

Real-time visual displays for voice tuition

David Paul Rossiter

A thesis conducted at the
Department of Electronics of the University of York
submitted for the degree of Doctor of Philosophy

January 1995

Abstract

The human voice is used regularly in day-to-day activity, particularly by teachers, politicians and singers. This can place a level of strain on their vocal mechanism. To counter this, professional voice users may undergo some form of vocal tuition. However, there is a prevalence of subjective, sometimes ambiguous tuition techniques. Vocal damage can therefore occur in spite of, or partially because of, the tuition techniques that are supposed to help the subject.

It is hypothesised that a computer system which can a) objectively measure a large set of parameters relating to vocal performance, b) appropriately translate these parameters to a smaller set of easily interpreted measures, and c) dynamically display this information in an appropriate graphical form may be able to provide a measure of progress to complement instructive techniques during vocal tuition.

A highly flexible, easy-to-use tool for the analysis and visualization of voice source and acoustic signals has been developed called ALBERT (Acoustic and Laryngeal Biofeedback Enhancement in Real-Time). The system can be used to provide real-time feedback on the state of a number of voice parameters. The visualization of these parameters is highly configurable, and includes 1D, 2D and 3D arrangements with optional colour mapping. ALBERT is novel in its ability to display many popular laryngeal and acoustic parameters in many informative graphical configurations. The system can be used in real-time to function as a tool for use by a voice tutor.

To test the hypothesis, three studies are presented of vocally developing male and female subjects as follows: i) a six week course of vocal expression, ii) a group of actors over an eight month period, and iii) a group of singers over a time span of 2.7 years. Two voice parameters in particular have been identified as indicative of vocal performance: the closed quotient of the vocal folds (CQ) and the spectral amplitude in the frequency band within which the phenomenon of the 'singer's formant' is known to occur relative to the spectral amplitude of the complete acoustic output (Ratio). ALBERT has been used as a tool to provide real-time visual feedback of these parameters in the context of singing tuition. Results suggest that ALBERT is a tool with considerable potential for use in a voice tuition process.

Contents

Abstract	i
Acknowledgements	xx
Declarations	xxii
1 Research aims and report structure	1
1.1 Introduction	1
1.2 Vocal use and abuse	1
1.3 Deficient voice tuition	2
1.4 Ambiguous tuition techniques	4
1.5 The reduction of ambiguity	5
1.6 Hypothesis	5
1.7 Thesis content and structure	6
2 Voice anatomy, physiology & acoustics	8
2.1 Introduction	8
2.2 The phonation circle	8
2.3 An overview of the vocal process	10
2.4 The subglottal system	12
2.4.1 An overview	12
2.4.2 Subglottal muscles	13
2.4.3 Respiration	14
2.5 The larynx	16
2.5.1 Introduction	16
2.5.2 The laryngeal framework	19

2.5.3	Laryngeal cartilages	19
2.5.4	An overview of the larynx muscles	20
2.5.5	The larynx muscles	21
2.5.6	Layers of the vocal folds	22
2.5.7	Laryngeal membranes and ligaments	22
2.5.8	Supraglottal cavity	23
2.6	The supraglottal system	24
2.6.1	The passive articulators	24
2.6.2	The active articulators	24
2.7	Acoustics of the voice	26
2.7.1	Glottal action	26
2.7.2	The spectrum of the glottal sound source	28
2.7.3	Supralaryngeal acoustics	32
2.8	Conclusions	35
3	Parameters for real-time assessment	36
3.1	Introduction	36
3.2	Voice parameters	36
3.3	Loudness	37
3.3.1	Introduction	37
3.3.2	The use of loudness in a real-time display	39
3.4	Acoustic signal changes	39
3.4.1	The 'singer's formant'	39
3.4.2	Physiological correlates	41
3.4.3	Assessment methods	42
3.4.4	A real-time measure	45
3.5	Fundamental frequency	45
3.5.1	Introduction	45
3.5.2	Fundamental frequency range	47
3.5.3	Fundamental frequency and intensity range	48
3.5.4	Range analysis in real-time displays	48
3.5.5	The use of fundamental frequency in a real-time display	50

3.6	Laryngeal closed quotient	52
3.6.1	Introduction	52
3.6.2	The applicability of closed quotient	52
3.6.3	Changes in closed quotient across different levels of intensity	52
3.6.4	Changes in closed quotient for different modes of singing . . .	53
3.6.5	The relationship between closed quotient and vocal training .	54
3.6.6	The suitability of closed quotient as a real-time measure . . .	56
3.6.7	Previous real-time closed quotient systems	57
3.7	Rates of larynx opening and closing	59
3.8	Conclusions	59
3.8.1	Real-time system design	59
3.8.2	Parameters indicative of vocal useage	59
3.8.3	Longitudinal assessment	60
4	Longitudinal voice studies	61
4.1	Introduction	61
4.2	Study 1 <i>A 6 week course of vocal expression</i>	63
4.2.1	Introduction	63
4.2.2	A course for vocal expression	63
4.2.3	Recordings	64
4.2.4	Data analysis	65
4.2.5	Results	65
4.2.5.1	Ratio	65
4.2.5.2	Fundamental frequency	69
4.2.5.3	Laryngeal closed quotient	70
4.2.5.4	The rate of closing of the vocal folds	71
4.2.5.5	The rate of opening of the vocal folds	72
4.2.6	Conclusions	73
4.3	An improvement to acoustic processing	75
4.4	Study 2 <i>Actors over an 8 month period</i>	76
4.4.1	Introduction	76
4.4.2	Subjects	76

4.4.3	Recordings	76
4.4.4	Accumulated data	77
4.4.5	Results	78
4.4.5.1	The rate of opening and closing of the vocal folds .	79
4.4.5.2	Fundamental frequency	83
4.4.5.3	Sound pressure level (SPL)	85
4.4.5.4	Laryngeal closed quotient and Ratio	85
4.4.6	The relationship between laryngeal closed quotient and Ratio	86
4.4.7	Contrasts between the two genders	88
4.4.8	Individual performance	90
4.4.9	Further correlations	93
4.4.10	Conclusions	98
4.5	The discard of two parameters	99
4.6	Study 3 <i>Singers over a 2.6 year period</i>	100
4.6.1	Introduction	100
4.6.2	Subjects	100
4.6.3	Recording	101
4.6.4	Results	101
4.6.4.1	Laryngeal closed quotient	101
4.6.4.2	Ratio	105
4.6.5	The correlation between laryngeal closed quotient and Ratio	107
4.6.6	Conclusions	110
4.7	Discussion of the three studies	111
5	A tool for biofeedback	113
5.1	Introduction	113
5.2	The benefit of visual feedback	113
5.3	ALBERT - Acoustic and Laryngeal Biofeedback Enhancement in Real-Time	115
5.4	ALBERT discussion of formal requirements	117
5.5	Access to the ALBERT User Guide	119
5.6	Functional units	119

5.6.1	Introduction	119
5.6.2	An overview of the forms	120
5.6.3	User control over form attributes	124
5.6.3.1	Introduction	124
5.6.3.2	Form visibility	124
5.6.3.3	Form size	126
5.7	An overview of the flow of information	126
5.8	Analysis of the acoustic signal	128
5.8.1	Deriving a measure of the singer's formant	128
5.8.2	Deriving a measure of the sound pressure level (SPL)	130
5.8.3	Display of acoustic signal information	131
5.8.4	User control over acoustic signal analysis	131
5.8.4.1	Introduction	131
5.8.4.2	Primary control form	131
5.8.4.3	Ratio parameter control form	133
5.9	Analysis of the electrolaryngograph signal	133
5.9.1	Introduction	133
5.9.2	Deriving a measure of fundamental frequency and closed quotient	135
5.9.3	Deriving a measure of the peak-to-peak signal amplitude	137
5.9.4	Deriving a measure of the rates of larynx opening and closing	138
5.9.5	The effect of quantisation	139
5.9.6	Display of the electrolaryngograph signal	139
5.9.6.1	Normalised period display	139
5.9.6.2	Animation of vocal folds	141
5.9.7	User control over electrolaryngograph signal analysis	141
5.9.7.1	Introduction	141
5.9.7.2	Primary control form	141
5.9.7.3	Threshold form	144
5.9.7.4	Refractory time form	144
5.9.7.5	Period form	144
5.10	Voice parameters	146

5.10.1	Introduction	146
5.10.2	User control over the parameters	148
5.10.3	Extending the list of voice parameters	149
5.11	Recording acoustic and electrolaryngograph signals	149
5.11.1	Introduction	149
5.11.2	Recording to file	150
5.11.3	Recording to memory	151
5.12	Methods for visualization	152
5.12.1	Introduction	152
5.12.2	An example application	152
5.12.3	User control over the parameters	155
5.12.4	Using colour for real-time feedback	157
5.12.5	The appropriate visual presentation of data	159
5.13	The rate at which information is updated	161
5.13.1	Introduction	161
5.13.2	User control over the update rate	162
5.14	Saving and loading ALBERT configurations	163
5.15	Some example real-time applications	164
5.15.1	Examples based on the state of one parameter	164
5.15.1.1	Adult singers	164
5.15.1.2	Child singers	164
5.15.1.3	Subjects with short attention span	165
5.15.2	Examples based on the performance of two voice parameters in combination	166
5.15.2.1	Introduction	166
5.15.2.2	The display of CQ with regard to SPL changes	166
5.15.2.3	The display of Ratio with regard to SPL changes	167
5.15.2.4	The display of development along a CQ/singer's for- mant continuum	167
5.16	An examination of the product compared to the original requirements	169
5.17	Conclusions	171

6	An assessment of biofeedback	172
6.1	Introduction	172
6.2	Tuition	173
6.3	Recording	174
6.4	Results	175
6.4.1	Fundamental frequency	175
6.4.2	SPL	178
6.4.3	CQ and Ratio	180
6.4.4	A comparison between performance of the speaking and singing voice	185
6.5	Discussion	185
6.6	Conclusions	189
7	Conclusions	191
7.1	An overview of research with regard to the original hypothesis . . .	191
7.2	Directions for future research	193
A	Study 1 - programme of work	194
B	Study 1 - assessment procedure	196
C	Study 2 - assessment procedure	198
D	Study 3 - assessment procedure	200
E	ALBERT User Guide	202
F	ALBERT parameter testing	245
F.1	Introduction	245
F.2	A computer program for testing ALBERT	245
F.3	Closed quotient	246
F.4	Fundamental frequency	247
F.5	Rate of opening and closing	247
F.6	Peak to peak amplitude of the signal	249
F.7	Ratio	249

F.8 SPL	250
F.9 Jitter	251
F.10 Electrolaryngograph period shimmer	252
F.11 Acoustic shimmer	253
F.12 Signal generator software code	254
G Progress along a linear correlation	261
G.1 Introduction	261
G.2 Derivation	263
G.2.1 Measurement of position along a continuum	263
G.2.2 Measurement of the distance of the data pair from the continuum	264
Bibliography	265

List of Figures

1.1	The relationship of voice teacher to the performer and to the specialists who support/assist in his/her work (<i>after Hollien, 1993</i>)	3
2.1	The phonation circle (<i>After Perkins & Kent, 1986</i>)	9
2.2	Production of the voice (<i>From Sataloff, 1992</i>)	10
2.3	The three physiological sections of human speech production (<i>After Lieberman & Blumstein, 1988</i>)	11
2.4	General arrangement of the vocal organs (<i>From Crystal, 1987</i>) . .	11
2.5	Physiology of the human head (<i>From Sataloff, 1992</i>)	13
2.6	Waveform measures of subglottal system (<i>From Stathapoulos & Sapienza, 1993</i>)	15
2.7	Three views of the larynx (<i>From Sundberg, 1977</i>)	16
2.8	Anatomy of the larynx (<i>After Perkins & Kent, 1986</i>)	17
2.9	Abduction and adduction of the larynx (<i>After Perkins & Kent, 1986</i>)	18
2.10	Vocal tract profiles for two vowels. <i>The shape is determined by the lip and jaw openings, the tongue shape, the velum, and the larynx. (After Fant, 1960; in Sundberg, 1987)</i>	26
2.11	The pattern of vocal fold movement (<i>From Sundberg, 1987</i>)	27
2.12	Idealised glottal area, flow and spectrum (<i>After Pickett, 1991</i>) . . .	29
2.13	Spectra of simple waveforms (<i>From Dodge & Jerse, 1985</i>)	30
2.14	Average source spectra from untrained voices (filled circles) and professional singers (open circles). <i>The spectra are represented by a curve showing how the spectrum contour deviates from the standard slope of -12dB/octave. (From Sundberg, 1987)</i>	32

2.15	The production of model vowels according to the source-filter theory. <i>An idealized spectrum slope of -12dB/octave is assumed for the glottal airflow spectrum. It is modified by the filter responses of the vocal tract, shown for four different vowels, to produce the final vowel spectrum on the right. (After Pickett, 1991)</i>	33
2.16	The effect of vocal effort on the vowel spectrum. <i>The overall slope of the vowel spectrum is steep for low effort and more gradual for high effort. Second and third formants are relatively more intense under high vocal effort. (After Pickett, 1991)</i>	34
3.1	Changes in spectrum level in four equally wide frequency bands associated with changes in SPL as observed in a singer (left) and non-singer (right) phonating the vowel /ae:/ at deliberately varied vocal loudness (<i>From Gauffin & Sundberg, 1989</i>)	37
3.2	Differences between SPL for covered and open singing of the lowest spectrum partial. The mean values of the difference are shown for six subjects. Fundamental frequency (F0) for each subject is also displayed. (<i>From Hertegard et al., 1990</i>)	38
3.3	Acoustic output spectra of trained and untrained (male) subjects singing [a]	40
3.4	Voice physiology. (a) Lateral view of ring phonation (b) Lateral view of constricted phonation (<i>From Acker, 1987</i>)	40
3.5	Bass, tenor, mezzo and soprano RB profiles. <i>The capitalisation of DOH marks the bottom note of an arpeggio. (From Watson, 1992b)</i>	43
3.6	Peak values for resonance balance measured from Scottish Opera singers. <i>Each bar is annotated with the actual RB level that it represents. The darker shading represents the /i/ vowel. The lighter shading represents the /u/ vowel. (From Watson, 1992b)</i>	44
3.7	Two illustrations of the change in fundamental frequency with age	46

3.8	Superimposition of speech F0-SPL characteristics and phonetograms for untrained male (top left) and female (top right) and trained male (bottom left) and female (bottom right) vocal groups. <i>Data points correspond to 10% increments of the F0 range (in Hz). (From Awan, 1993)</i>	49
3.9	SINGAD screen displays <i>(From Rossiter et al., 1993)</i>	51
3.10	Open quotient against age and intensity. <i>(After Stathopoulos & Sapienza, 1993)</i>	53
3.11	<i>Adult male subjects</i> - Mean closed quotient (with standard deviation bar measure) for each subject producing the read passage (left) and the sung scales (right). <i>Subjects are ranked in order of decreasing vocal experience. Mean CQ (%) values are mapped to the y-axis. (From Howard et al., 1990)</i>	54
3.12	<i>Adult female subjects</i> - Mean closed quotient means in each third octave band for the trained and the untrained groups. <i>(From Howard, 1994)</i>	55
3.13	Screen displays of real-time CQ system <i>(From Rossiter et al., 1993)</i>	58
4.1	Relative length and timing of the three studies. <i>Key: xM=number of male subjects. xF=number of female subjects.</i>	61
4.2	Ratio mean and standard deviation measures for all subjects plus tutor. <i>Data assessed prior to training is shown by white bars. Data assessed after training is shown by grey bars. Key: P1=passage 1, P2=passage 2, V=vowel sequence</i>	67
4.3	Fundamental frequency (F0) mean and standard deviation measures for all subjects plus tutor <i>Key: P1=passage 1, P2=passage 2, V=vowel sequence</i>	69
4.4	Closed quotient (CQ) mean and standard deviation measures for all subjects plus tutor <i>Key: P1=passage 1, P2=passage 2, V=vowel sequence</i>	70
4.5	Rate of closing mean and standard deviation measures for all subjects plus tutor <i>Key: P1=passage 1, P2=passage 2, V=vowel sequence</i> . .	72

4.6	Rate of opening mean and standard deviation measures for all subjects plus tutor <i>Key: P1=passage 1, P2=passage 2, V=vowel sequence</i>	73
4.7	This figure shows example time-synchronised acoustic and electro-laryngograph signals (labelled 'SP' and 'LX' respectively) with vertical bars used to indicate buffer start and end times. Processing of a buffer of acoustic data only takes place if voiced phonation is detected in the corresponding buffer of electro-laryngograph data. <i>In this example a buffer update rate of 25Hz is assumed.</i>	75
4.8	Study 2 - data recorded and analysed	78
4.9	Study 2 - recording intervals	79
4.10	Results for each of the six parameters for male, female and both genders combined. <i>The x axis represents the number of the recording and may be regarded as an approximate indicator of time and training.</i>	80
4.11	Results for each of the six parameters for male subjects. <i>The x axis represents the number of the recording and may be regarded as an approximate indicator of time and training.</i>	81
4.12	Results for each of the six parameters for female subjects. <i>The x axis represents the number of the recording and may be regarded as an approximate indicator of time and training.</i>	82
4.13	Correlation between CQ and spectral ratio from longitudinal data set. <i>The G data represents the increment in Ratio for a 1% increase in CQ.</i>	87
4.14	Pattern of change of mean CQ/Ratio for all male subjects, for all five assessment tasks. <i>Spline interpolation (Hearn & Baker, 1986) is used to indicate one possible path of development between the data points. The third and fourth data points in the series correspond to a period of high trauma.</i>	89
4.15	Distribution of CQ/Ratio means for each male subject for each vocal task	91

4.16	Individual patterns of change of all five assessment tasks mean CQ/Ratio data. <i>Spline interpolation is used to indicate one possible path of development between the data points. For each subject, the first data point is marked 'start' and the last data point marked with the number of the recording.</i>	92
4.17	Scatterplot and linear correlation of F0 and SPL	94
4.18	Scatterplot and linear correlation of SPL and CQ	95
4.19	Scatterplot and linear correlation of CQ and F0	96
4.20	Scatterplot and linear correlation of SPL and Ratio	97
4.21	Scatterplot and linear correlation of F0 and Ratio	98
4.22	Measures of CQ across the recording sessions for the singing voice assessment	102
4.23	Measures of CQ across the recording sessions for the speaking voice assessment	103
4.24	Scattergrams of CQ against F0 for subject M1. <i>The number underneath each figure indicates the number of weeks that have elapsed since the first recording.</i>	104
4.25	Measures of Ratio across the recording sessions for singing voice assessment	105
4.26	Measures of Ratio across the recording sessions for speaking voice assessment	106
4.27	Scattergrams of Ratio against F0 for subject M1. <i>The number underneath each figure indicates the number of weeks that have elapsed since the first recording.</i>	107
4.28	Individual patterns of singing voice CQ/Ratio development, with superimposed data ranges for speaking voice. <i>The first recorded assessment is marked 'Start'. The last recorded assessment is marked with the number of weeks since recording began. Spline interpolation has been used to indicate one possible path of progress between the data points.</i>	108
4.29	Plot of Ratio against CQ for assessment of the spoken passage	109

4.30	Scatterplot and line of best fit of Ratio and CQ for assessment of the singing voice, for the 6 male subjects assessed in study 2 and the 3 male subjects assessed in study 3.	111
5.1	The phonation circle of Perkins & Kent (1986) with the domain of ALBERT superimposed	115
5.2	An overview of the ALBERT system process	116
5.3	An example configuration of ALBERT for voice source analysis . . .	121
5.4	An example configuration of ALBERT for acoustic output analysis .	122
5.5	User control over form visibility, width and height. The control on the left can be used for managing the visibility of each form. The control on the right can be used for managing the width and height of each form.	125
5.6	Overview of the system process. <i>Key: ADC = Analogue to digital converter</i>	127
5.7	Fourier transform on a sequence of windows. <i>R represents the sampling frequency of the signal. R/2 is called the Nyquist frequency. It is the theoretical limit on the highest frequency that can be represented in a sampled digital signal (From Dodge and Jerse, 1985).</i>	128
5.8	Spectrogram of a trained singer, with time aligned graph measure of the accumulated fft coefficient amplitudes in the region 2-4KHz relative to the accumulated amplitudes across the range 0-5KHz. . .	129
5.9	The three standard display modes for voice source and acoustic output signals illustrated for 'trained' and 'untrained' modes of phonation	132
5.10	Acoustic analysis control form	133
5.11	Use of the electrolaryngograph	134
5.12	Cross-section of the neck at the level of the larynx indicating current path across the vocal folds for open and closed glottis (<i>After Garner, 1993</i>)	134
5.13	The derivation of fundamental frequency and closed quotient from electrolaryngograph and airflow waveforms	136

5.14	Deriving a measure of the rate of larynx opening and closing from the electrolaryngograph waveform	138
5.15	Accuracy of deriving fundamental frequency from the electrolaryngograph output for phonation at 1KHz and 125Hz digitally sampled at 48KHz	140
5.16	Animation sequences illustrating two contrasting levels of CQ	142
5.17	Voice source control form	143
5.18	Voice period control form	145
5.19	ALBERT parameter control window	147
5.20	User control form for file record and playback	150
5.21	Two example visualization configurations and a complimentary colour scheme with application to the real-time display of jitter	154
5.22	ALBERT visualization control window	156
5.23	Real-time graph display of fundamental frequency with levels of closed quotient (CQ) expressed by colour changes	158
5.24	Mother-child interaction (<i>From Trevarthen, 1983</i>)	160
5.25	User control form for saving and loading ALBERT configurations . .	163
5.26	An example colour scheme for use in mapping <i>time</i> as a mechanism to encourage attention	165
5.27	Mean longitudinal CQ and Ratio data for 6 male subjects singing the nursery rhyme 'Baa Baa Black sheep' across an eight month period of vocal training (from voice study 2)	168
6.1	Bar graph display used for real-time biofeedback	173
6.2	Measures of F0 across the recording sessions for the spoken passage relative to the lowest value in the data set	175
6.3	A comparison of mean speaking voice F0 with singing voice F0 range assessed after the final tuition session	177
6.4	Measures of SPL across the recording sessions	178
6.5	Measures of CQ across the recording sessions	180
6.6	Measures of Ratio across the recording sessions.	181

6.7	Plot of Ratio against CQ for assessment of the spoken passage with line of best fit. <i>Spline interpolation is used to indicate one possible pattern of development between data points.</i>	182
6.8	Plot of Ratio against CQ for assessment of the sung scale. <i>Spline interpolation is used to indicate one possible pattern of development between data points. The domain of the two parameters for speaking voice assessment shown in the previous figure are superimposed.</i> . .	183
6.9	Representative images of electrolaryngograph and spectrum activity across the seven recordings for subject B	184
6.10	Plot of Ratio against CQ for assessment of the spoken passage with data from longitudinal voice study 2 shown superimposed.	186
6.11	Plot of Ratio against CQ for singing assessment with data from a longitudinal voice study 2 shown superimposed	186
6.12	Changes in CQ and Ratio from one assessment to the next for subject B	188
G.1	Deriving a measure of advancement along a line in 2D space	262

List of Tables

3.1	The relationship of vocal technique ratings to singing range, from very deficient technique (top row) to good technique (bottom row). <i>(F)</i> denotes subjects were female only. Singing range is in semitones. <i>(After Teachey et al., 1991)</i>	47
4.1	Experimental subjects for study 1	63
4.2	Summary of mean change for passage 1, passage 2, and vowel sequence for male, female and both genders combined	66
4.3	Experimental subjects for study 3	100
5.1	Example visualization configurations and function presented in the scientific literature, and example configurations and applications in which ALBERT may be used. <i>Order of entries: y-axis followed by x-axis, coloured black unless specified as scattergram. 'LX' refers to the signal derived from an electrolaryngograph. 'Ratio' refers to the level of spectral amplitude in the singer's formant region relative to all voice spectrum output. 'SNR' stands for signal-to-noise ratio.</i> . .	153
F.1	Assessment data for CQ values at input CQ levels of 5%, 50% and 95% across a range of input F0 values	246
F.2	Assessment data for F0. Mean values are shown across a range of input F0 values	247
F.3	Assessment of rates of opening and closing for a selection of frequency values	248
F.4	Assessment of the peak-to-peak amplitude for a selection of input peak-to-peak amplitude values	249

F.5	Assessment of the Ratio parameter	250
F.6	Assessment of the sound pressure level	250
F.7	Assessment of the Jitter parameter	251
F.8	Assessment of the electrolaryngograph Shimmer parameter	253
F.9	Assessment of the Shimmer parameter	254

Acknowledgements

I would like to thank my wife Nancy Sui Wan and daughter Emma for their considerable understanding and patience during the evolution of this work.

