k/m))),

end

% The above equatıon ıs the standard formula for a Hammıng wındow
```

```
% A Matlab program which extracts the spectral envelope from a set
% of data
% This program operates on the principle that a peak point is detected
% when there is a slope change about a point going from positive to
% negative These peaks when detected form the spectral envelope and
% the frequency values at these peaks are returned
%
% Filename specenvel m
%
% By Orla Cooney
%
% fsam = sampling frequency
% datarr = array of data used
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function specenvel (fsam, datarr)

lendat = length (datarr),

r = 1,

slopea = datarr ( (1 lendat-2) + 1) – datarr (1 lendat – 2),

slopeb = datarr ( (1 lendat – 2) +2) – datarr ( (1  lendat – 2) + 1),

% The above two commands check for a change in slope from positive
% to negative between two data points

If (slopea > 0 and slopeb < 0 ),

Specen (r) = datarr ( (1 lendat – 2) +1),

End

Cal = 2 * lendat,

Call = (0   ( ( lendat) – 1) ),

Cal3 = cal * call,

Freqs = fsam/cal3,
```

For k = 1   (r – 1),

Freqvals (k) = freqs (n),

End

```
% A Matlab program which enables the user to input formant track data
% via the mouse and graph  and return the formant frequency
% information as a vector
%
% Filename  Trackvec m
%
% By  Orla Cooney
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


function [trackdata] = trackvec ()
p = 16,

[x,y] = ginput (p),

trackdata = y,

% The first four points in the vector that is returned by this program
%
% represent F1, the second four points are F2, the third four are F3
%
% and the last four are F4
```