The work presented in this thesis has benefitted from the good advice of many people. Predominant amongst them are those from my supervisor, Dr David Howard, who provided consistently helpful and astute guidance on innumerable occasions. Others who have helped over the years include Tim Anderson, Chris Barnes, Mike Edgington, Michelle Evans, Paul Garner, Ian Gibson, Andy Hunt, Anna Katrami, Paul McGilly, Mark Pearson, Paul Stringer, Miles Osborne, John Tuffen, and Sean Watkinson. I am very grateful to them all.

I have received support in different forms from several people in the voice community, including Martyn Clark, Roz Comins, Mike DeCosta, Susan Stern, and Colin Watson. Their ideas and encouragement have helped me on many occasions.

I received excellent constructive criticism concerning thesis proofs from David Howard and Miles Osborne, to whom I am grateful. I would also like to thank Andy Hunt, Martyn Clark, Robert Dormer and Paul Garner for useful comments.

Over the years I have received much support and helpful advice from staff at the York University Computing Services Department, including John Byrne, Dave Atkin, and especially Mike Brudenell and Rob Fletcher.

Thanks are due to people arising from particular areas of the research:

Chapter 4, longitudinal voice study 1.

The invitation by Roz Comins of the Voice Care Network UK to record her students at the University of Warwick, and her subsequent guidance in describing course content and assessment, reproduced in Appendix A, were of great benefit. Thanks must also go to the subjects of the survey for kindly letting me take up their valuable time, and to Michelle Evans for help in recording the first session.

Chapter 4, longitudinal voice study 2. Thanks must go to the subjects and staff at the acting school where the recordings took place for their tremendous cooperation during the course of this study.

Chapter 4, longitudinal voice study 3. I would like to express my thanks for the subjects of this study for agreeing to be recorded over such a length of time, often at some inconvenience to themselves. Thanks also to Michelle Evans for help with recording and to James Angus for technical assistance.

Chapter 5. The development of ALBERT has been influenced by several people. Contributions for code were given by Mark Pearson, Ian Gibson, Stuart Priest and Paul Danset. Conversations with David Howard, Mike DeCosta, John Tuffen, Tim Brookes and others contributed to the design of ALBERT. Their input was most valuable. The contents of Appendix G are based upon notes made by Tim Brookes. Peta White, of the Roehampton Institute, was very helpful in providing information on developing child singers used as the basis for one of the example applications.

Chapter 6. Mike DeCosta, of the DeCosta Singing Academy in York, played a key role in agreeing to use the ALBERT system in the course of his tuition process. His enthusiasm and ideas were of great benefit. I would also like to express my gratitude to Valerie Tiplady and Andy Hunt who played supporting roles in the biofeedback study, and to Tim Brookes and Tim Anderson, who agreed to be subjects.

Declarations

I hereby declare that this thesis is not the same as that of any that I have submitted for a degree, diploma, or other qualification at any other University or similar institution.

I declare that no part of my dissertation has already been or is being concurrently submitted for any degree, diploma, or other qualification.

I further state that this research and its results are for the most part my own. Where I have adopted theories or results of others I state this clearly in the text and provide references.

I declare that part of this research has previously been published and that part is currently in submission for journal publication. These papers are listed below.¹

1. D ROSSITER, D M HOWARD & R COMINS, (1994) Objective measurement of voice source and acoustic output change with a short period of vocal tuition. *Accepted for publication by Voice* [Chapter 2 - Voice study 1]
2. D ROSSITER & D M HOWARD, (1994) An investigation into the vocal development of a group of actors. *Submitted to the Journal of Speech and Hearing Research* [Chapter 2 - Voice study 2]
3. D ROSSITER & D M HOWARD, (1994) Voice source and acoustic output qualities for singing synthesis. *Proceedings of the International Computer Music Conference*, pp. 483-484 [Part of paper included in Chapter 2 - Voice study 3]
4. D M HOWARD & D ROSSITER, (1992) Results from a pilot longitudinal study of electrolaryngographically derived closed quotient for adult male singers in training. *Proceedings of the Institute of Acoustics*, vol. 14, no. 6, pp. 529-536 [A subset of Chapter 2 - Voice study 3]
5. D ROSSITER & D M HOWARD, (1994) What happens as singers develop their voices? A study into longitudinal voice development. *Submitted to the Journal of Voice* [Chapter 2 - Voice study 3]

¹Published papers are available for inspection via the world wide web at <http://www.york.ac.uk/~elec10/>

6. D ROSSITER & D M HOWARD, (1994) ALBERT - A system for interactive analysis and display of voice source and acoustic parameters. *Proceedings of the Institute of Acoustics*, vol. 16, no. 5, pp. 301-308 [Chapter 5]
7. D ROSSITER & D M HOWARD, (1994) The design, implementation and usage of ALBERT, a real-time visual feedback computer tool for voice analysis and training. *Submitted to the Journal of Voice* [Chapter 5]
8. D ROSSITER & D M HOWARD, (1994) An approach for the use of computers for singing tuition. *To appear in Musical Praxis*, vol. 1, no. 2 [Parts of paper included in Chapters 1, 2 and 5]
9. D ROSSITER & D M HOWARD, (1994) Animation of the vocal folds derived from an electrolaryngograph signal. *Voice*, vol. 3, no. 2, pp. 86-91 [Part of paper included in Chapter 5]
10. D ROSSITER & D M HOWARD, (1994) The application of multimedia and visualization techniques to the enhancement of voice analysis and development. *Proceedings of the 12th Eurographics UK conference*, pp. 5-17 [Part of paper included in Chapter 5]
11. D M HOWARD & D ROSSITER, (1993) Real-time visual displays for use in singing training: an overview. *Proceedings of the Stockholm Music Acoustics Conference*, pp. 191-196 [Part of paper included in Chapter 5]
12. D ROSSITER & D M HOWARD, (1992) Cyberspace visualization of vocal development data. *Proceedings of the Institute of Acoustics*, vol. 14, no. 6, pp. 241-248 [Contains some voice visualization ideas expressed in Chapter 5]
13. D ROSSITER, D M HOWARD & M DECOSTA, (1994) Voice development with and without the use of visually presented biofeedback. *Submitted to the Journal of the Acoustical Society of America* [Chapter 6]

Chapter 1

Research aims and report structure

1.1 Introduction

This chapter introduces the context of the research presented in this thesis. The reader is introduced to the predominance of vocal abuse in many professional fields, and the ambiguous techniques frequently employed in the process of vocal tuition to counter this. A need for objective assessment is identified, and a proposal presented for the role of real-time feedback. An overview of the research presented in this thesis follows: 1) the identification of voice parameters which change with the process of vocal tuition; 2) the design and implementation of a suitable real-time system; and 3) a formal assessment of the benefits of the real-time system based upon (2) when used as a tool for vocal tuition.

1.2 Vocal use and abuse

The voice is used in everyday life in many different contexts. For some people, their profession demands a higher level of voice usage. Examples include doctors, politicians, teachers, and telephone operators. These people often experience a high level of unintended vocal abuse, often with consequential vocal damage.

Teachers are prime examples of professional voice users who can suffer from vocal health problems. Comins (1992) found that 34% of voice patients in some speech

and language clinics were teachers. After a recent landmark tribunal decision was made to compensate a teacher who lost her voice during the course of her work, it was reported that “a flood of similar claims” was predicted (Morrow, 1994). One comment from a speech and language therapist highlights the extent of the problem;

“These are not just numbers; they represent people who have not been advised about a problem to which teachers are prone. They represent classes without their teacher, or taken by a supply teacher at a cost to the school of over £90 a day. Speech and language therapists know from cases that some teachers will damage their voices permanently by continuing to teach with laryngitis, whilst others may, mistakenly, be given a course of antibiotics, when their basic problem is not infection but inefficient voice production.” (Comins, 1992)

1.3 Deficient voice tuition

Almost without exception, all instances of abuse could be avoided with appropriate education and training (Morrow, 1994). However, vocal training is itself open to much ambiguity and abuse. The voice teacher needs to embody principles from a wide field of expertise (see figure 1.1). In practice, however, it is often found that there is a bias towards subjective appreciation (generally shown as the left hand side of figure 1.1) rather than objective assessment (the right hand side of figure 1.1).

For example, the English National Curriculum has recently been expanded to require the objective assessment of the oral skills of pupils. Teachers attempting to satisfy this requirement were reported as realising that they did not have the knowledge or ability to suitably assess the qualities under consideration (Stables, 1992). Quoted comments include “...if we knew enough about the processes involved we could actually give that child the help” and “I really question what we are looking for in terms of actual proficiency”. This is one indication of considerable deficiencies in conventional voice tutoring systems. Vocal damage may be a direct consequence. In one formal assessment 47% of voice students were shown to have sought medical help for voice problems (Sapir, 1993).

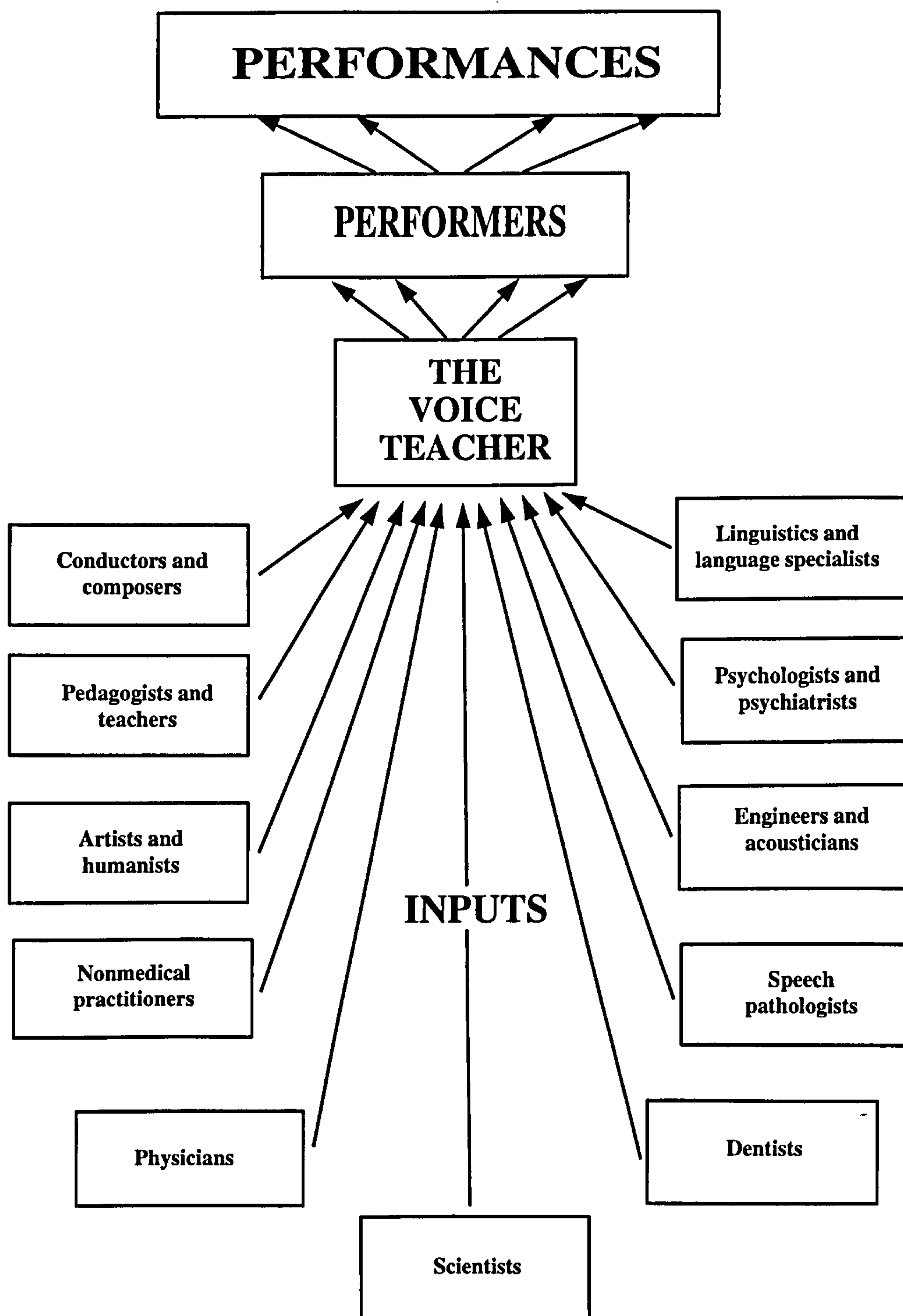


Figure 1.1: The relationship of voice teacher to the performer and to the specialists who support/assist in his/her work (after Hollien, 1993)

1.4 Ambiguous tuition techniques

The problems indicated as prevalent in the field of voice tuition may well be due largely to the various styles of tuition adopted. Some forms of tuition are ambiguous and subjective in their approach. This is illustrated in an example taken from the field of singing tutorage:

“I was once told to imagine the smell of a freshly cooked pork chop before embarking on ‘Un Aura Amorosa’ from Mozart’s Cosi fan tutte, ...the memory of pork chops I have known still disturbs my concentration when I sing the part of Ferrando.” (Lavender, 1994)

Techniques employing mental imagery are frequently employed. This is illustrated in the following extract:

“I have frequently used the following mental images in dealing with the difficulties in the phrases immediately preceding the high C-flat in the “O Don Fatale” aria of Eboli from Don Carlos: On the third beat (a rest), two bars preceding the high C-flat, immediately preceding the triplets which lead up to the high note, I imagine a water wheel with water flowing over it. Directly preceding the high note, during the inspiration, I imagine a well at the back of my head into which the sun is shining. The water in the well feels warm, the rays of the sun (the breath during inspiration) are causing the space at the back of my head.” (Quoted in Immoos, 1993)

There are indications that the use of mental imagery in the process of vocal tuition may not be as appropriate as its prevalence of use might suggest (Anderson & Helstrup, 1993). For example, Reed & Johnsen (1975) found that it is more difficult to find a hidden part in a mental image of a complex figure than in a physical percept of the same figure. Chambers & Reisberg, (1985) reported that subjects were unable to “see” the alternative interpretation of classical ambiguous figures in their mental images of the figures, but were able to see this as they drew their images on paper. Clearly, these points raise questions concerning the suitability of the use of mental imagery as a form of reliable guidance.

Orton (1938) epitomises the desire for objectivity:

“Sims Reeves wrote: ‘The tones in the head register vibrate, especially at the bridge of the nose, and the highest notes convey a sensation as of an electric thrill in the head.’ Such a description, and many similar ones, surely demonstrate, by their diversity and ambiguity, the need of tangible guides and tests in attempts at voice culture.” (Orton, 1938)

1.5 The reduction of ambiguity

Ambiguities present in the tuition process may be resolved through the use of an automated system able to *objectively* measure voice usage, and to offer appropriate visual incitement for vocal development.¹ A description of *objective* in this context may be the expression or use of facts without distortion by personal feelings or prejudices.

Such a system would be able to complement the central figure of the voice teacher.

However, in order to create such a system, there must first be some understanding of the natural progression of voice development. To this end, studies are required of people as they undergo vocal training so that parameters indicative of vocal development which may then be embodied in the creation of a voice assessment and development system are identified.

A formalisation of these proposals is contained in the following thesis hypothesis.

1.6 Hypothesis

It is hypothesised that computer systems which

1. objectively measure a large set of vocal attributes,
2. appropriately translate these attributes to a smaller set of easily interpreted measures relating to vocal efficiency, and

¹In this context, visual feedback is a form of *biofeedback*.

3. dynamically display this information in an appropriate graphical form during the process of vocal tuition

may be able to provide a useful objective measure of progress for the advancement of vocal development.

In order to satisfy this hypothesis, the following objectives are to be resolved:

1. the analyses of developing voice users in order to identify parameters indicative of voice development.
2. the development of a software tool for vocal tuition able to embody these parameters in the process of real-time analysis and subsequent visual display.
3. a formal evaluation of the usefulness of the real-time visual feedback system.

1.7 Thesis content and structure

Chapter 2 considers the nature of the voice in order to set the context for a subsequent assessment of vocal parameters which appear relevant to the process of vocal tuition. It describes the anatomy of the voice, and the acoustic operation of the physiological mechanisms of which the voice comprises. Attention is given to the articulatory controls used in phonation, and to the role of the larynx.

Chapter 3 considers voice source and acoustic output parameters which may be suitable for use in a real-time system intended for use as a tool to help improve vocal development. Previous related research addressing the application of these parameters in tools for vocal tuition are discussed.

Chapter 4 presents three studies considering trends in the parameters identified in chapter 3 in voice users as they undergo tuition. In study 1, male and female subjects were analysed before and after a six-week course of vocal expression. Study 2 presents a longitudinal assessment of male and female actors in training over a period of 8 months. Subjects undertook a number of vocal tasks including singing and three different modes of vocal projection. Developmental parameters and gender differences are discussed. Study 3 presents a longitudinal assessment of three male singers in training over a maximum time span of 33 months for both speaking and

singing voice. Taken in aggregate, the studies indicate two parameters in particular which appear to reflect the progress of vocal development.

Chapter 5 presents the design and implementation of a highly configurable assessment and visual feedback system which is able to provide and update a visual representation of the parameters identified in chapters 3 and 4 in real-time as a tool used by a voice tutor for the tuition process.

Chapter 6 assesses the performance of the real-time system when used to provide visual feedback on the developmental parameters identified in chapter 4 in the context of singing tuition.

Chapter 7 presents an overview of conclusions concerning the research, and discusses directions in which future research could go.

The Appendices include the assessment procedures used for the voice studies, the User Guide for the software presented in chapter 5, an assessment of the accuracy of the software in assessing voice parameters, and a description of how the real-time system discussed in chapter 5 may be used to present a real-time display based upon a linear correlation between two parameters.

Chapter 2

Voice anatomy, physiology & acoustics

2.1 Introduction

This chapter provides the background for a subsequent examination of parameters that may be appropriate for analysis by a voice development system. After an overview of the vocal process, the anatomy of the three sections of the vocal system is described.¹ The acoustic consequences of the physiological mechanism of the vocal system is then discussed. Conclusions are drawn concerning the operation of the voice with regard to the design of a real-time visual feedback system for the enhancement of vocal performance.

2.2 The phonation circle

All the processes involved in human speech can be considered in a circular pattern, for example, the “phonation circle” (see figure 2.1) of Perkins & Kent, (1986). This thesis describes work concerned with complimenting the section of the phonation circle concerned with aural reception and related neural processing (top left and top of figure 2.1) with feedback based on real-time analysis of the voice, presented in a

¹There are several instances in this chapter where there is more than one name for a particular object. Where this has been encountered, the primary label is used, with any other name(s) mentioned in brackets when the word is first used.

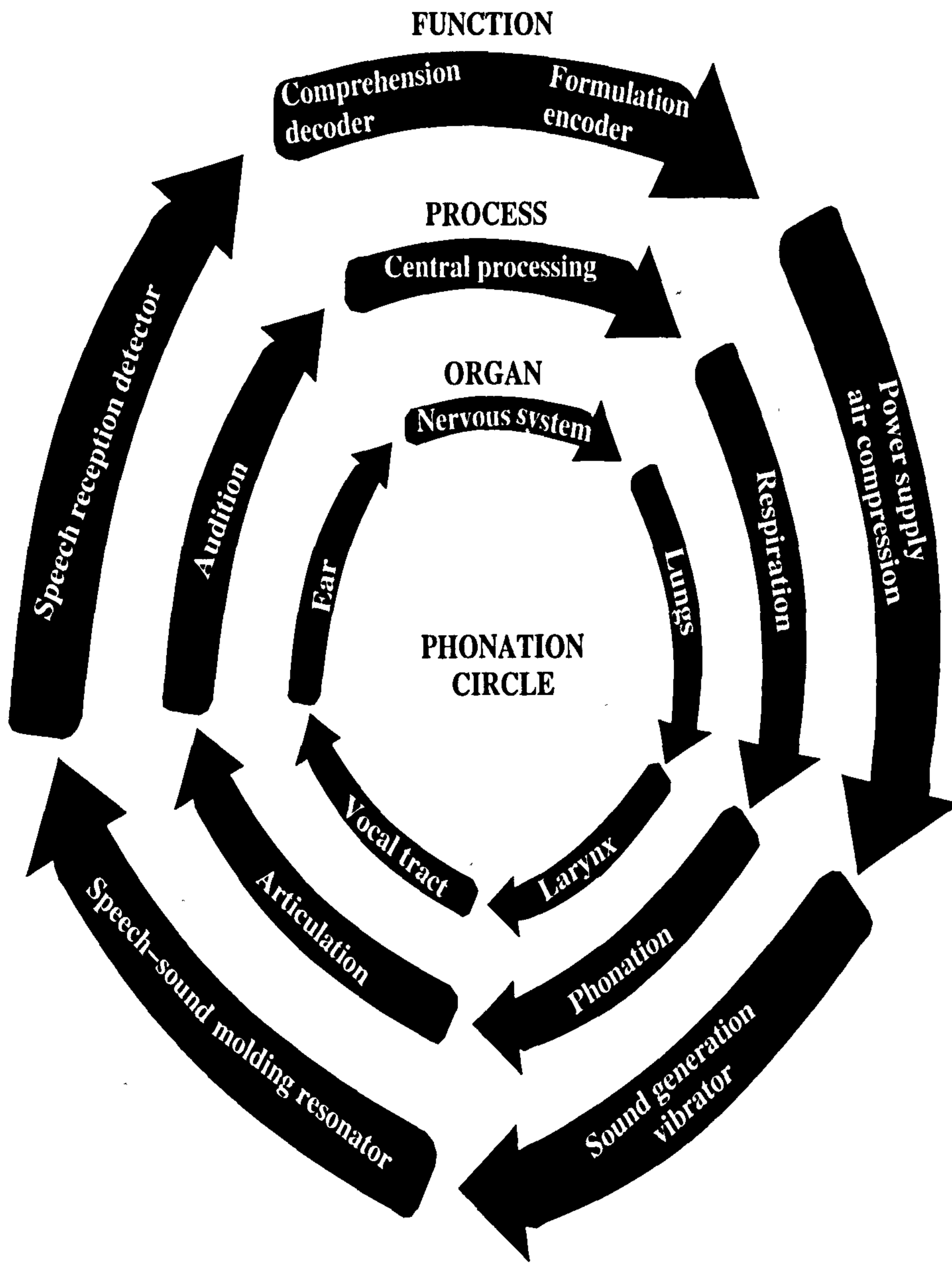


Figure 2.1: The phonation circle (After Perkins & Kent, 1986)

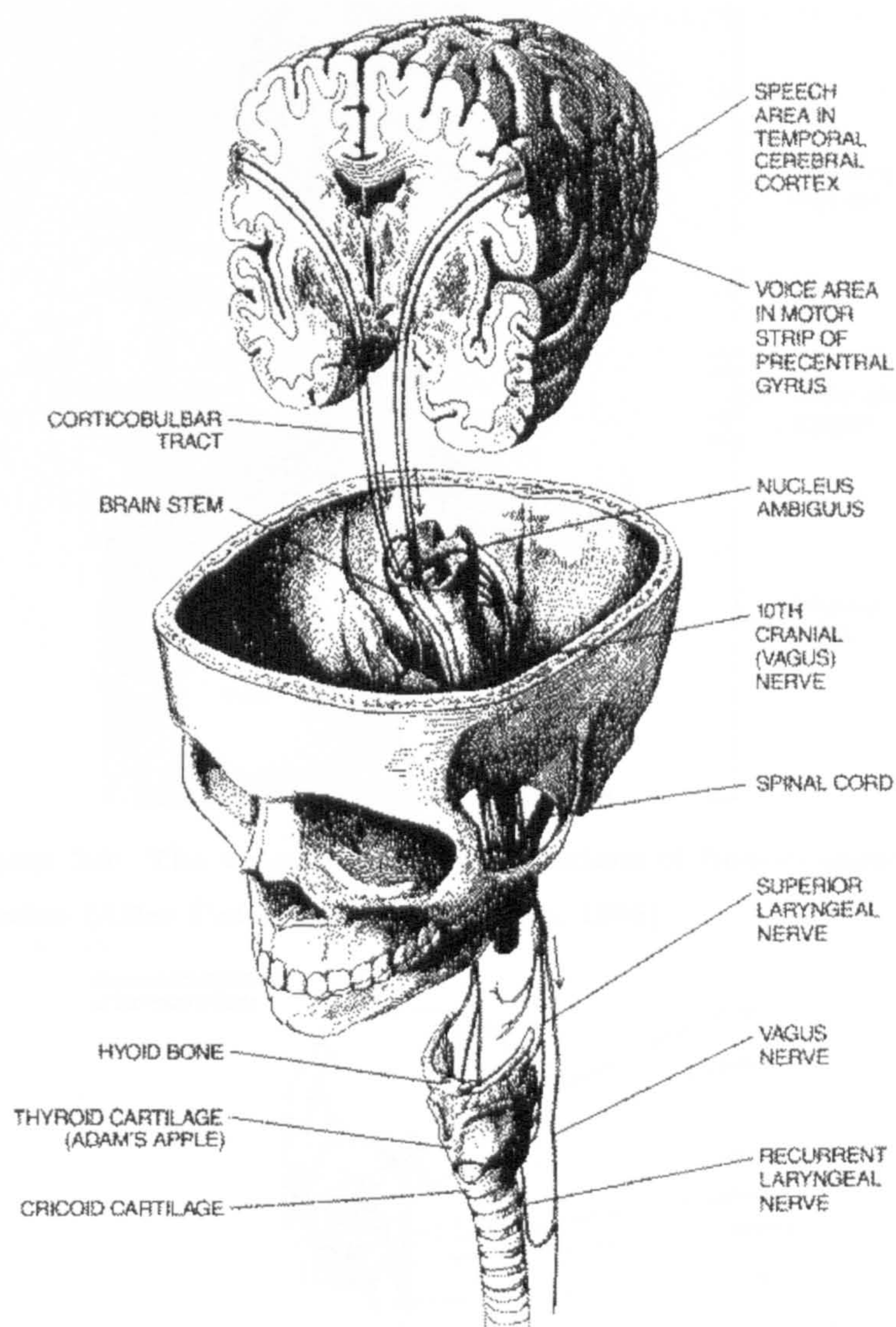


Figure 2.2: Production of the voice (From Sataloff, 1992)

visual form. In preparation for this, the nature of the vocal process must be carefully considered so that parameters relating to vocal tuition which can be monitored by such a system may be subsequently identified.

2.3 An overview of the vocal process

The first signal within the chain of processes for voice production occurs in the cerebral cortex of the brain (for example, in the speech area). This is shown in figure 2.2. It is transmitted to the vocal organs by the nervous system. The organs involved with the production of speech and singing may be divided into three units

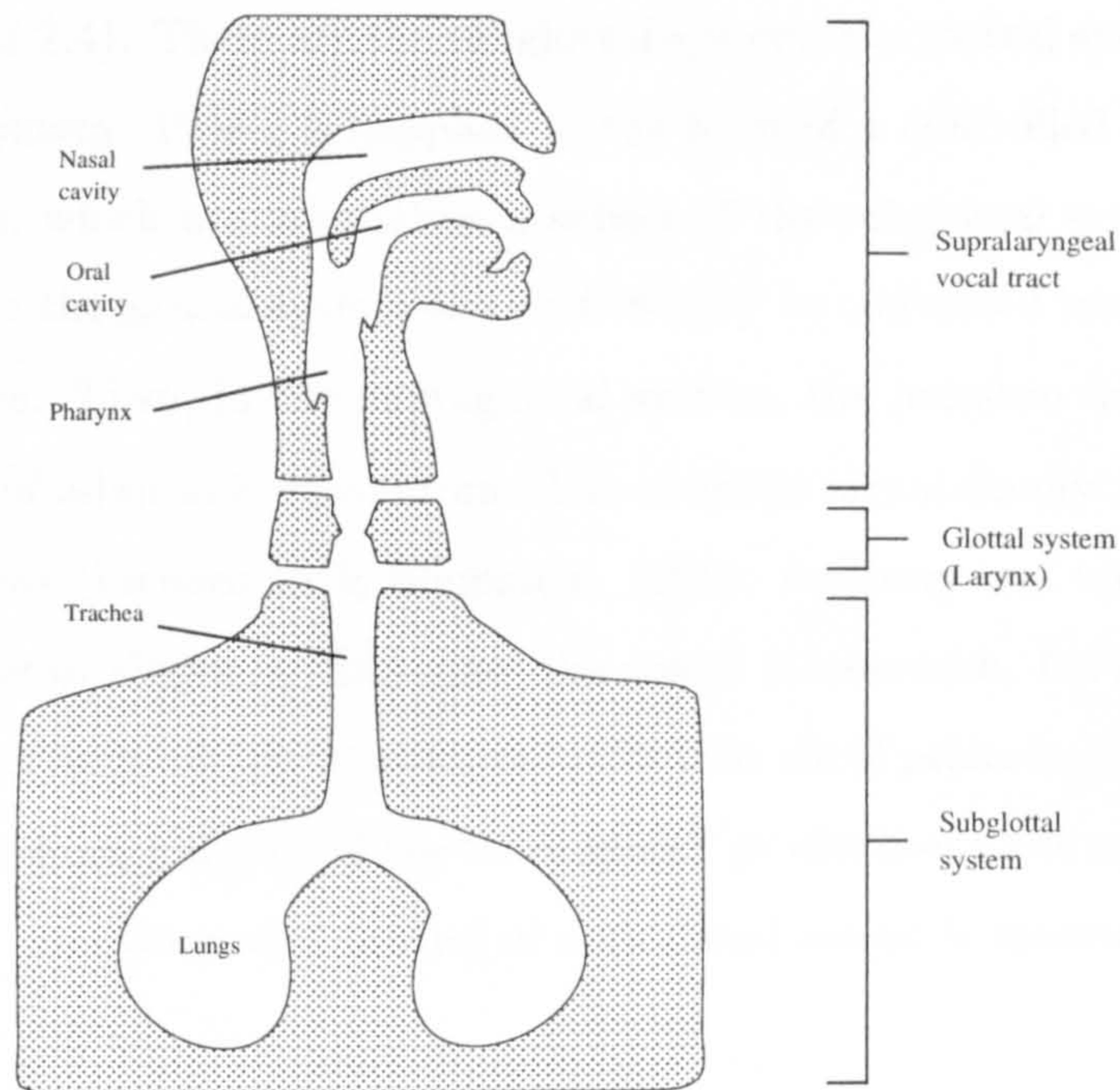


Figure 2.3: The three physiological sections of human speech production (After Lieberman & Blumstein, 1988)

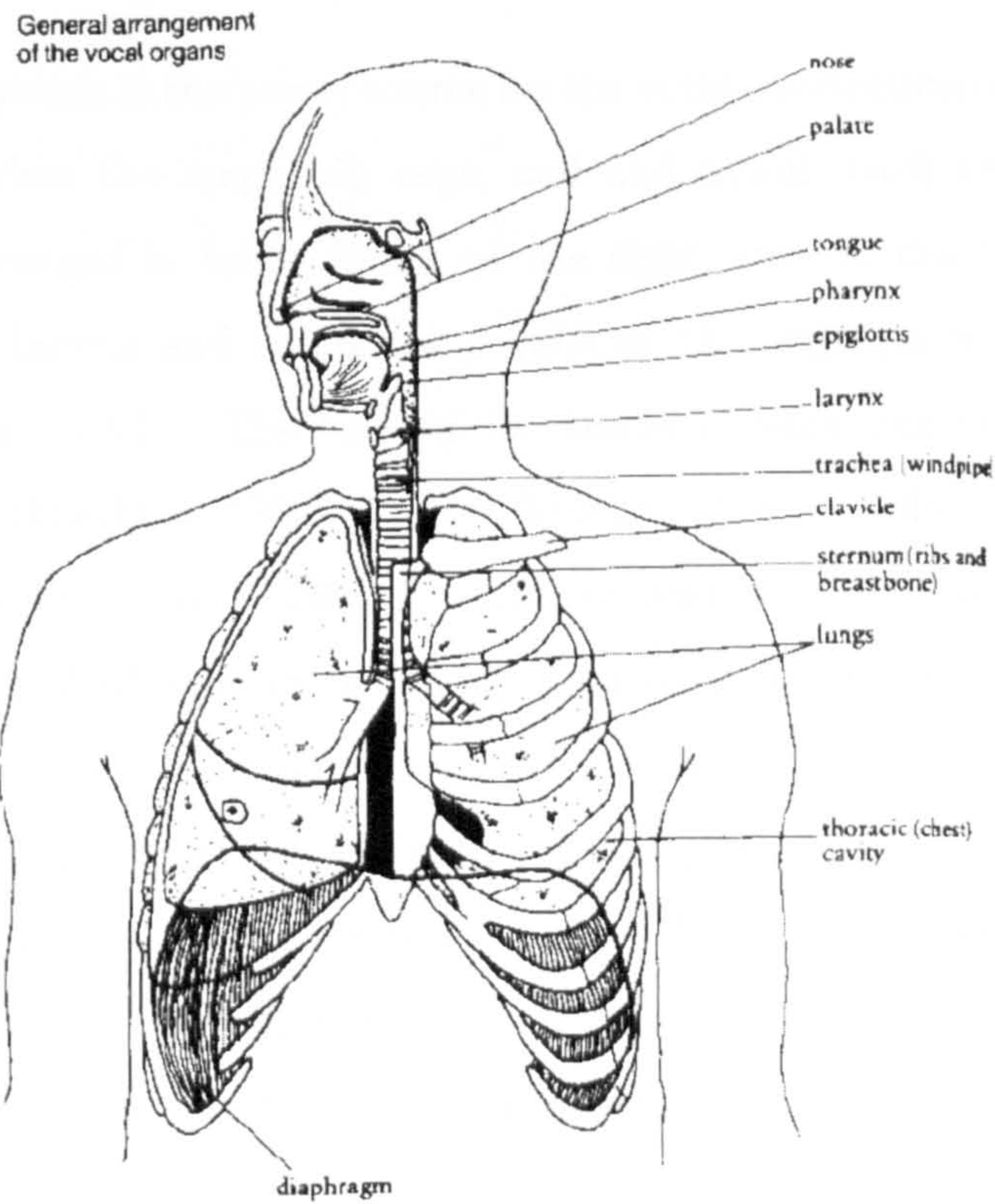


Figure 2.4: General arrangement of the vocal organs (From Crystal, 1987)

(see figure 2.3 and 2.4). These are the subglottal system, the glottal system, and the supralaryngeal system. Power is supplied in the form of a controlled release of air flow by the lungs, which are themselves a subset of the subglottal system. As this air passes through the glottal system the air flow may be converted into a controlled sequence of pulses. Then, in the supraglottal system, the pressure wave is shaped through the use of adaptable resonators. The acoustic signal finally radiates from the mouth and nose (Lieberman & Blumstein, 1988). Auditory and tactile feedback enable the speaker or singer to 'fine-tune' the sound (Lieberman, 1977).

In the following sections, the anatomical units that are of prime importance in the physiological functioning of each of the three speech production units are considered. Their role in the generation and control of articulated sound is assessed.

2.4 The subglottal system

2.4.1 An overview

The subglottal system is the power source for the vocal production of sound (Proctor, 1980). It comprises the lungs, rib cage, and abdominal, back and chest muscles. Each lung is arranged in lobes, three on the right, two on the left. The trachea is joined to the larynx and comprises a tube of 18 cartilages which are enclosed by the trachealis muscle. The base of the trachea stems into two smaller tubes, called bronchi (Stradling, 1968). These bronchi then sub-divide iteratively into progressively smaller tubes, called bronchioles and ducts. Alveolar air sacs are situated at the end of each duct. They pass oxygen into the blood stream on inhalation and release carbon dioxide on exhalation.

The upper cavity of the subglottal system is called the thoracic cavity (chest). This is shown in figure 2.4. The lower section is called the abdomen. The thorax is a barrel-shaped bone and cartilage cage containing the pulmonary system, respiratory passages and the heart. The thoracic cavity contains the mediastinum (mainly the heart, blood vessels, and esophagus) and the two thin, air-tight pleural membranes in which the lungs are encased. The abdomen contains the digestive system, and several other organs that do not play a part in the vocal process, such as the kidneys and liver. The abdomen and thorax are separated by a dome-shaped muscular

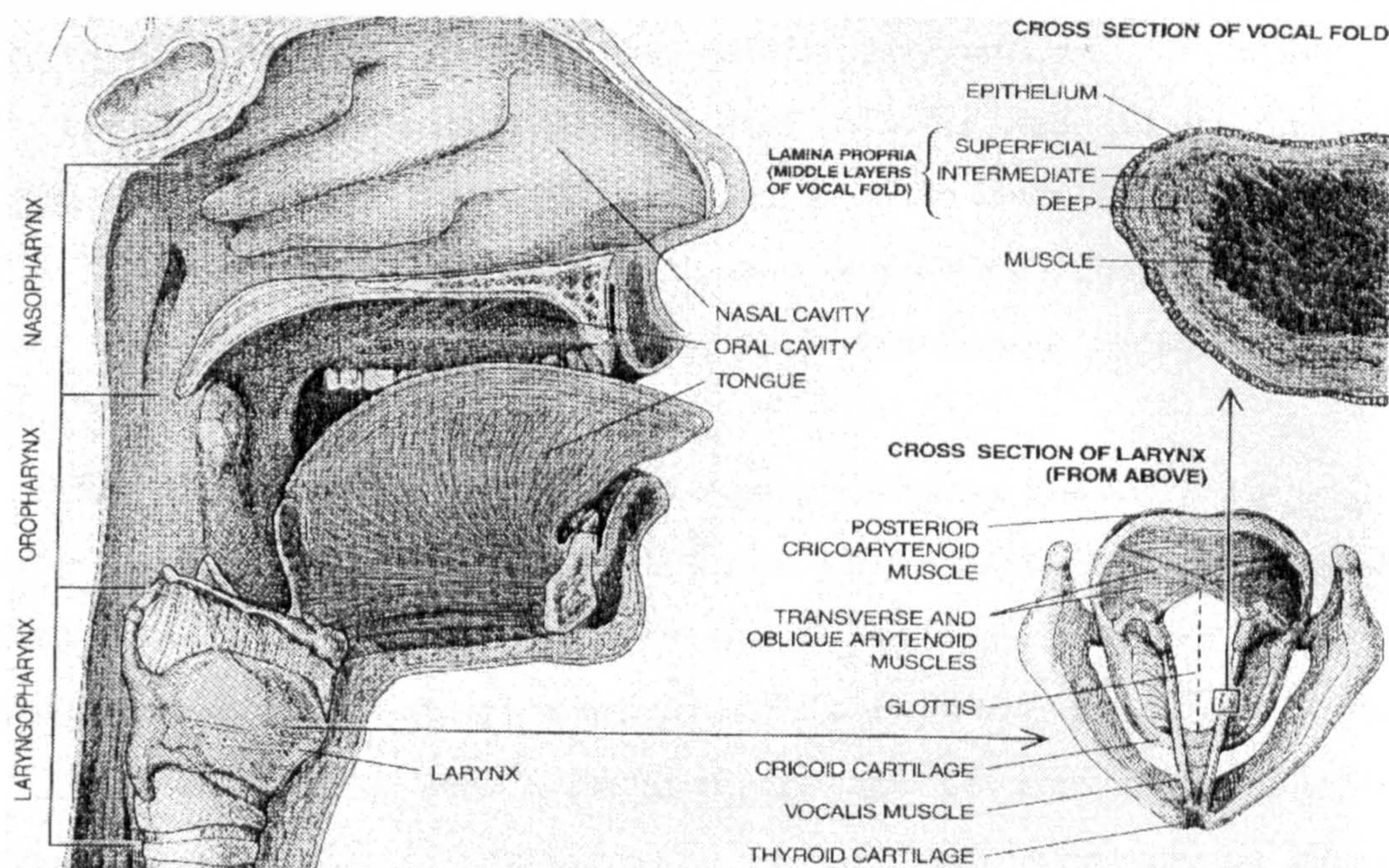


Figure 2.5: Physiology of the human head (From Sataloff, 1992)

structure called the diaphragm.

The chest wall consists of the rib cage, the diaphragm, the abdominal wall, the abdominal viscera, and the muscles of the chest and abdomen. The pulmonary system consists of the network of tubes leading from the trachea (windpipe) to the alveolar air sacs at the outermost reaches of the lungs. The respiratory pump is the term given to the chest wall and the pulmonary system (Hixon, 1987).

The thorax is formed by 12 pairs of ribs attached by cartilage to the 12 thoracic vertebrae of the backbone. In front, the upper 10 ribs attach by cartilage to the sternum (breastbone). The lowest two are unattached. The lower ribs slope downwards from their connection to the sternum. Accordingly, the ribs move laterally outwards when they are raised. The upper ribs are hinged in such a way that they move forward slightly when raised, but do not provide as much expansion as is achieved with elevation of the lower ribs.

2.4.2 Subglottal muscles

The diaphragm is a flat sheet of muscle and tendon connected to the lower border of the rib cage (Proctor, 1980). It is shaped like an inverted bowl. When the rim

muscles contract, they flatten the domes on each side, thereby achieving vertical expansion, and pull up on the lower ribs, moving them outward.

The external intercostal muscles run from the lower border of one rib to the upper border of the rib below, filling the outside of the eleven spaces between the ribs. They function as if they were in a single sheet of muscle pulling all of the lower ribs toward the first rib, which is attached to the first thoracic vertebra and to the base of the skull.

The serratus posterior muscles are flat muscles slanting down from the neck to insert into the second through to the fifth ribs. They help in lifting the upper ribs. This results in slight expansion of the thorax and stabilises the upper ribs for the pull of the external intercostals which provide the major force for raising the ribs.

Other muscles often assist in raising the rib cage. The levatores costarum (rib elevators) are a series of 12 small muscles, each of which inserts into the rib just below the vertebra from which it originates. When contracted they pull up on the ribs, although their influence is small due to their size and poor leverage position (Perkins & Kent, 1986).

2.4.3 Respiration

The lungs are positioned within the thoracic cavity. They are the main organs of respiration. We inhale and exhale in order to move oxygen into the lungs and carbon dioxide from them (Proctor, 1980; Hixon, 1987). For speech and singing, we use this inspiratory-expiratory function to make the chest function as a pressure pump. In order for pressure to be increased, the lungs must be inflated. This is accomplished by creating a partial vacuum which requires expansion of the volume of the chest. The chest may be expanded in three dimensions; vertically, transversely (sideways) and anteroposteriorly (front to back). Vertical expansion is achieved by lowering the diaphragm, which is the floor of the chest cavity. Lateral and anteroposterior expansion result from raising the rib cage. To apply exhalatory pressure, the lungs are squeezed by lowering the rib cage and pushing the diaphragm upwards by muscular contraction of the abdominal muscles (Perkins & Kent, 1986).

The predominant muscles for inspiration (inhalation) are the diaphragm and the external intercostal (rib) muscles. Expiration (exhalation) is largely passive during

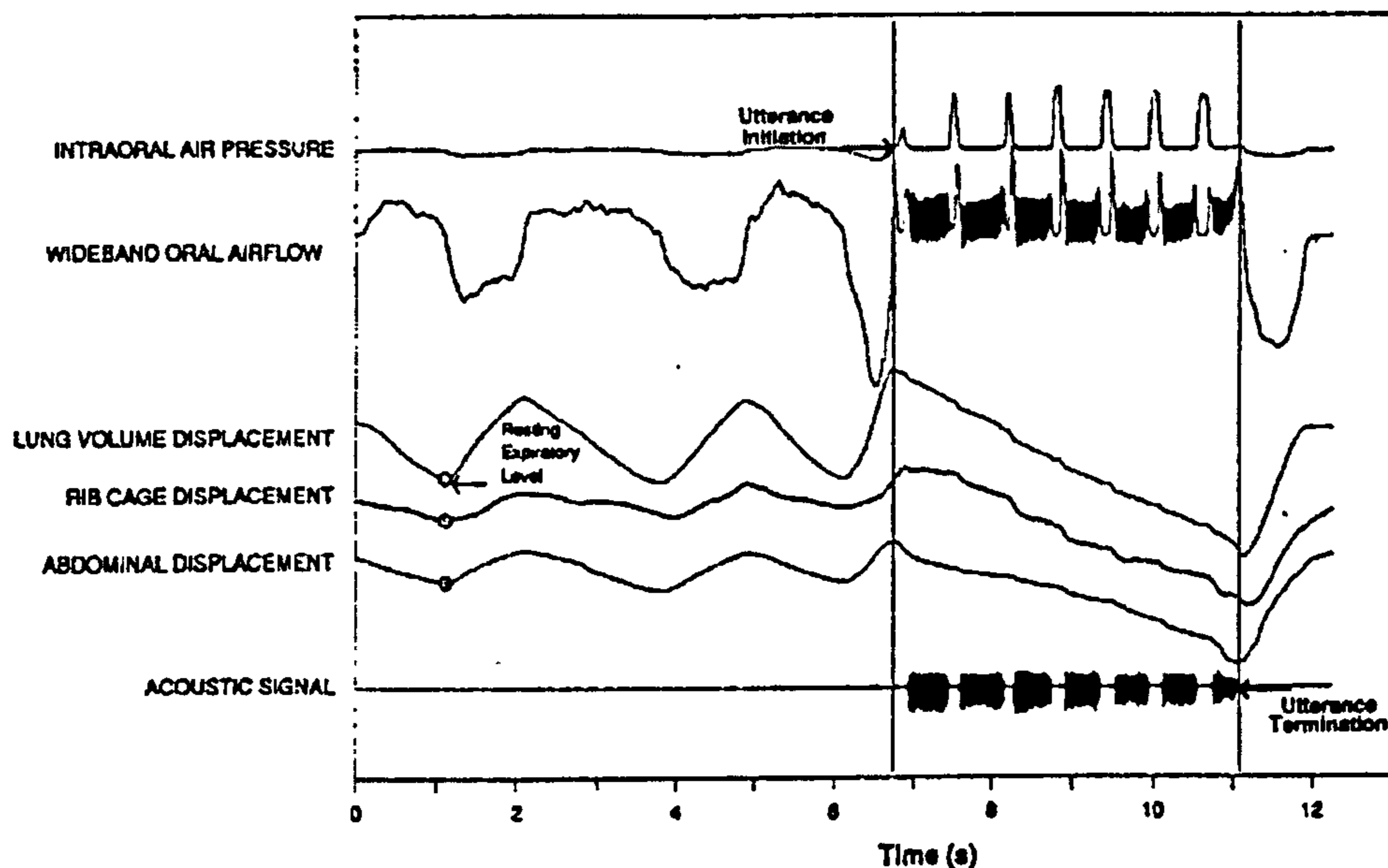


Figure 2.6: Waveform measures of subglottal system (From Stathopoulos & Sapienza, 1993)

normal respiration. The elastic properties of the lungs and rib cage typically force air out of the lungs after inhalation. During active expiration, abdominal muscles may raise the pressure within the abdomen and thereby force the diaphragm upward. Alternatively, they may lower the ribs and sternum to decrease the dimensions of the thorax (MacNeilage et al., 1979).

Figure 2.6 illustrates the pattern of movement for several parameters of the subglottal system. Lung volume, rib cage and abdominal displacement are shown (in addition to intraoral air pressure, oral airflow and the acoustic signal which are measures from the glottal system and supralaryngeal vocal tract). These signals are shown for a period of normal breathing followed by a sequence of several spoken words. In this figure, the strong correlation between lung volume displacement, rib cage displacement and abdominal displacement during the process of breathing can be seen. With inspiration, each of the measures rises, reflecting the use of the rib muscles and diaphragm to expand the lungs and draw air into the body. The abdominal displacement occurs due to the downward movement of the diaphragm. With expiration, each of the measures falls as the elastic forces in the muscles cause

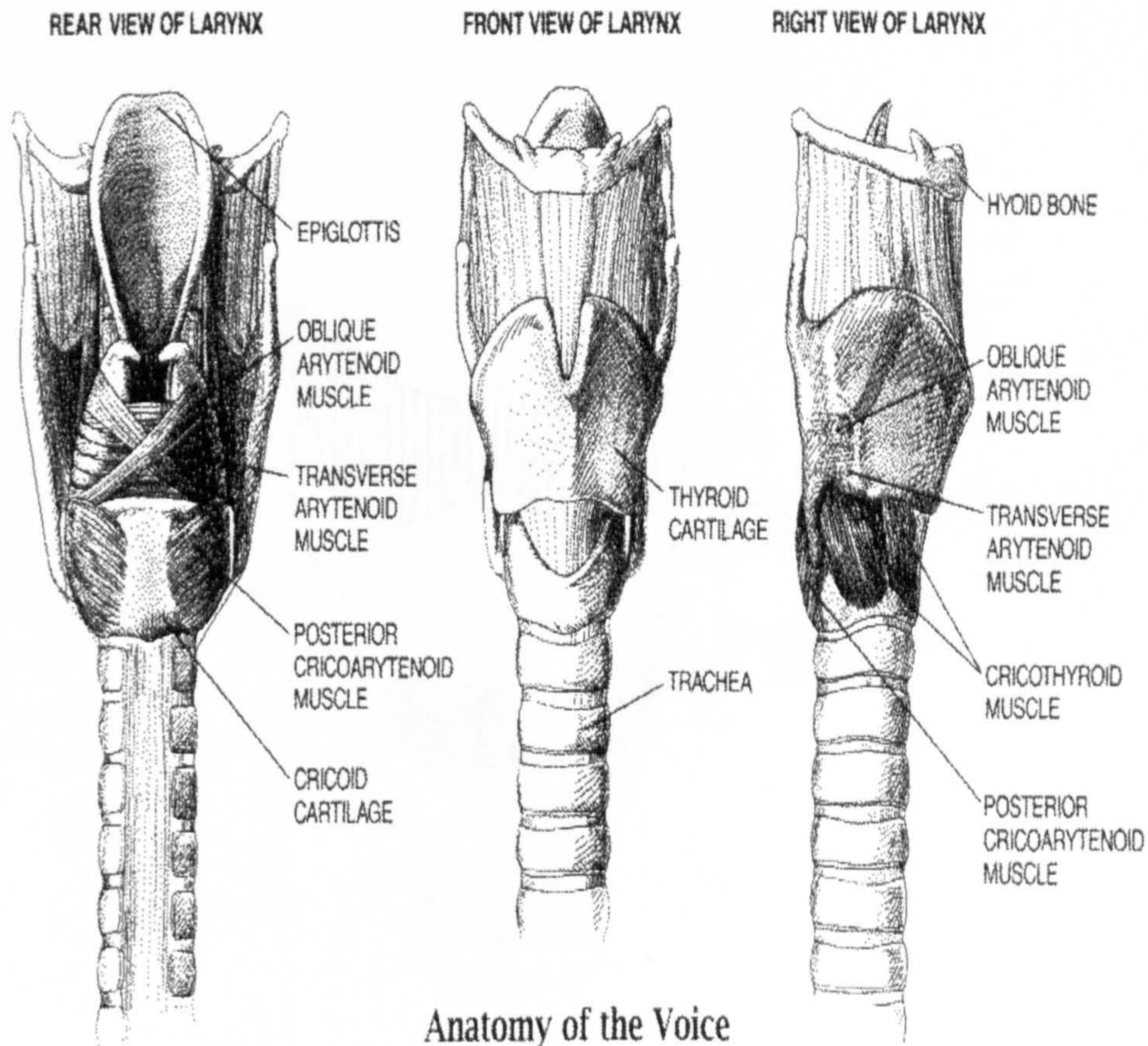


Figure 2.7: Three views of the larynx (From Sundberg, 1977)

the diaphragm and rib structure to return to their relaxed state. Immediately prior to phonation, maximum lung volume displacement is recorded. This reflects inspiration by the subject of a sufficient quantity of air for the subsequent sequence of utterances. During the sequence, a gradual decrease in lung volume displacement, rib cage displacement and abdominal displacement occurs as the air supply required for phonating is used and the volume of air in the lungs decreases. At the end of the sequence, inspiration immediately occurs again in order to refill the lungs.

2.5 The larynx

2.5.1 Introduction

The larynx, illustrated in figures 2.7, 2.8, and 2.9, functions as a valve that connects the respiratory system to the airway passages of the throat, mouth and nose (Rossing, 1990). It has evolved in part to help prevent anything but air from

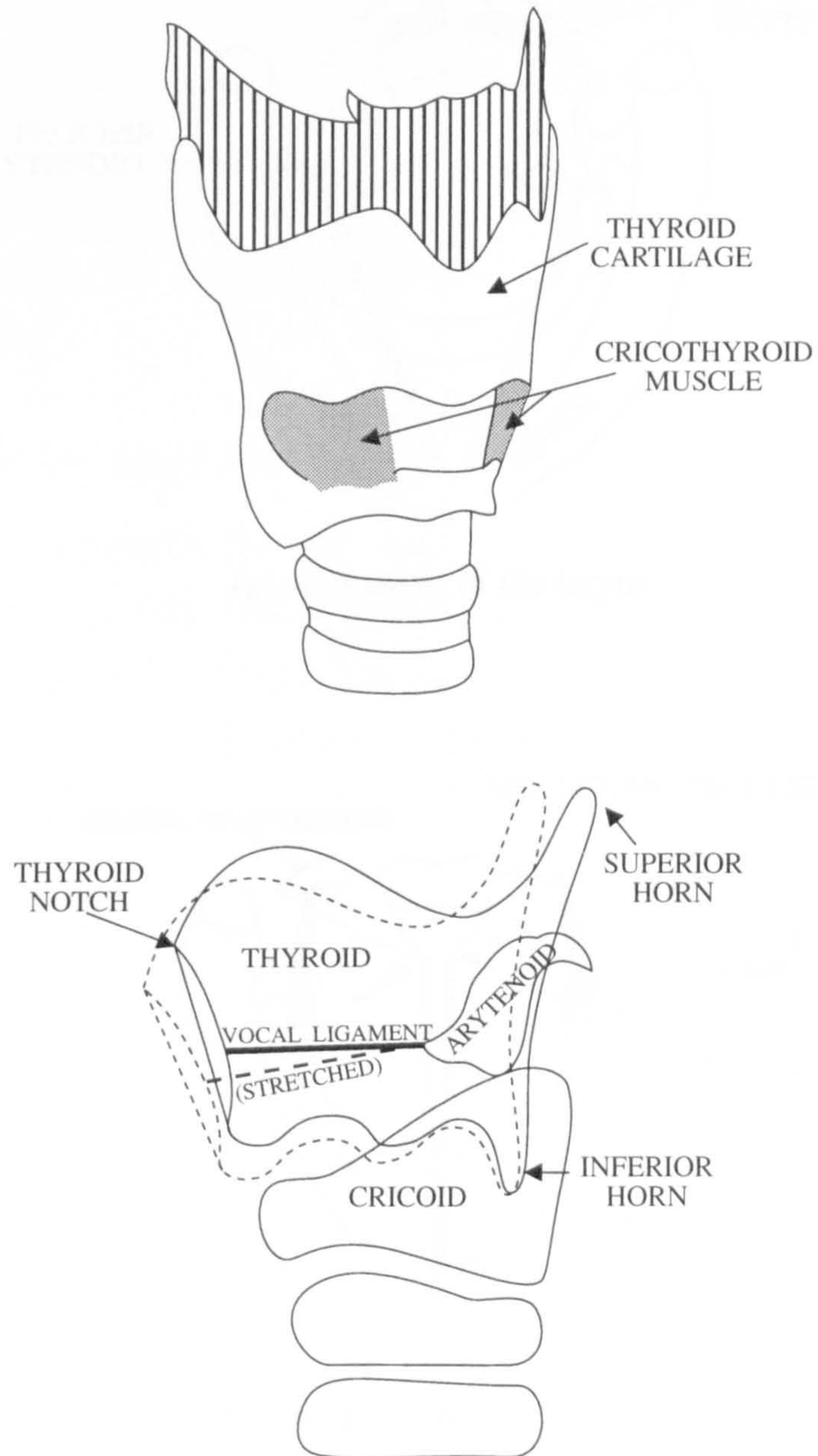


Figure 2.8: Anatomy of the larynx (After Perkins & Kent, 1986)

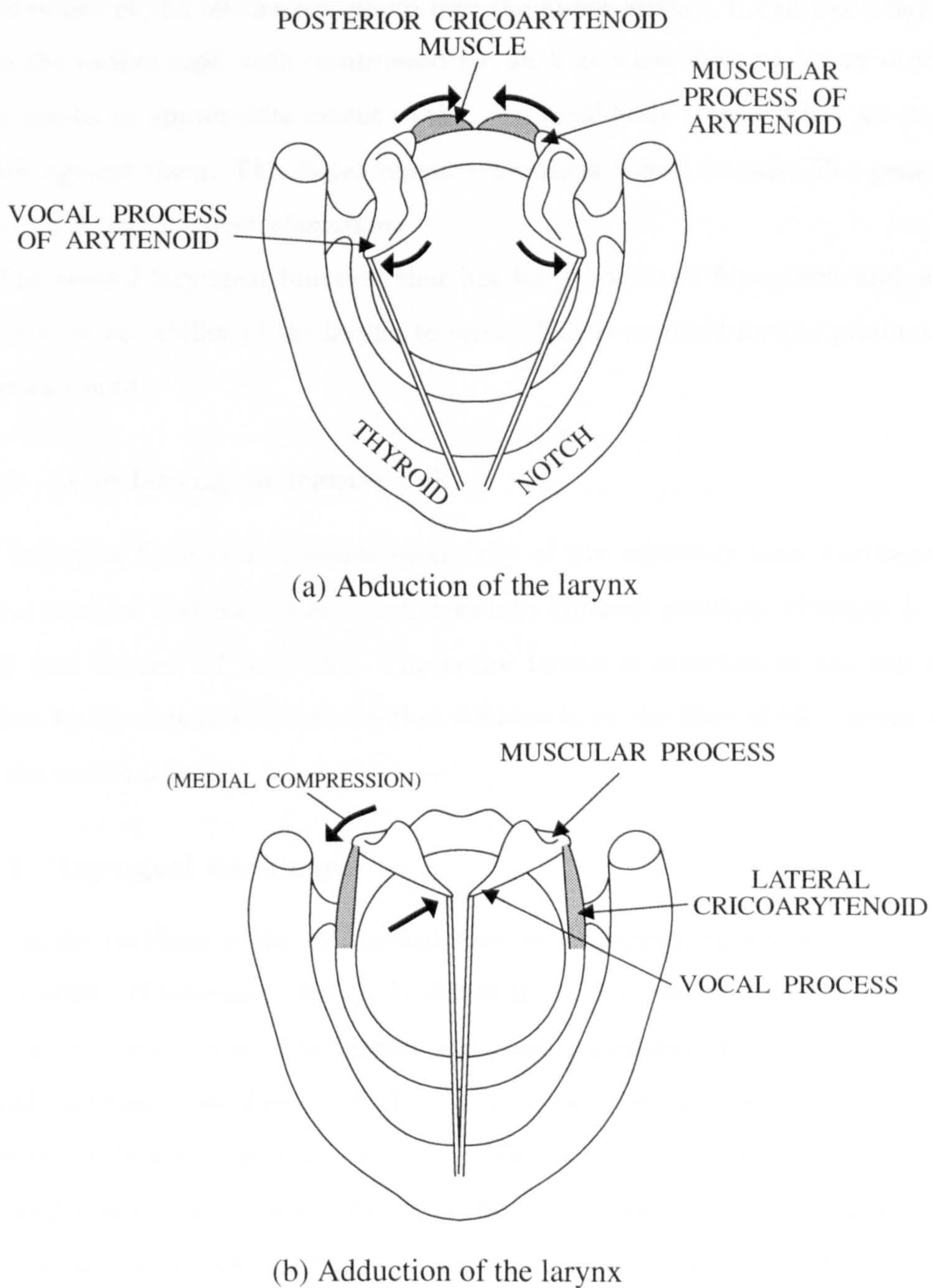


Figure 2.9: Abduction and adduction of the larynx (After Perkins & Kent, 1986)

entering the lungs. It also serves as a pressure device for such biological activities as coughing, sneezing and elimination. It is attached to the top of the trachea and is the outlet of the respiratory pump into the upper airway. It can close tightly to make the thorax rigid with compressed air, such as when lifting a heavy object. It can close to an appropriate extent so that the vocal folds vibrate when air pressure pushes against them. This latter function produces voiced sounds. The generation of voiced sound is called *phonation*.

The second laryngeal function that has been borrowed for speech and singing functions is the ability of the larynx to open. This is required for the production of voiceless sounds.

2.5.2 The laryngeal framework

The laryngeal framework consists essentially of five relatively large cartilages and several muscles that move these cartilages into different positions (Perkins & Kent, 1986) (see figures 2.7 and 2.8). The entire larynx is attached to the top of the trachea by ligaments and muscles that connect it to the base of the tongue above and the sternum below.

2.5.3 Laryngeal cartilages

The cricoid cartilage is the anatomical base of the larynx, to which the other cartilages attach (Lieberman, 1977). It differs from all other tracheal rings in that it forms a complete circle. The largest and most prominent of the cartilages is the thyroid cartilage. See figure 2.8. It is best known for its prominence, commonly called the 'Adam's apple' (thyroid notch). To the rear, it stretches up to a superior horn and down to an inferior horn. It is the lower horn that provides a pivotal attachment to the cricoid which permits the two cartilages to rock back and forth to each other. This is illustrated in figure 2.8. This ability is the basis for changing the length of the vocal folds in pitch adjustments. When these two cartilages are pulled towards each other, the vocal folds are lengthened; when tilted apart, the folds are shortened.

The arytenoids, a pair of small cartilages shaped roughly in the form of a pyramid, are mounted opposite each other on the rim of the signet portion of the cricoid

cartilage. They possess two forms of motion; rotation and gliding, which occur together. When they rock, their front sides tilt forward or away from each other. When they glide, their back sides slide toward or away from each other. With these two movements, the vocal folds may be abducted (opened) or adducted (closed) (Sawashima & Hirose, 1981; Ladefoged, 1973). This is illustrated in figures 2.9(a) and 2.9(b) respectively. Each arytenoid has two processes (prongs), one projecting forward, the other laterally. The anterior one is called the vocal process, to which the vocal folds attach. The lateral projection is the muscular process, to which several of the muscles responsible for rotation and gliding attach. The upward projection is the apex.

The epiglottis (shown in figure 2.7) is the least important of the main cartilages of the larynx. It provides the anterior wall of the tube leading from the vocal folds to the throat (Perkins & Kent, 1986).

2.5.4 An overview of the larynx muscles

The larynx has four basic anatomical components: a cartilaginous skeleton, intrinsic muscles, extrinsic muscles and a mucosal layer. This is illustrated in figure 2.7. The most important parts of the laryngeal skeleton with regard to the generation of voiced sound are the thyroid cartilage, the cricoid cartilage and the two arytenoid cartilages. Extrinsic muscles connect these cartilages to other throat structures; intrinsic muscles run between the cartilages themselves.

One pair of intrinsic muscles extends from the arytenoid cartilages to a point just below and behind the Adam's apple. These muscles are called the thyroarytenoid muscles. They form the body of the vocal folds. The space between them is the glottis (shown in figure 2.5).

The intrinsic muscles change the relative positions of the cartilages. These changes alter the shape, position and tension of the vocal folds. In particular, the cricothyroid muscle acts within the context of pitch control by increasing the longitudinal control of the vocal folds (Perkins & Kent, 1986).

The extrinsic muscles raise and lower the laryngeal skeleton. The resulting effect also changes the angles and distances between the cartilages and alters the resting lengths of the intrinsic muscles.

2.5.5 The larynx muscles

A brief description of the larynx muscles follows.

- **Abductor muscles**

The posterior cricoarytenoids (shown in figure 2.7) are flat muscles arising from the back wall of the cricoid cartilage. They insert into the muscular processes of the arytenoid cartilages. They are the major muscles responsible for rocking and gliding the arytenoids apart and for the abduction of the vocal folds (Sawashima & Hirose, 1981).

- **Adductor muscles** The lateral cricoarytenoid muscles arise from the lateral borders of the cricoid cartilage (Perkins & Kent, 1986). They insert into the muscular processes at the outside corners of each arytenoid cartilage. They pull in essentially the opposite direction from the posterior cricoarytenoids and consequently their effect is the opposite; they adduct the arytenoids, and are also able to squeeze the anterior tips of the vocal processes tightly together in a condition of medial compression.

The interarytenoids are formed by two muscles, the transverse arytenoid and the oblique arytenoids (shown in figure 2.7). The transverse arytenoid extends horizontally across the backs of the pair of cartilages. It pulls their middle edges together from bottom to top. The oblique arytenoids extend from the muscular processes of one arytenoid to the apex of the other. Their action is to pull the upper tips of the arytenoids together.

- **Vibrator muscles** The muscles that are set into vibration to produce sound are part of a larger pair of muscles, the thyroarytenoids (Titze et al., 1989). This consists of two parts; the internal thyroarytenoids (the vocal folds) and the external thyroarytenoids, which are lateral to the vocal folds. The thyroarytenoid can be considered as protruding into the upper airway to produce an arch structure when the vocal folds meet.
- **Vocal folds** The internal thyroarytenoids are the vibratory positions of the vocal folds. This paired muscle is normally referred to as the vocalis. Its location is shown in figure 2.5. The action of this muscle when it contracts is

to pull the vocal process, to which it attaches, into a straight line toward the thyroid notch (the point of origin). This action shortens the vocal folds and adducts them (assuming they were abducted). In addition, there is a stiffening of the vocal folds (Hollien, 1960).

- **Vocal fold lengthening muscles** The cricothyroid muscles (shown in figure 2.8) originate from the front of the cricoid cartilage and insert into the lower border of the thyroid cartilage. Their contraction is responsible for pulling the two cartilages together, thereby lengthening and stiffening the vocal folds.

2.5.6 Layers of the vocal folds

Five distinct layers of the vocal fold structure have been identified (Hirano, 1975). This is illustrated in figure 2.5. Beneath the thin, lubricated epithelium on the surface lie the superficial, intermediate and deep layers of tissue called the lamina propria. Underneath this is the thyroarytenoid muscle (vocalis muscle). Each of the five layers have their own physical properties that combine to produce the smooth shearing motions which are fundamental to vocal fold vibration. The vocal folds can be regarded as behaving physically like a three layer structure, consisting of a cover (the epithelium and superficial layer of the lamina propria), a transition layer (intermediate and deep layers of the lamina propria) and a body (the thyroarytenoid muscle) (Titze & Strong, 1975).

2.5.7 Laryngeal membranes and ligaments

The laryngeal cartilages are attached to joints and covered with membranes. In this section, the most prominent membranes in the process of speech generation are discussed. These are:

- **False vocal folds** These are situated above the true vocal folds. They consist mainly of thick folds of mucous membrane that protude into the airway, but not as far as the true vocal folds. The false folds originate just below the attachment of the epiglottis and insert into the lateral edges of the arytenoids below the apex. They do not normally vibrate through conscious control. Typically, their role is passive.

- **Laryngeal ventricle** Between the false and true vocal folds is a deep indentation in the mucous membrane wall of the larynx called a ventricle. This is the laryngeal ventricle, also called the ventricle of Morgani. This extends nearly the complete length of the vocal folds. It is bound laterally by the external thyroarytenoid muscle. Within the ventricle are several mucous glands. These provide lubrication of the true folds.
- **Conus Elasticus** This is a broad elastic membrane which covers the entire inner wall of the larynx. The lower portion of this membrane extends from the glottal edges to the cricoid cartilage. It is covered with mucosa.
- **Vocal ligament** This is an integral part of the mucosal cover of the vocalis muscle called the lamina propria. This consists of a superficial layer that provides a loose connection between the outer epithelial covering of the glottal edge and the intermediate and deep layers of the lamina propria that attach to the vocalis muscle. The intermediate layer contains elastic fibres. It blends into the deep layer, which contains collagenous fibres (thread-like strands). These two layers together form the vocal ligament. This construct permits the mucosal membrane to vibrate more or less independently of the vocalis muscle, which vibrates synchronously but less vigorously (Perkins & Kent, 1986).

2.5.8 Supraglottal cavity

The tube immediately above the larynx is called the laryngeal collar (supraglottal cavity) (Rossing, 1990). This cavity is formed anteriorly by the epiglottis, posteriorly by the arytenoids, and laterally by the aryepiglottic folds.

Embedded within these membranous folds are the aryepiglottic muscles (Perkins & Kent, 1986). These appear to be a continuation of the oblique arytenoid muscles, which, after inserting into the arytenoid cartilages, extend upward and forward to insert into the epiglottis. Contraction of the aryepiglottic muscle, coupled with the oblique arytenoids, tends to tilt the arytenoids against the epiglottis and to pull the epiglottis down. To push the epiglottis back, the thyroepiglottis contracts.

2.6 The supraglottal system

The supraglottal system (supralaryngeal vocal tract; see figure 2.3) consists of the various passages from the glottis to the lips (Gauffin & Sundberg, 1978). It involves the pharynx, mouth and nasal cavities, including the tongue, teeth, velum, and lips. The production of speech sounds through these organs is known as articulation (MacNeilage et al., 1979).

There are parts of the vocal tract that are immobile (passive articulators) and those that can move under the control of the speaker (active articulators).

2.6.1 The passive articulators

Passive articulators include the upper teeth, which are used in conjunction with the tongue to form a constriction for several sounds such as the first sound of 'thin', the ridge behind the upper teeth (called the alveolar ridge) against which many speech sounds such as [t] and [s] are made, and the bony arch behind the alveolar ridge, known as the hard palate, which is used in the articulation of a few sounds, such as the first sound of 'you'.

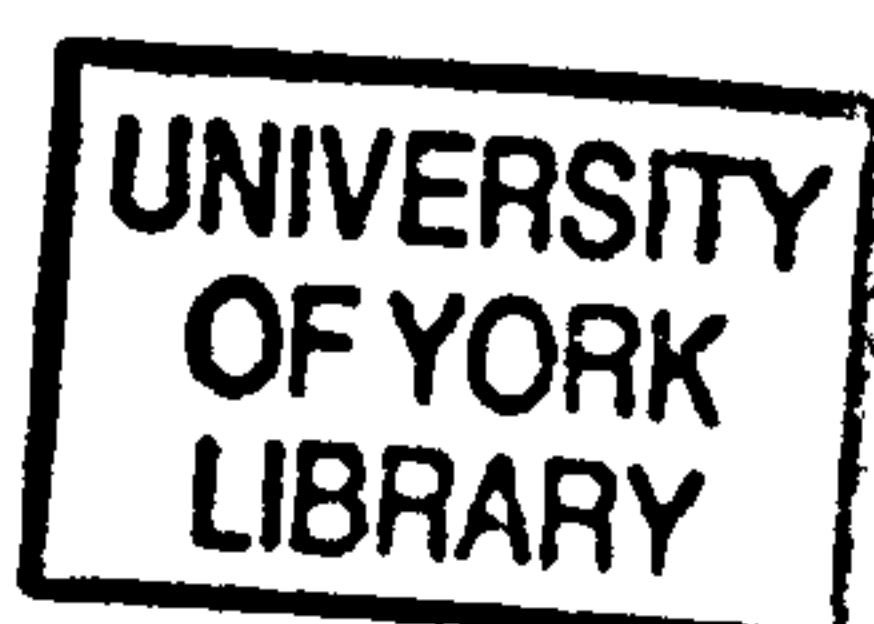
2.6.2 The active articulators

A brief description of the active articulators follows.

- **Pharynx** - This is a long muscular tube leading from the laryngeal cavity to the back part of the oral and nasal cavities. The areas adjacent to these cavities provide a means of dividing the pharynx into sections: the laryngopharynx, oropharynx, and nasopharynx (see figure 2.5). The pharynx can be narrowed or widened. Certain types of consonant can be produced by making a constriction here, and movements of the larynx, soft palate, and tongue may also involve pharyngeal modifications that affect the quality of the sound.
- **Soft palate (velum)** - This is a broad band of muscular tissue in the rear upper region of the mouth, whose most noticeable feature is the uvula, which is an appendage that hangs down at the back of the mouth. In normal breathing, the soft palate is lowered, to permit air to pass easily through the nose. In speech there are three main positions that affect the quality of sounds:

- i) The soft palate may be raised against the nasopharyngeal wall to make a 'velopharyngeal closure' so that air escapes only through the mouth. This produces a wide range of oral sounds, such as all the vowels and most of the consonants of English.
 - ii) The soft palate may be lowered to allow air to escape through mouth and nose. This is the position required to produce nasalized vowels, as used in, for example, the French word 'bon'.
 - iii) The soft palate may be lowered, but the mouth remains closed. In this case, all the air is released through the nose, as in nasal consonants such as [m] and [n].
- **Lips** - The orbicularis oris is the main muscle controlling lip movement, though several other facial muscles are also involved. The lips may be completely closed (i.e [p] and [m]) or held apart in varying degrees to produce the different kinds of rounding or spreading used on vowels or the friction of certain kinds of consonant.
 - **Jaw** - The mandible bone permits a large degree of movement. It controls the size of the gap between the teeth and strongly influences the position of the lips.
 - **The tongue** - The tongue is the most versatile of the articulatory organs. It can move in any of three main directions through the action of the various 'extrinsic' muscles: upwards/forwards (such as for [i]), upwards/backwards (such as for [u]), and downwards/backwards (such as for [a]). In addition, several 'intrinsic' muscles determine the shape of the tongue, in any position. For example, some muscles raise or lower the tongue tip, or move it to the left or to the right.

Two examples of the articulatory state required for pronouncing vowels are shown in figure 2.10. For the vowel [i], the tongue is arched forward and upwards (Fant, 1960). For the vowel [u], it is arched backwards, and the lips are relatively closer. The acoustic role of the articulators and voice source are discussed in the following sections.



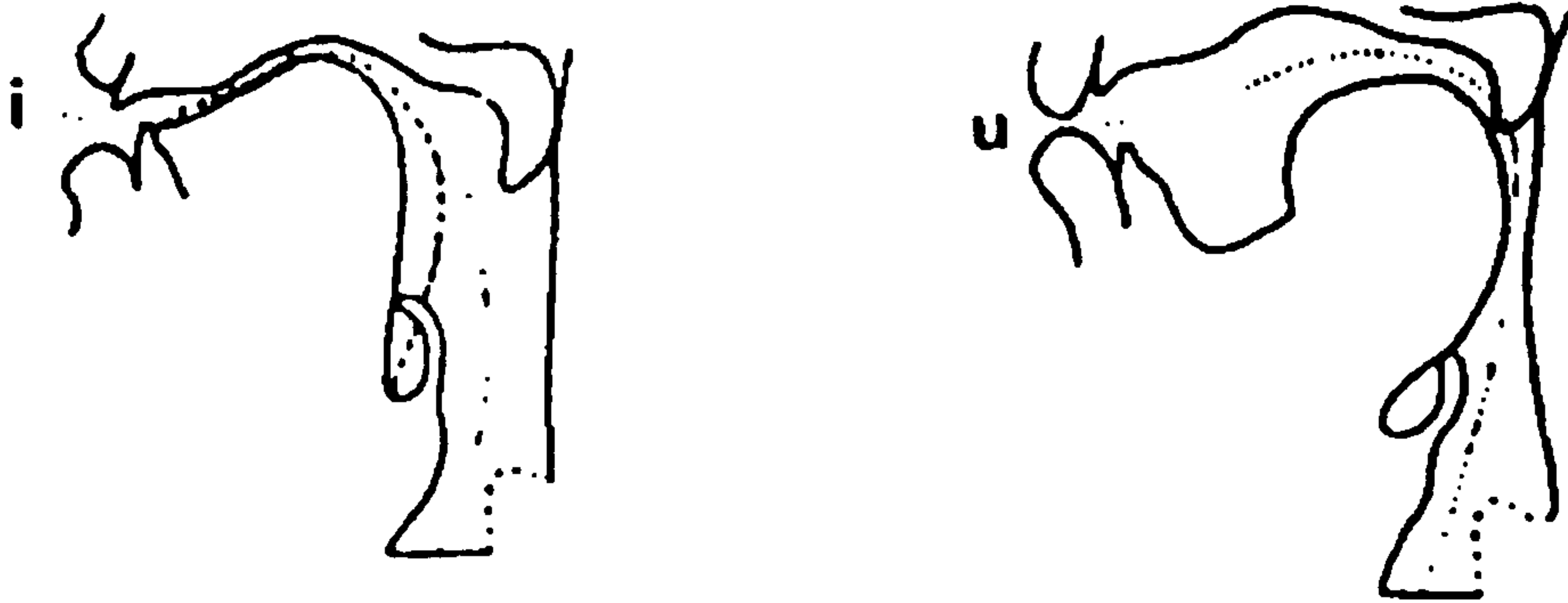


Figure 2.10: Vocal tract profiles for two vowels. The shape is determined by the lip and jaw openings, the tongue shape, the velum, and the larynx. (After Fant, 1960; in Sundberg, 1987)

2.7 Acoustics of the voice

2.7.1 Glottal action

To recall, the action of the glottis to produce a sound is called *phonation* (Lieberman, 1977). Phonation depends on how the airflow interacts with the muscular and elastic tensions of the vocal folds.

During breathing, the arytenoid cartilages are held outward, keeping the glottis open in a wide-open position. When phonation is about to begin, the arytenoids move inward and bring the vocal folds to a position of adduction (figure 2.9). The chest and lungs press inward and the diaphragm contracts, causing the air pressure below the glottis to rise. When this subglottal pressure is sufficiently higher than the air pressure above the glottis, the closing tension on the folds is overcome and they begin to open in a glottal slit between the folds. As the glottal slit opens, air begins to flow out through the glottis. The subglottal pressure continues to force the glottis to open wider, and there is an increasing airflow through the glottis, until the natural elastic tension of the folds, which increases dynamically with increased opening, balances the separating force of the air pressure. Then the glottal opening and the rate of airflow through the glottis have reached their maximum value. At this point the kinetic energy that the vocal folds received during the opening move-

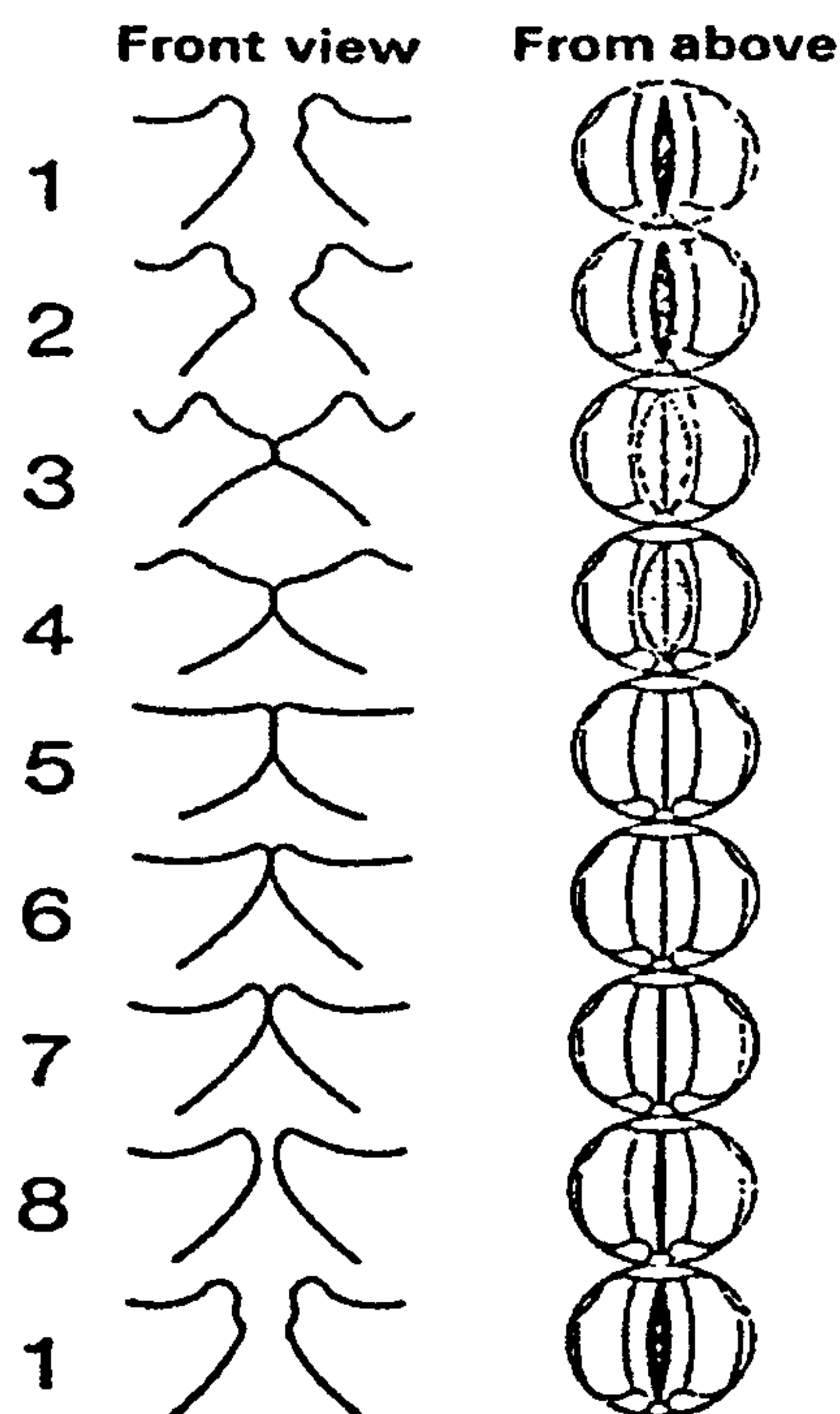


Figure 2.11: The pattern of vocal fold movement (From Sundberg, 1987)

ment is stored as elastic recoil energy, resulting in a restoring force large enough to overbalance the separating force of the airflow. This stored energy also causes the folds to move inward. The inward movement gathers momentum and continues. When the glottis becomes sufficiently narrow, the high velocity of movement of the air particles within the narrow glottis creates a suction effect, which tends to accelerate the vocal folds towards each other. This phenomenon is known as the Bernoulli effect (Ladefoged, 1973; Laver & Hanson, 1981). Both the elastic restoring force and the Bernoulli force act to close the glottis. Elastic restoring forces during this collision of the vocal folds, in conjunction with the sub-glottal pressure, then start a new cycle of action (Stevens, 1988). The acoustic main excitation is at the instant of closure of the vocal folds (Sundberg, 1987). The pattern of vocal fold movement is shown in figure 2.11. The rolling motion along the medial parts of the vocal folds is referred to as the mucosal wave (Sundberg, 1987).

As long as subglottal pressure remains at a sufficiently high level and the arytenoid cartilages hold the vocal folds together, voicing phonation occurs and the

glottis will continue to emit a rapid series of air pulses. This produces the *glottal waveform*.

2.7.2 The spectrum of the glottal sound source

The pulse train is the basic sound source for vowels and all other voiced sounds. The spectrum of this sound source depends on just how the glottis forms the pulses. The exact form of the airflow through the glottis has an effect on the glottal sound spectrum (Fry, 1979).

The sound produced by the larynx is a complex tone which contains the fundamental frequency of the sound together with many partial components (Perkins & Kent, 1986).

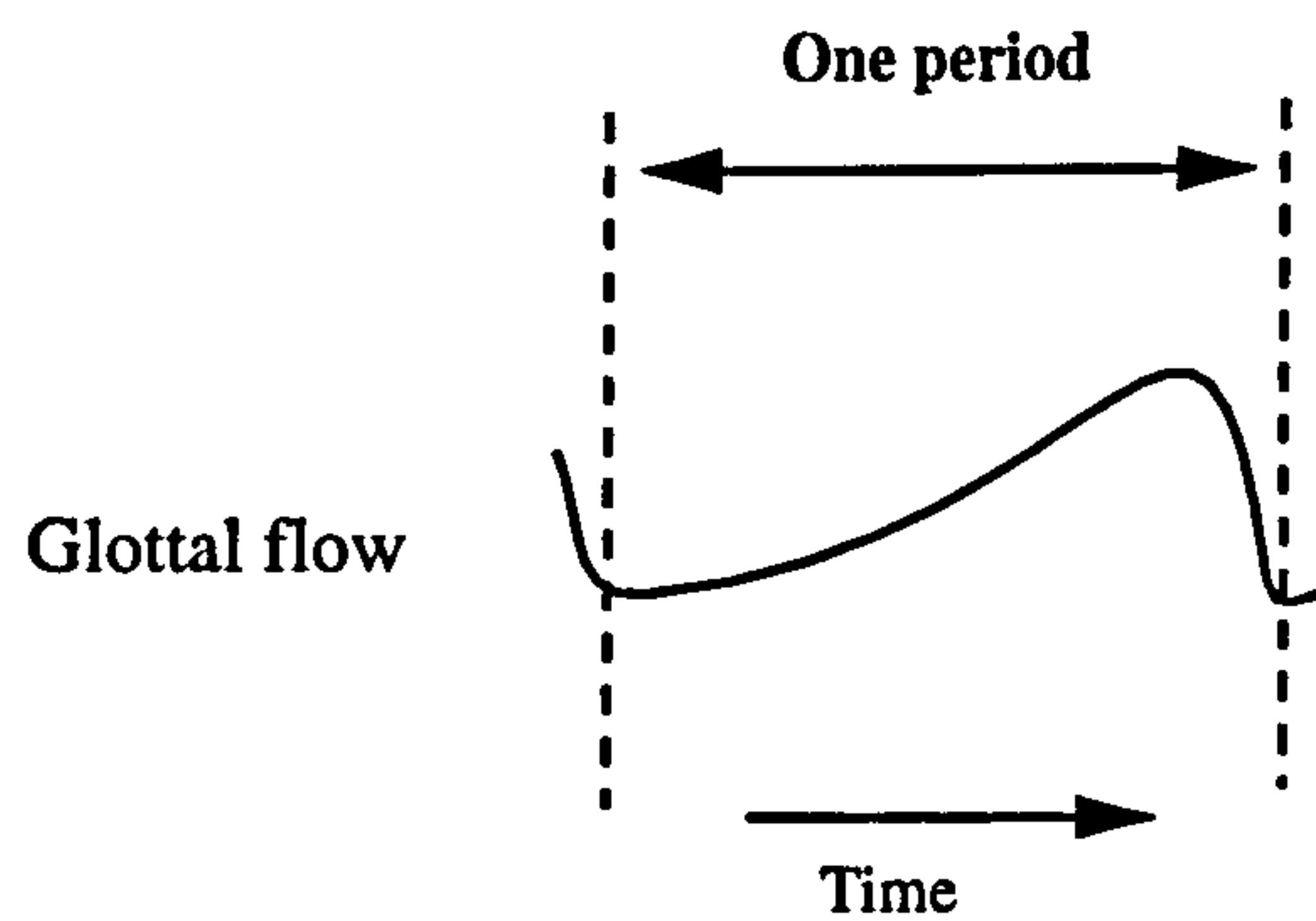
There are two primary characteristics of the glottal spectrum (Berke & Geratt, 1993). These are;

1. the frequency spacing of the fundamental and harmonics, and,
2. the amplitude pattern of the components over frequency.

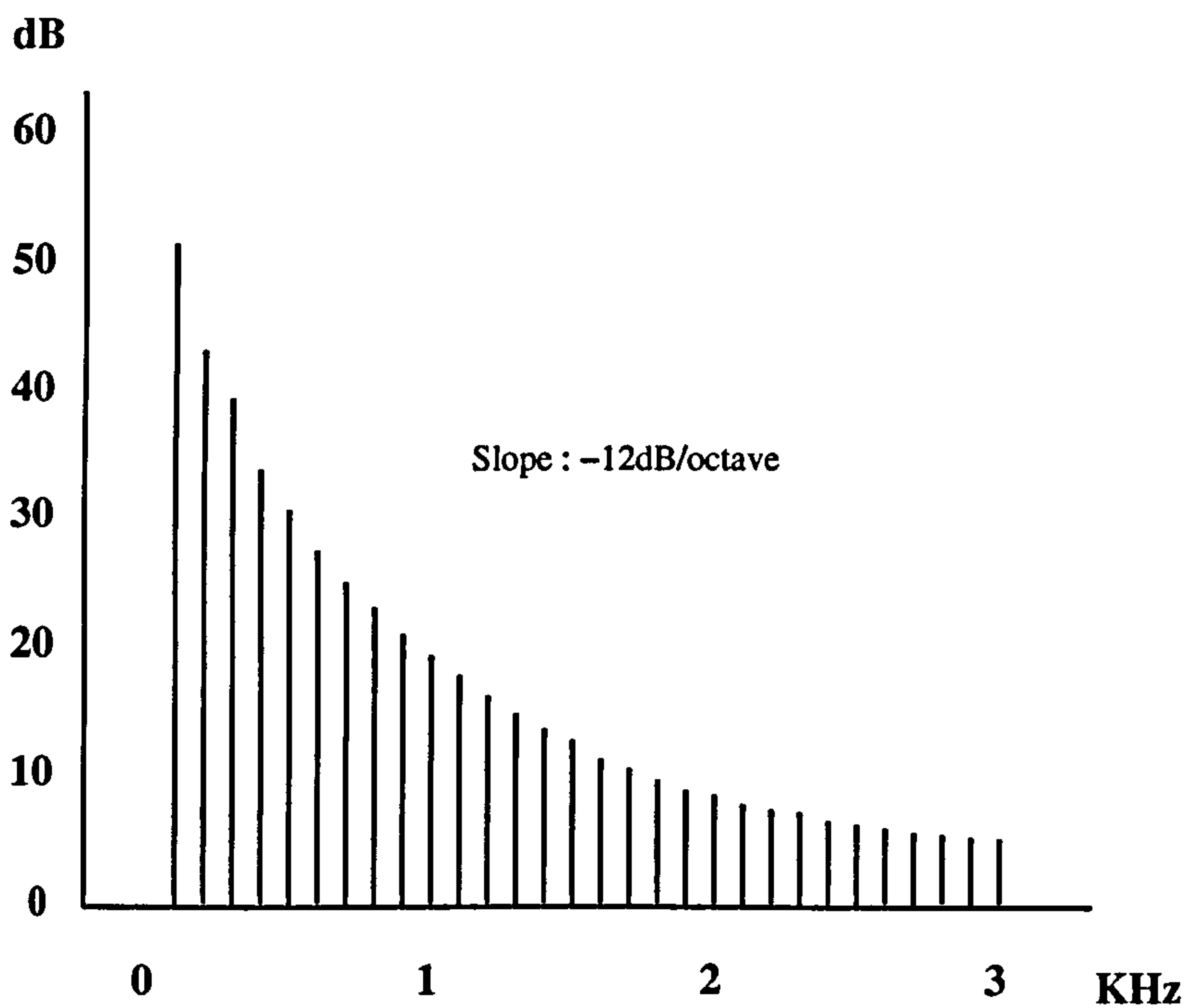
The glottal spectrum is affected by several glottal parameters, including the duration of the open portion and the ratio of the open portion to the closed portion (Laver & Hanson, 1981). These parameters are discussed in detail in the next chapter.

The frequency spacing depends on the repetition rate of the pulses in the glottal wave. The amplitude and phase characteristics of the spectral components depend on the exact shape of the pulses. The spectrum of the glottal pulse reveals the spectral amplitude distribution as a result of the glottal pulses. Figure 2.12 shows an idealised example of a glottal wave and corresponding spectrum. The fundamental frequency of the spectrum shown is 100Hz. Each harmonic component occurs at integer multiples of this frequency.

The amplitude of the components in the glottal spectrum have a pattern that generally decreases from low frequency to higher frequency harmonics at an average of 12dB per octave. Component amplitudes of the glottal spectrum relate to the degree of rounding of the glottal waveform. Examples illustrating the basis for this relationship are shown in figure 2.13. A pure sine wave has only one spectral



(a) The variation in airflow through the glottis across one period.



(b) Spectrum of the glottal flow wave shown in (a) above

Figure 2.12: Idealised glottal area, flow and spectrum (After Pickett, 1991)

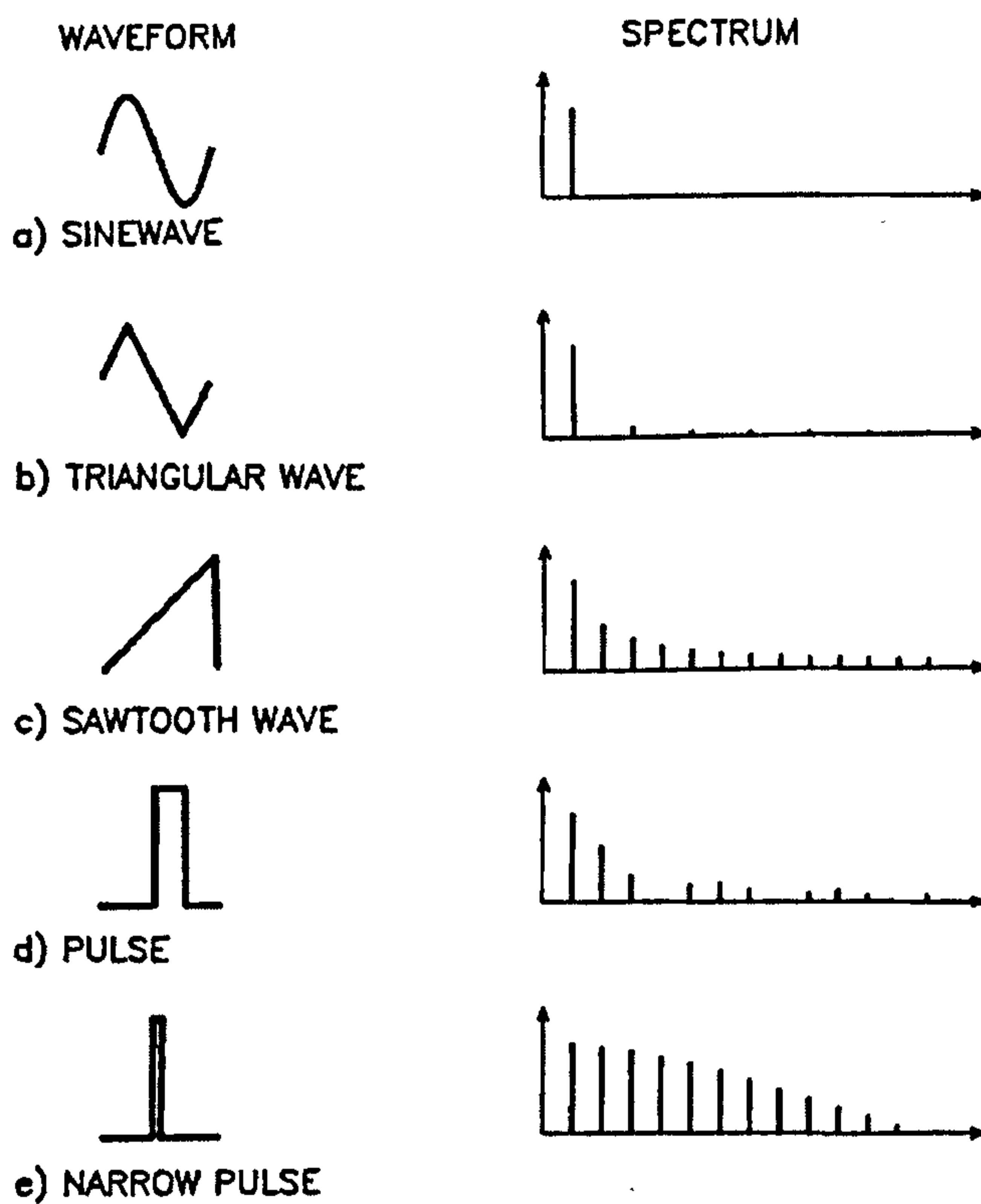


Figure 2.13: Spectra of simple waveforms (From Dodge & Jerse, 1985)

component (figure 2.13(a)). Any non-sinusoidal period signal (such as all those produced by the action of the larynx) introduce other spectral harmonics at higher frequencies. This is illustrated by figures 2.13(b)-2.13(e). The number of spectral components, and in general their amplitude, is proportional to the rate of opening and closing of the glottal wave, and to the relative length of open and closing times. An illustration of a greater degree of higher partials resulting from a sharper rate of opening and closing can be seen by comparing the waveform shown in figure 2.13(a) with that shown in figure 2.13(b). An illustration of the effect of differences in the relative length of the open and close times can be seen by comparing the waveform in figure 2.13(d) with that shown in figure 2.13(e). Although both have the same rate of opening and closing, there is a greater degree of high frequency partials in the spectrum of the latter waveform than of the former, due to the relatively narrower pulse.

These parameters relate physiologically to the style and force of speaking and to individual characteristics of vocal fold behaviour (Kent, 1993). Several studies have indicated the validity of applying these signal processing techniques for models of the voice. For example, Cheng and Guerin (1985) conclude that the open quotient plays "a comparatively important role" and utilise the notion of the voice as a source-filter model (to be discussed in the following sections) to support the theory that the energy in the high frequencies of the source spectrum increase with CQ.

The 12dB slope used for glottal spectrum roll-off is a fair approximation for average, relaxed phonation (Fant, 1960). Phonation of a greater degree of relaxation results in glottal pulses that are relatively more rounded on the corners, causing the glottal spectrum to slope downward more steeply (typically -15dB/octave). For forceful speaking, the glottal waveform can have a relatively sharper rising and falling edge because of higher subglottal pressure on the opening of the vocal folds and a higher level of Bernoulli suction before the closing. An example roll-off in this mode of phonation would be -9dB/octave (Fant, 1970).

The fundamental frequency of the glottal pulsing is perceived as pitch. The former depends on the tension on the vocal folds, the effective mass of vocal folds, and on the subglottal pressure (Laver & Hanson, 1981). The effective mass depends on the size of the vocal folds, which are themselves dependent on sex, age and

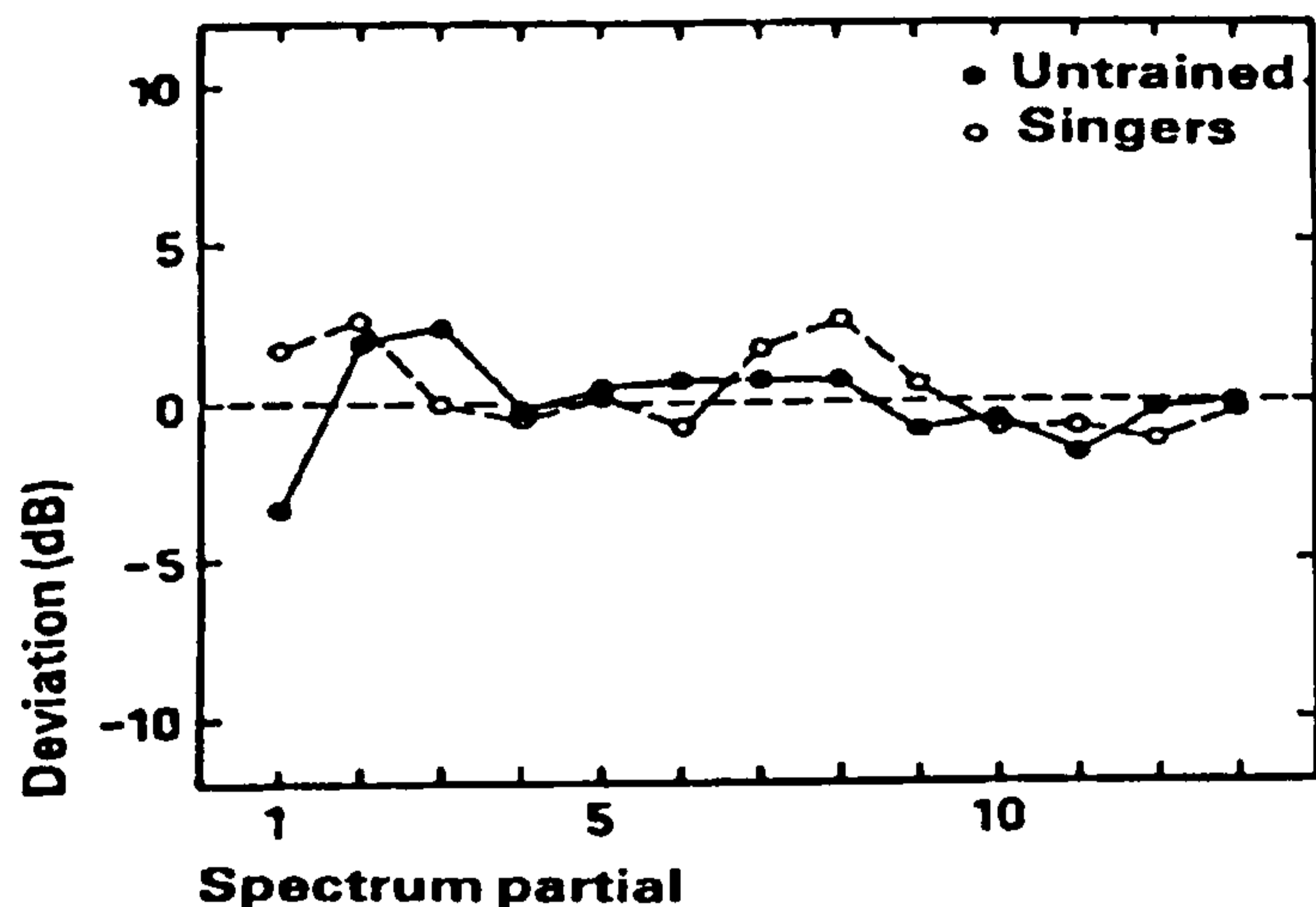


Figure 2.14: Average source spectra from untrained voices (filled circles) and professional singers (open circles). The spectra are represented by a curve showing how the spectrum contour deviates from the standard slope of -12dB/octave . (From Sundberg, 1987)

individual characteristics.

The average voice source spectrum characteristic has been shown to exhibit differences between adult untrained speakers and professional singers. Deviations from the standard glottal spectrum slope of -12dB/octave are shown plotted in figure 2.14. This figure uses results from two different investigations which used the same method of analysis (Carr and Trill, 1964; Cleveland, 1976; discussed in Sundberg, 1987). The curves shown in the figure represent the average obtained from the vowels /u:, i:, a:/. The amplitude of the fundamental (labelled in the figure as partial 1) is clearly stronger for professional singers. In addition, the amplitude of partials 7 to 9 are clearly greater for the singers. These voice source contrasts may be appropriate for consideration as part of an assessment criteria for the voice in a system for voice development. These, and other attributes, are discussed in depth in the next chapter.

2.7.3 Supralaryngeal acoustics

The sound produced by the glottis is propagated through the supralaryngeal

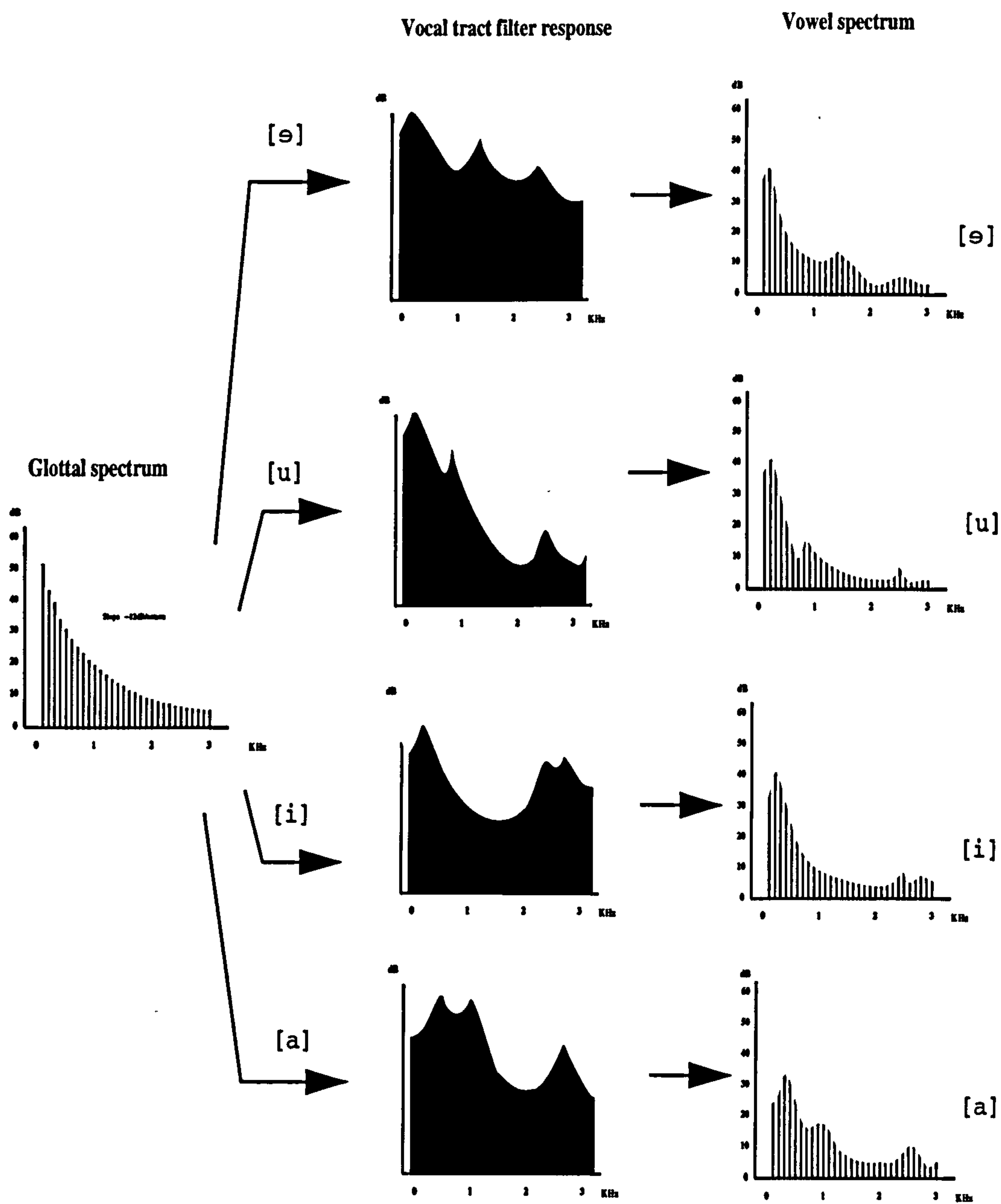


Figure 2.15: The production of model vowels according to the source-filter theory. An idealized spectrum slope of -12dB/octave is assumed for the glottal airflow spectrum. It is modified by the filter responses of the vocal tract, shown for four different vowels, to produce the final vowel spectrum on the right. (After Pickett, 1991)

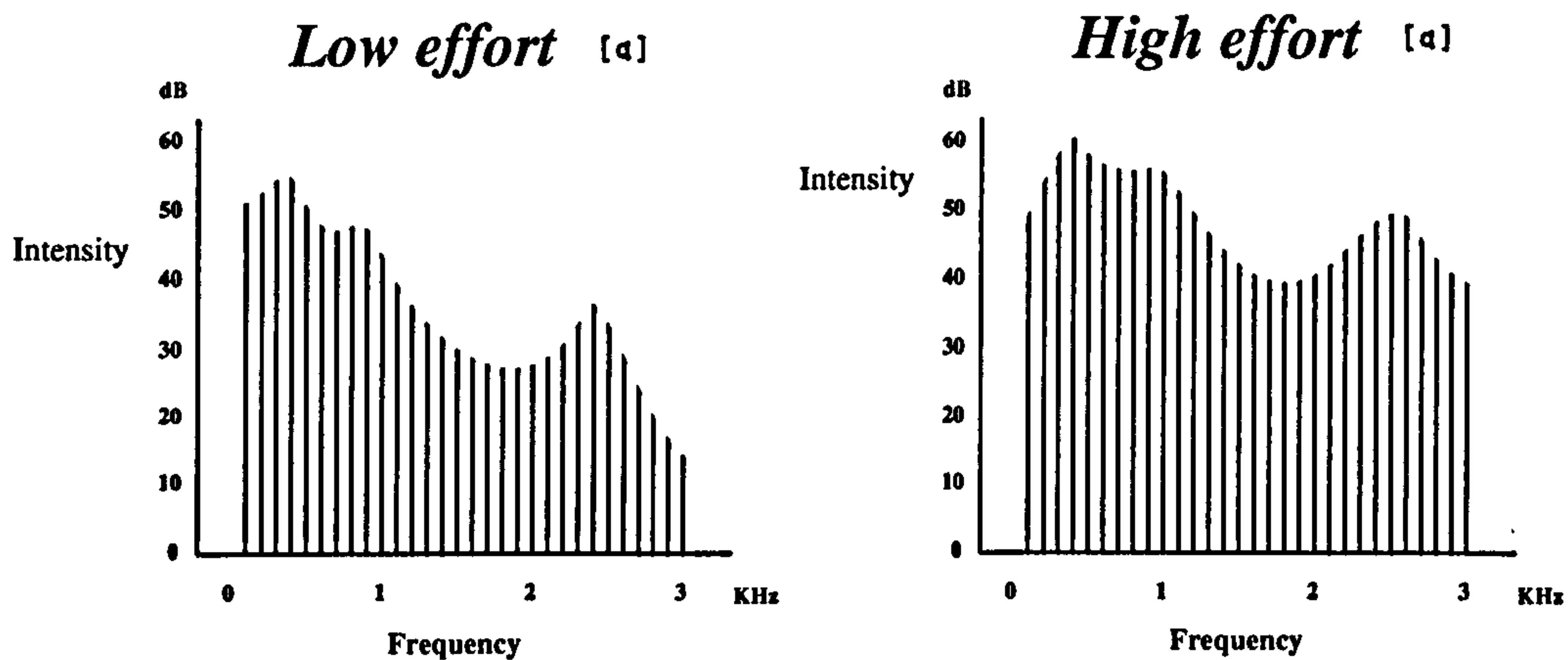


Figure 2.16: The effect of vocal effort on the vowel spectrum. The overall slope of the vowel spectrum is steep for low effort and more gradual for high effort. Second and third formants are relatively more intense under high vocal effort. (After Pickett, 1991)

vocal tract to the outside air. The tract can be considered to act as a filter which emphasises spectral areas of the source sound at frequencies at and near the resonant frequencies of the tract (Fant, 1983). The spectrum of the vocal output can be altered by changing the activity of the articulatory organs within the vocal tract, so that the resonant characteristic of the signal changes (Hillenbrand & Gayvert, 1983). This approach is called the *source-filter* theory of vowel production (Fant, 1960). Examples of the different position of the vocal tract articulators are shown in figure 2.10. The resonant peaks of the vocal tract determine the filtering curve of the tract. When this filter curve is applied to the spectrum of the glottal sound source, the vowel sound spectrum is produced. This is illustrated in figure 2.15, in which diagrams show how the spectrum of the glottal sound source is modified according to filter curves of the oral-pharyngeal tract to form the vowel sound spectrum (Cleveland & Sundberg, 1983).

Formants are spectral areas in which there is a particular embellishment. It has been shown that four or five formants are relevant for the majority of voiced sounds (Sundberg, 1987). Any alteration in the cross-sectional area of the vocal tract results in shifts of formant frequency and strength (Johansson et al., 1983). Vowels can be seen to have distinctive patterns of formant structure, as figure 2.15

illustrates.

An illustration of the effect of changes in the glottal spectrum in combination with supralaryngeal vocal tract qualities is shown in figure 2.16. For low vocal effort the spectrum is relatively steeper in slope (approximately -15dB/octave) and the level is relatively low in amplitude. However, for higher effort the general amplitude of the partials has increased, reflecting a higher level of acoustic output. Although the spectrum peaks are at the same frequency location the spectral slope has changed considerably to one of approximately -9dB/octave (Pickett, 1991). The spectral slope is shallower with higher amplitudes at the high frequencies and higher spectral amplitudes than those shown for less effort, but the fundamental component is lower in amplitude. These increases in the amplitude of resonant oscillations are due solely to the increase in spectral energy from a shift in glottal action.

2.8 Conclusions

This chapter has considered the process of phonation. The operation of the subglottal system, larynx, and articulators have been discussed in depth.

With regard to monitoring the human voice, it has been indicated that two key loci are appropriate for the monitoring of any act of phonation. These are:

- **the larynx**, at which the glottal pulse is applied, and
- **the acoustic output**, at which articulatory effects may be assessed.

The next chapter assesses parameters expressed at these two loci which are likely to be appropriate for monitoring as part of a analysis and visual feedback system for voice development able to operate in real-time.

Chapter 3

Parameters for real-time assessment

3.1 Introduction

In the last chapter a comprehensive overview of the process of phonation was provided. Two key loci for monitoring vocal production were identified as

1. the voice source (larynx), and
2. the acoustic output of the subject.

In this chapter the parameters expressed at these two loci which relate to vocal development are assessed for suitability for inclusion within a system for vocal assessment and tuition able to operate in real-time. Previous research concerning the application of these parameters for real-time display is considered.

3.2 Voice parameters

From an assessment of the scientific literature, the following parameters have been identified as major indicators of vocal usage.

From the acoustic output:

- Loudness
- Qualities of the acoustic spectrum

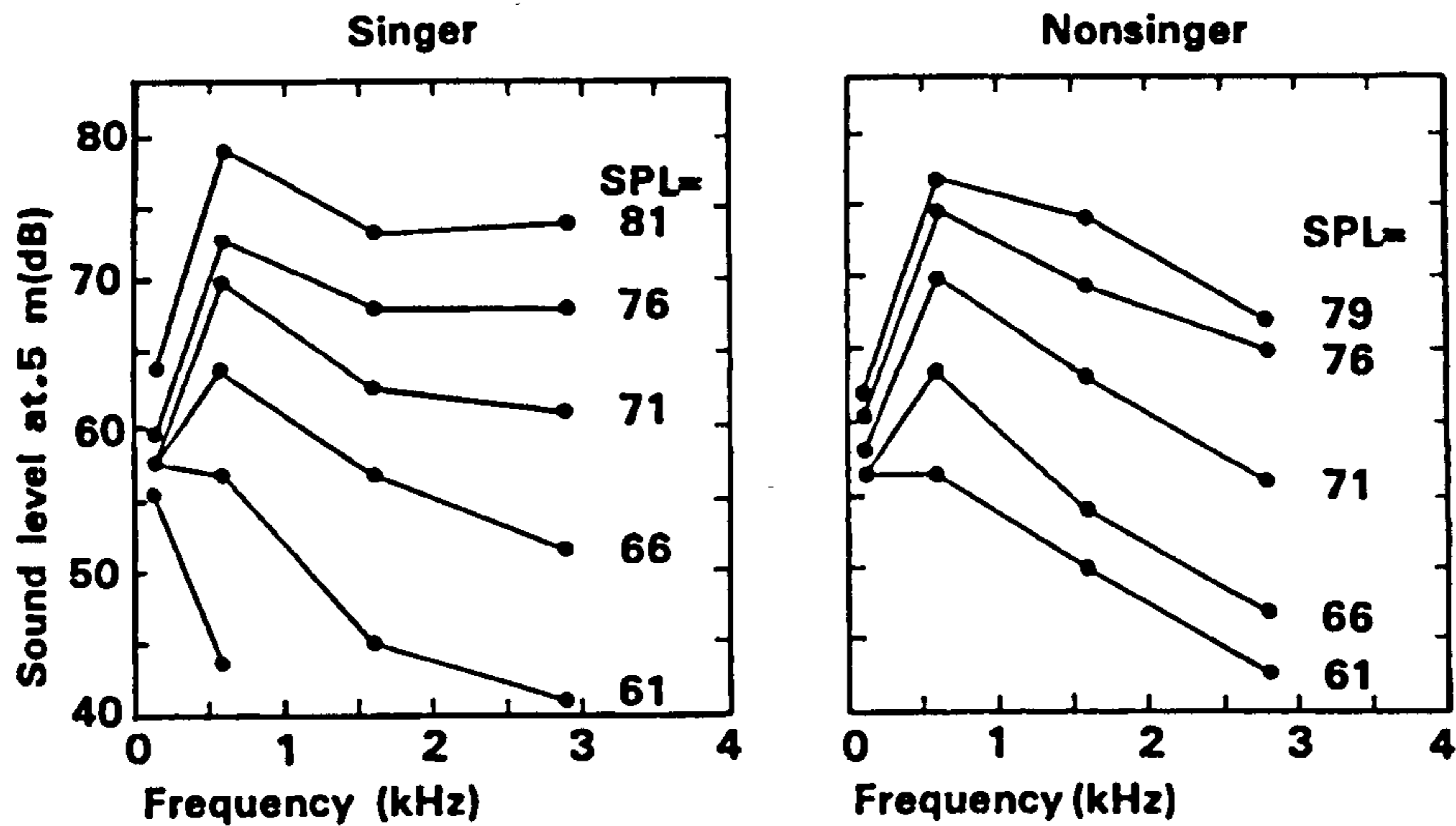


Figure 3.1: Changes in spectrum level in four equally wide frequency bands associated with changes in SPL as observed in a singer (left) and non-singer (right) phonating the vowel /ae:/ at deliberately varied vocal loudness (From Gauffin & Sundberg, 1989)

From the voice source (larynx):

- Fundamental frequency
- Closed quotient of the laryngeal period

In the following sections these parameters and their potential for use as part of a real-time system for voice development will be discussed in detail.

3.3 Loudness

3.3.1 Introduction

Differences have been observed in patterns of loudness (usually measured in terms of the sound pressure level (SPL)) between trained and untrained voice users. For example, Gauffin & Sundberg (1989) observed voice source variations in a singer and a nonsinger under different levels of loudness. Figure 3.1 illustrates the sound level of the fundamental and all formants in the frequency range 2-4KHz of a singer and a non-singer who both phonated the vowel /ae:/ at different levels of loudness,

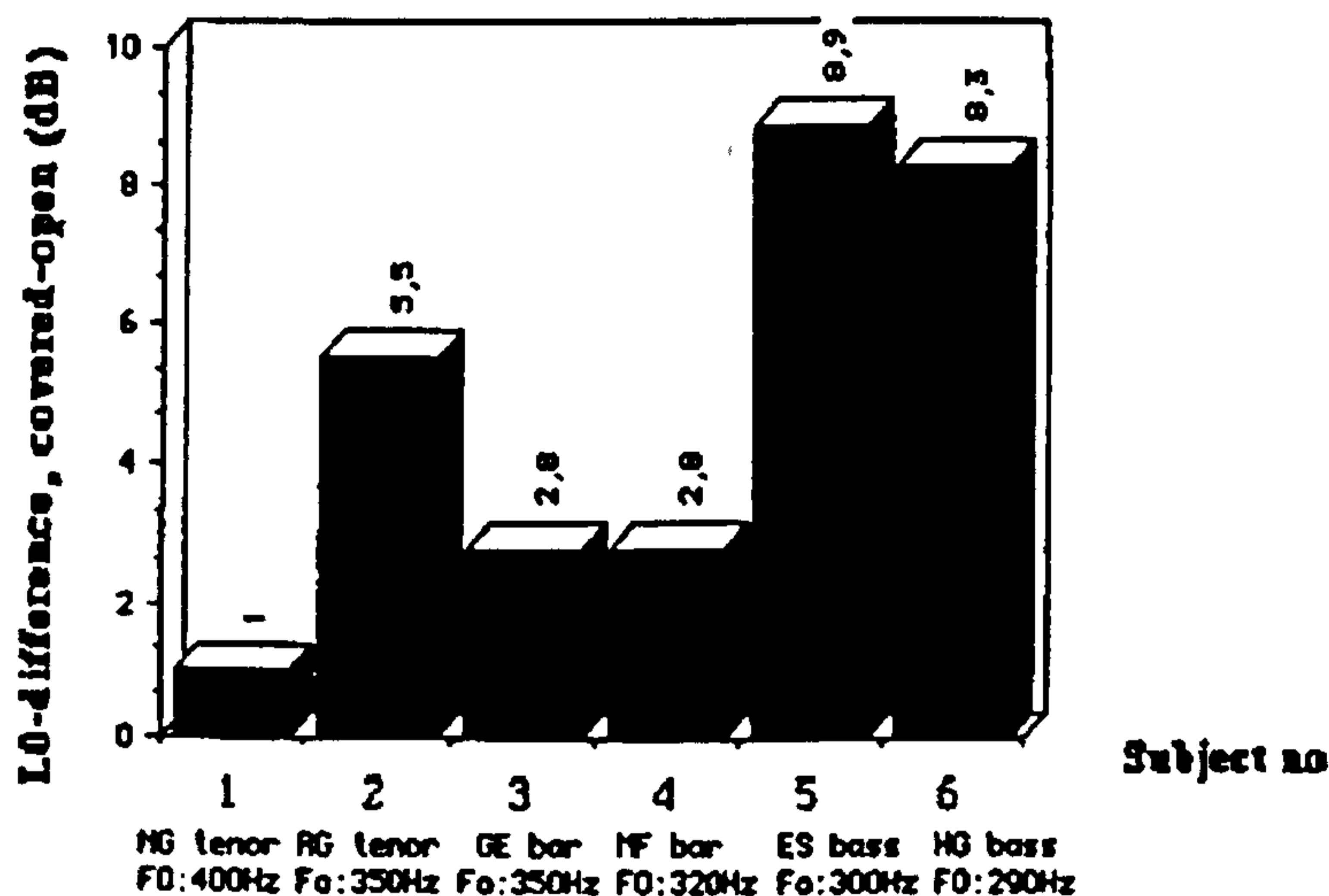


Figure 3.2: Differences between SPL for covered and open singing of the lowest spectrum partial. The mean values of the difference are shown for six subjects. Fundamental frequency (F0) for each subject is also displayed. (From Hertegard et al., 1990)

but at constant fundamental frequency. The graph indicates that the singer is able to maintain the generally even distribution of formants regardless of SPL. The non-singer is less able to maintain this consistency. Gramming et al. (1988) noted that when reading in noise singers raised their mean sound level relatively less than non-singers, implying an ability to convey their voice efficiently through means other than the raising their SPL alone.

Differences in SPL output between certain modes of voice usage have also been identified. For example, figure 3.2 shows the relative difference in SPL between the fundamental of the note when singing in a 'covered' style¹ to give the impression of just one single register over the entire range compared to singing in the contrasting 'open' style² (Hertegard et al., 1990). The level is consistently higher for covered than for open singing.

¹defined by the author of the study as characterised by a raised jaw, no lip protrusion, a small lip opening, retracted tongue tip, raised tongue dorsum, and the subjective sensation of a squeezed or constricted throat.

²defined by the author of the study as characterised by a low jaw and tongue position and a pitch-dependant size of mouth opening.

The difference in the level of SPL from one instant to the next is known as *shimmer*, and it is often used as a measure of voice perturbation (i.e., Ludlow et al., 1987; Orlikoff & Kahane, 1991; Slavit & Maragos, 1994). Gelfer et al. (1991) indicated that trained singers exhibit a more consistent level of shimmer than untrained singers. A real-time display of the measure may be of benefit in encouraging vocal stability.

3.3.2 The use of loudness in a real-time display

Methods supported by computer instruction would be likely to help support the development of acoustic output level control in many different contexts of voice usage. However, the application of a real-time display of this parameter does not appear to be widespread. This may be because it is relatively easy to control, which reduces the usefulness of a real-time display of the parameter.

In one example, a real-time display of the level of acoustic output was used in the treatment of an adult male subject (Cott, 1994). This subject was a professional academic who was frequently required to use his voice for talks and lectures. As soon as his acoustic output reached a particular volume level, his voice would effectively cease functioning. A portable device was created which monitored his SPL output and presented the information visually via a series of LED displays which were worn on his wrist. The subject could then inconspicuously look at the display during the process of his presentations and was able to consciously control the output level of his voice to ensure his voice usage remained within the SPL range known not to cause problems.

3.4 Acoustic signal changes

3.4.1 The 'singer's formant'

Sundberg (1974), amongst others, noted that a spectral envelope peak is resident in the region of 3KHz for voiced sounds produced by male opera and concert singers in Western music culture. A simple example illustrating differences in spectra between trained and untrained singers is shown in figure 3.3. He originally attributed it to a clustering of the third, fourth and fifth formants, resulting in increased impact

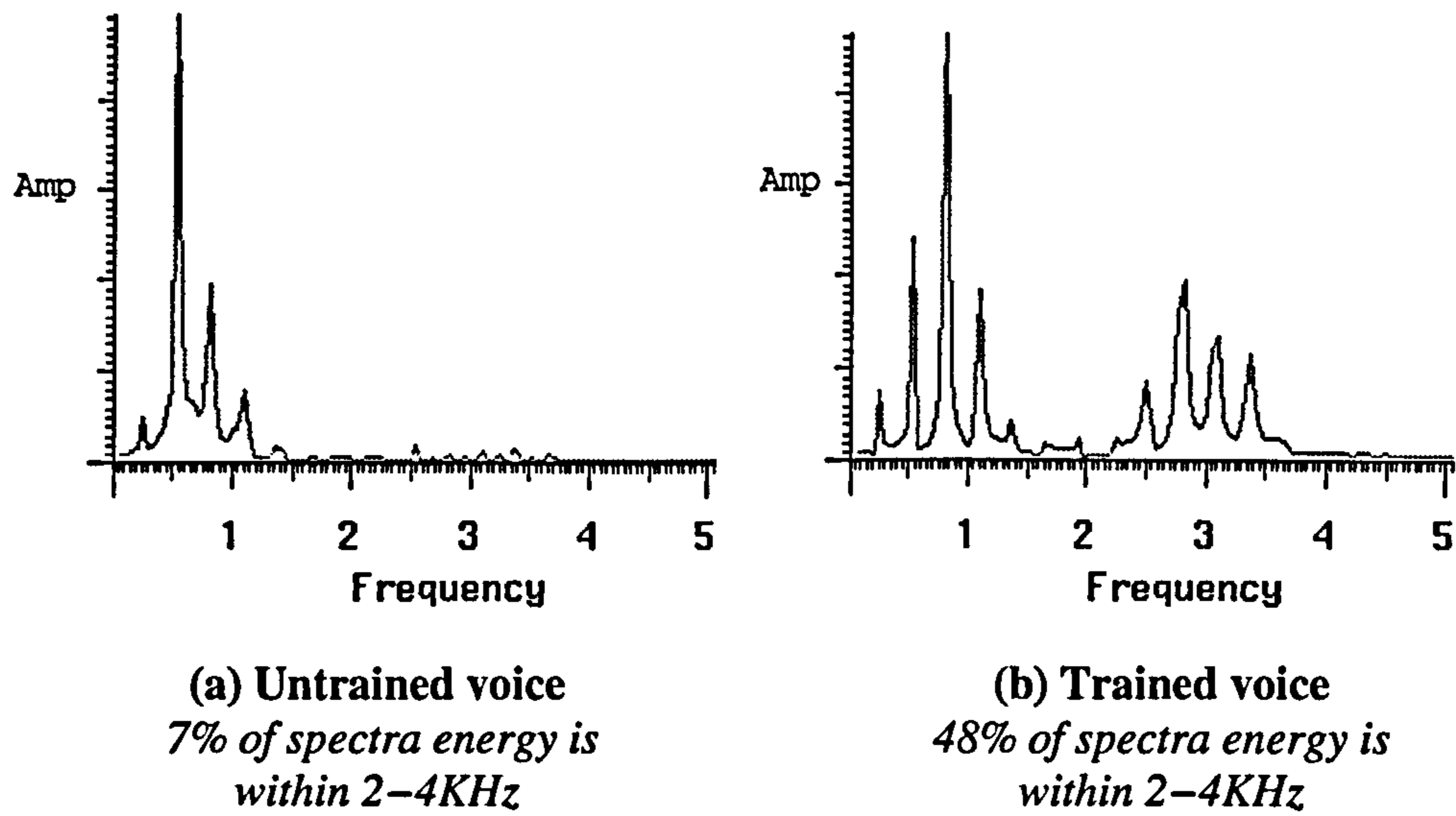


Figure 3.3: Acoustic output spectra of trained and untrained (male) subjects singing [a]

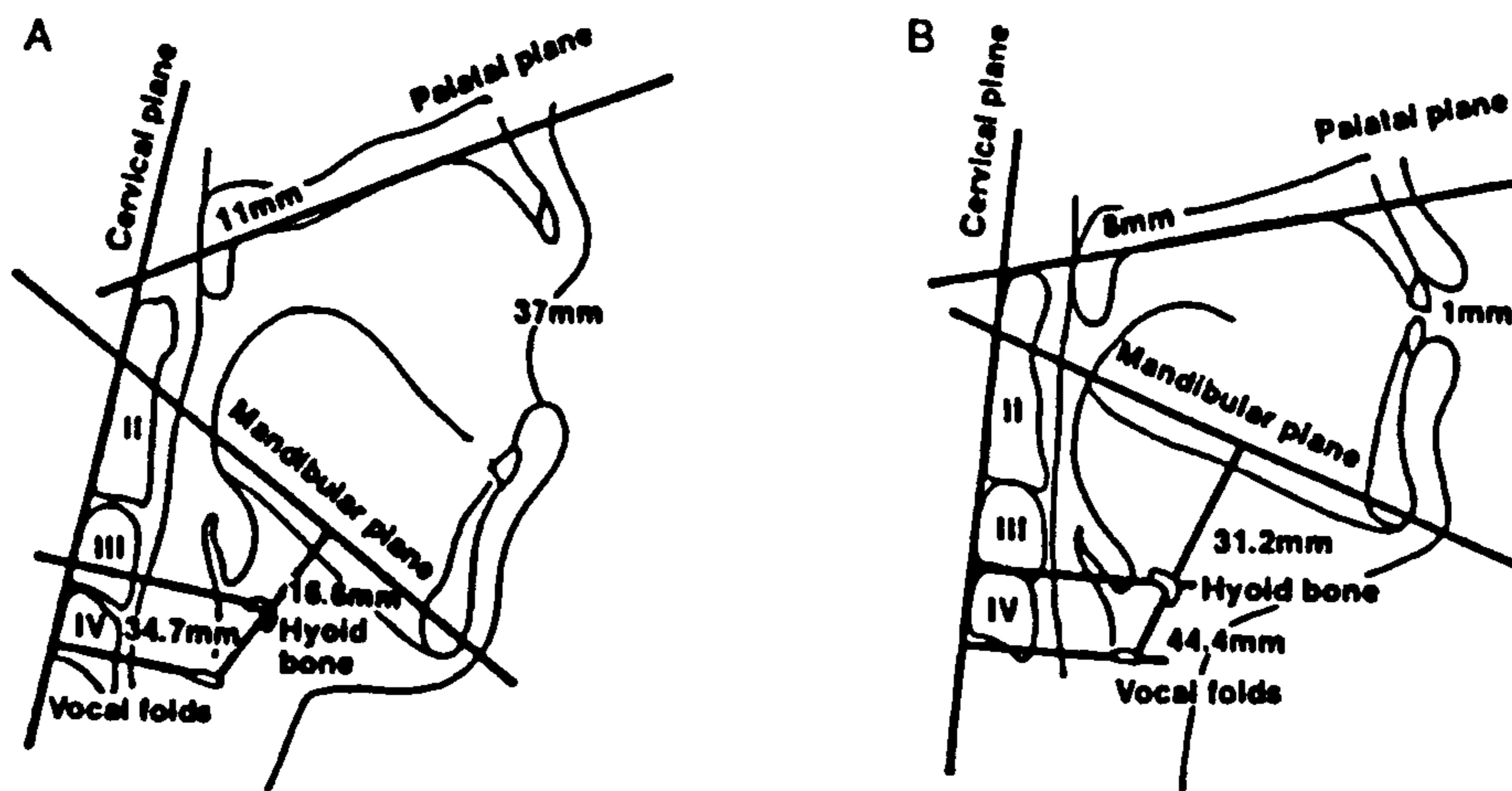


Figure 3.4: Voice physiology. (a) Lateral view of ring phonation (b) Lateral view of constricted phonation (From Acker, 1987)

of the frequency spectrum in these areas. This peak is generally called the singer's formant (Sundberg, 1987). Other names also exist for it, including "head resonance" or "ring". Several other studies have observed this phenomenon (i.e. Seidner et al., 1983; Walker, 1992) including an analysis of CD recordings of professional tenors (Lindsey and Howard, 1989). This resonance has also been recorded using vibration sensors in various locations around the body, with a spectral resonance observed for trained singers, but absent for untrained singers (Sakakura and Takashi, 1988).

The centre frequency of the singer's formant is dependent on the type of voice, pitch, and vowel (Sundberg, 1987). For example, Seidner et al. (1983) report that the center frequency of the singer's formant varies between 2.3kHz and 3kHz for basses, and 3kHz and 3.8kHz for tenors.

3.4.2 Physiological correlates

Sundberg (1974) originally postulated that the singer's formant was due to a lengthening of the pharynx and a widening of the cross-sectional area in the pharynx at the level of the larynx tube opening to more than six times the area of the opening. A widening of the pharynx base appears to have been verified with the use of acoustic pulse reflectometry techniques³ (Clark, 1992). Others (i.e. Yanagisawa et al., 1989) have noted a strong relationship between constriction of the aryepiglottic sphincter and the production of the singer's formant.

Acker (1987) studied vocal tract adjustments for the projected voice. He considered the perceptual, acoustic and physiological features of the spectral 'ring', and constricted vocal modes (defined as a loud vocal production, using a raised jaw, no lip protrusion, a small lip opening, retracted tongue tip, raised tongue dorsum, lowered soft palate, and the subjective sensation of a squeezed or constricted throat) in a listening test with spectrographic evaluations and conventional radiography. In a perceptual test, 80% selected the ring vocal productions as the louder sound, although analysis revealed both ring and constricted conditions had a similar pattern of overall amplitude levels. The only spectral differences between ring and

³The technique is described by the author of the study in the following way. An acoustic pulse is introduced to the mouth, and analysis of the reflected sound allows reconstruction, using complex algorithms, of the *area-distance* function.

constricted modes that appeared were the higher amplitudes in the spectral regions between 1.3kHz and 2kHz and at 3.6kHz. Although the vocal sound utilising the singer's formant appeared to be louder than the constricted form of vocal articulation, the recorded SPL was only slightly higher. This implies that the effect of the singer's formant gave the impression that the vocal sound was more predominant. This would clearly be an important quality to encourage in singing training.

The different physiological states employed by the subject between the different modes of singing assessed by Acker (1987) are shown in figure 3.4. The mouth opening was recorded as being considerably wider. The soft palate was further above the palatal plane for the ring mode than for the constricted mode. The ring phonation mode exhibited a relatively greater cross-sectional area at the point in the vocal tract above the vocal folds.

Sundberg (1987) demonstrated that a singer can change the shape of the vocal tract to align formants near to harmonic frequencies. He observed that this can be used as a strategy of tuning the resonator to increase the sound pressure level (SPL) of a vowel, particularly in the female voice. This technique may be used by a professional singer to enhance the spectral region of the singer's formant.

Although use of these techniques would be hard to encourage directly at a physiological level, it may be that a form of graphical visualization based on acoustic analysis, operating in real-time, could incite the appropriate physiological actions that are required for this phenomenon.

3.4.3 Assessment methods

The spectrogram presents a visual representation of spectral amplitude from which the singer's formant phenomenon can be observed. It is established in a diagnostic context (i.e., Sataloff, 1992) and there are also reports of its use as a tool for inciting vocal technique in certain contexts (such as the training of deaf children; e.g., Stark, 1972; Goldstein & Stark, 1976; Stewart, 1976; Huggins, 1980). However, in many cases the high degree of information presented to the user is more likely to confuse rather than inform. Analysis must be carefully applied in order to derive valid and useful information for improving vocal technique (Miller and Schutte, 1990; Brooks et al., 1981), which greatly reduces the applicability of the spectrogram as an easy-

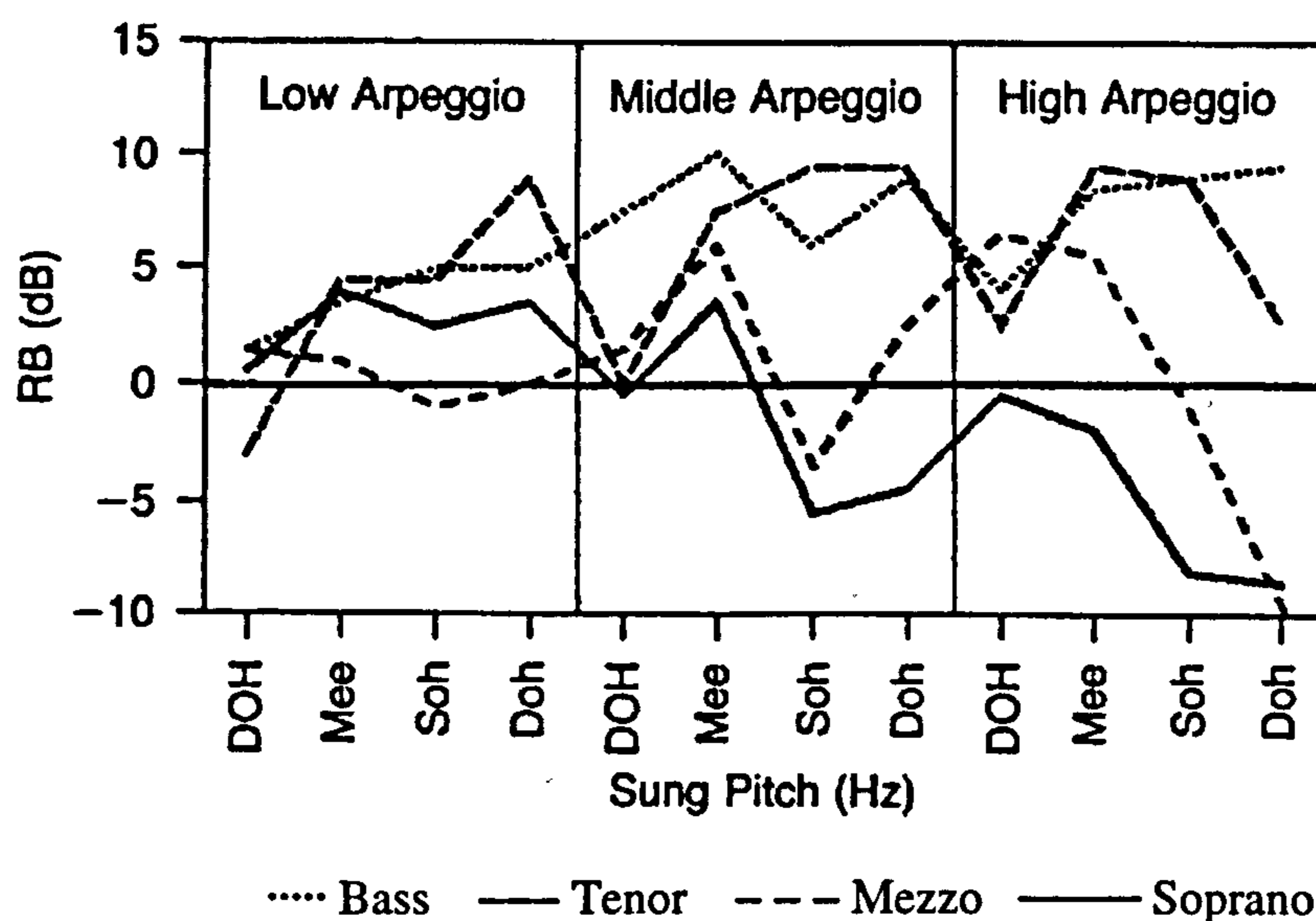


Figure 3.5: Bass, tenor, mezzo and soprano RB profiles. The capitalisation of DOH marks the bottom note of an arpeggio. (From Watson, 1992b)

to-use tool for real-time feedback. This is summarised in the following quote :

“Another computer-based aid currently in use is the Speech Spectrographic Display...but the complexity of the display may be an inherent disadvantage...” (Brooks et al., 1981)

A method for reducing the acoustic spectrum information to a set of easily interpreted parameters related to vocal development, such as the measure of the singer’s formant discussed previously, is therefore likely to be useful for the developing voice user.

Watson (1992a; 1992b) developed one such measure. He concluded that the ‘singer’s formant’ is a misleading description and chose to evaluate higher partial enhancement in terms of the difference between the mean of the two largest partials in the 50-1800Hz band and the mean of the two largest partials in the higher partial band 2200-5000Hz. He used the term ‘resonant balance’ (RB) to describe the methodology. This measure was also used to determine objective profiles of different types of singing. Figure 3.5 illustrates the RB profiles for bass, tenor and soprano. Distinctive contrasts between the voice types are visible.

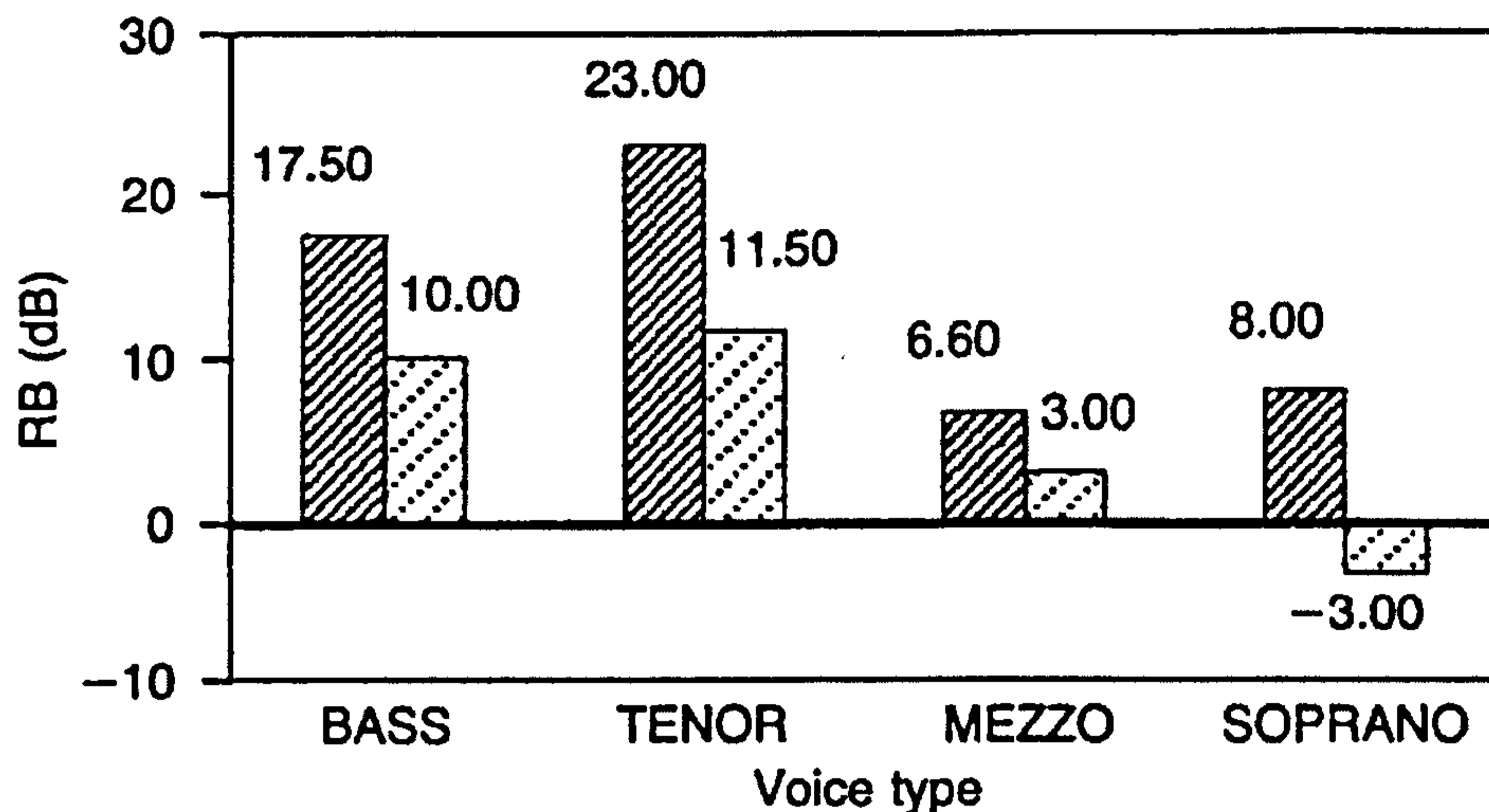


Figure 3.6: Peak values for resonance balance measured from Scottish Opera singers. Each bar is annotated with the actual RB level that it represents. The darker shading represents the /i/ vowel. The lighter shading represents the /u/ vowel. (From Watson, 1992b)

However, a real-time measure of RB would require complex contextual analysis for which enough data has not currently been derived. For example, maximum RB values for each class of singer are shown in figure 3.6 for the vowels /u/ and /i/. The pattern of RB is distinctly different for each vowel. Other vowels may exhibit different trends, but these remain uninvestigated. In addition, the specific approach measure does not have a wide research base. This algorithm would also be fairly demanding to be successfully implemented in a real-time digital computer system. The original research was carried out by hand and eye judgement (Watson, 1993). An automated implementation would require a high resolution fourier transform in order to ensure accurate detection of the appropriate partials in the spectral signal. The time taken for such a high level of processing in combination with subsequent visual display of an appropriate form is likely to be beyond the real-time performance of currently available computer systems.

3.4.4 A real-time measure

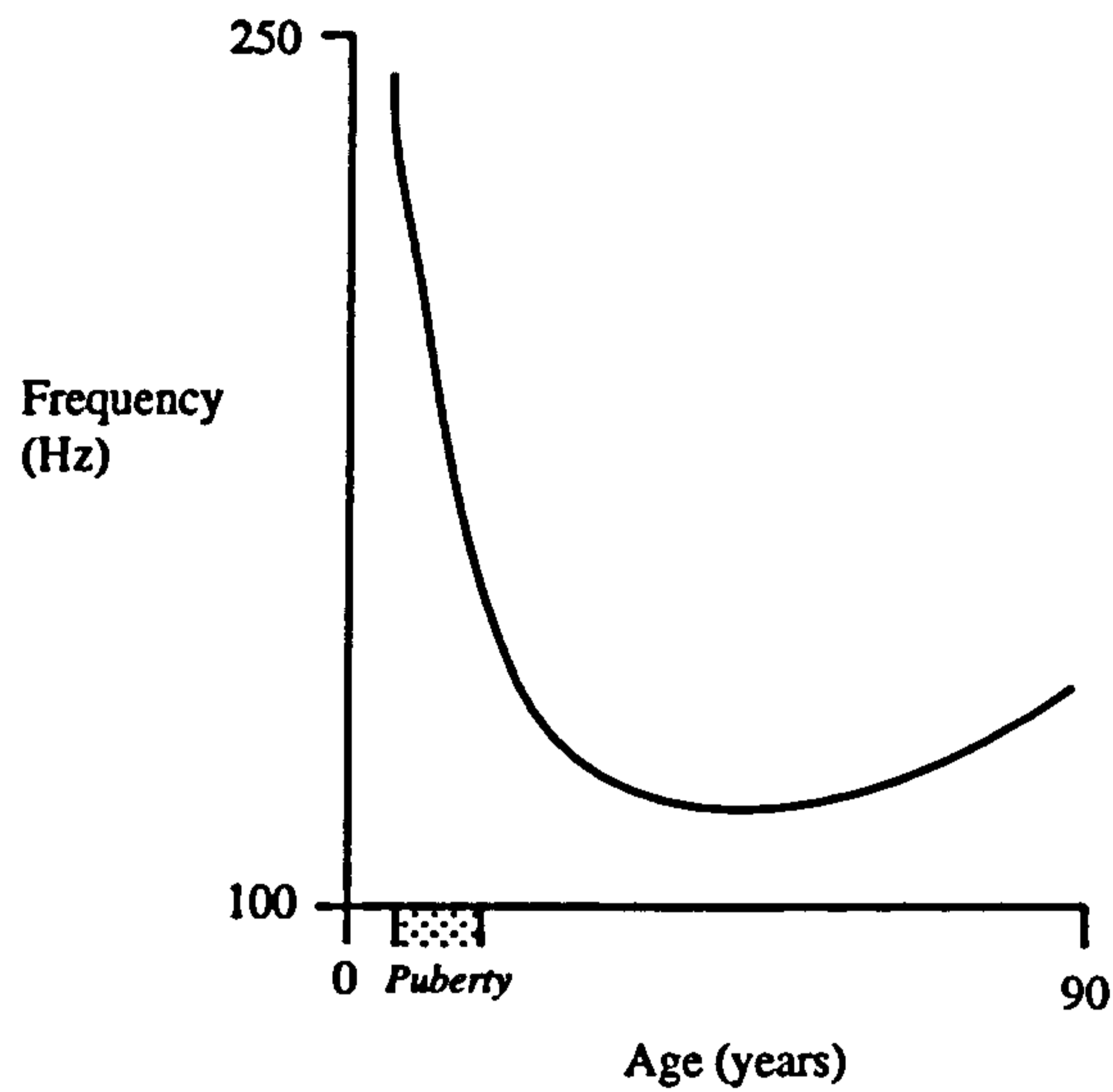
The display of a single parameter representative of the singer's formant phenomenon, presented in a feedback context other than the display of the entire acoustic spectrum might therefore be a useful attribute for a real-time voice tuition system. To achieve this, a method used to evaluate the phenomenon of the singer's formant objectively from the digital representation of the acoustic signal was developed. This is described in section 5.8.1.

3.5 Fundamental frequency

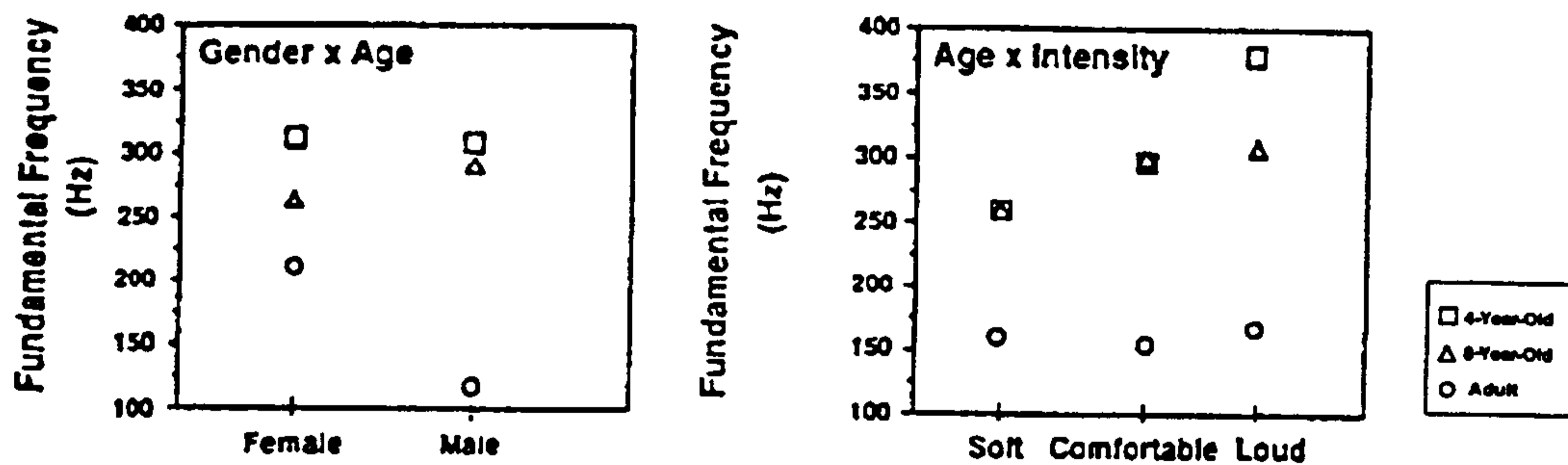
3.5.1 Introduction

The fundamental frequency of the voice (often referred to as 'F0') has been identified as a highly important parameter of speech and singing (Crystal, 1988; Baken & Orlikoff, 1987). It is affected by a wide variety of issues, including sex, age (Aronson, 1985) and health together with the acoustic environment (Summers et al., 1988) and the nature of the speech task. An example of the latter is indicated by the results of a study indicating that speakers undergoing police interviews frequently exhibit a lower mean and/or modal F0 value than for a disputed telephone recording for the same speaker (Hirson et al., 1993). In addition, it is well known that all languages rely to some extent on intonation to convey information (Crystal, 1988). Abberton & Fourcin (1978), for example, demonstrated that listeners were able to identify familiar speakers solely on the basis of suprasegmental laryngeal information (predominantly mean fundamental frequency and the fundamental frequency contour shape).

There is a natural pattern of change in average fundamental frequency with age. This is illustrated for males (10-90 years) in figure 3.7(a) (Kent & Bukard, 1981). This trend is verified and a more gradual rate of decline indicated for females in the results of an examination into average fundamental frequency for male and female subjects aged 4, 8 and at adult age (Stathopoulos & Sapienza, 1993). This is shown in figure 3.7(b). Other studies have verified considerable changes evident at the age of puberty (Penderson et al., 1982; 1984). Figure 3.7(b) also illustrates differing levels of fundamental frequency with intensity for males and females across age.



(a) Male fundamental frequency against age. The data points are a composite of averages taken from various published studies. For females, the level is stable during middle age, decreasing later. (After Kent & Burkard, 1981)



(b) Fundamental frequency against age and intensity. (From Stathopoulos and Sapienza, 1993)

Figure 3.7: Two illustrations of the change in fundamental frequency with age

<i>Rating</i>	<i>Number of subjects</i>	<i>Mean singing range</i>	<i>% frequency perturbation</i>	<i>Lesion</i>
Very deficient technique	9	25	1.05	3
Deficient technique	9	24.5	.75	6
Expected technique	6	27	.91	5
Above expected technique	3 (F)	42	.65	1
Good technique	3 (F)	65	.30	3 (early)

Table 3.1: The relationship of vocal technique ratings to singing range, from very deficient technique (top row) to good technique (bottom row). (F) denotes subjects were female only. Singing range is in semitones. (After Teachey et al., 1991)

Fundamental frequency has also been shown to be indicative of mood (Williams & Stevens, 1972). For example, one study found that 80% of 112 listeners were able to determine the difference between voice source only recordings of people immersed in mental depression and the same group of people after recovery (Nilsson & Sundberg, 1985). This has clear connections to areas of vocal performance such as acting where, for example, the impression of mode and tone is critical to the vocal delivery.

The difference in F0 from one instant to the next is called *jitter*. It is used as a measure of voice perturbation, often in combination with shimmer (i.e., Ludlow et al., 1987; Orlikoff & Kahane, 1991; Slavit & Maragos, 1994). Gelfer et al. (1991) concluded that trained singers exhibit a more consistent pattern of jitter performance than untrained singers. A real-time display of the measure may be of benefit in encouraging vocal stability.

In the following sections, the role of F0 as an indicator of changes in the physiology of the developing voice and the potential applicability of this voice parameter to a system intended to help enhance voice development are considered.

3.5.2 Fundamental frequency range

The fundamental frequency age span for males aged 20-60 years is approximately

107-120Hz (Hollien and Ship, 1972). The corresponding data for women in the same age range is 214-227Hz (Kelley, 1985). It has been demonstrated that frequency range tends to be directly proportional to a rating of vocal quality, as based on an acoustic judgement of the subjects voice (Teachey et al, 1991). For example, deaf children have also been shown to speak with a narrower fundamental frequency range than their normally hearing peers, across all ages and for both sexes (Beijk, 1992). Table 3.1 tabulates a correlation between singing range and the proficiency of vocal technique for a number of subjects who had some form of medical complaint relating to voice usage. Subjects with better technique exhibit considerably higher F0 range, in addition to less F0 perturbation and generally less lesions.

3.5.3 Fundamental frequency and intensity range

The *phonetogram* is a graphical display of the fundamental frequency and intensity capabilities of the voice. Many studies (i.e, Gramming et al., 1988; Gramming, 1991; Awan, 1993) have indicated that the information displayed by a phonetogram appears to have a wide range of applicability in the assessment of singing and speech. Awan (1993) noted that there were significant correlations between the total intensity range and the intensity range used in speech for trained female vocalists. His study demonstrated that trained singers exhibit improved overall F0 and intensity capabilities in comparison to their untrained counterparts. These phonetograms are shown in figure 3.8. A major conclusion of the study was that “maximum performance” tests can provide important information about phonation activities.

3.5.4 Range analysis in real-time displays

Clearly, the range of F0 and vocal intensity are important measures of vocal performance in certain contexts. But by its very nature the range of a certain vocal parameter is not a suitable measure for *real-time* display. (Similarly, a minimum speech sample duration of 90 seconds is required in order to attain a stable F0 distribution (Barry et al., 1990)). However, it would be possible to derive the range of a vocal attribute during the process of a tuition session. This measure could then be suitable as a long-term indication of vocal development rather than as a short-term measure suitable for real-time display, although once the subject's voice range is

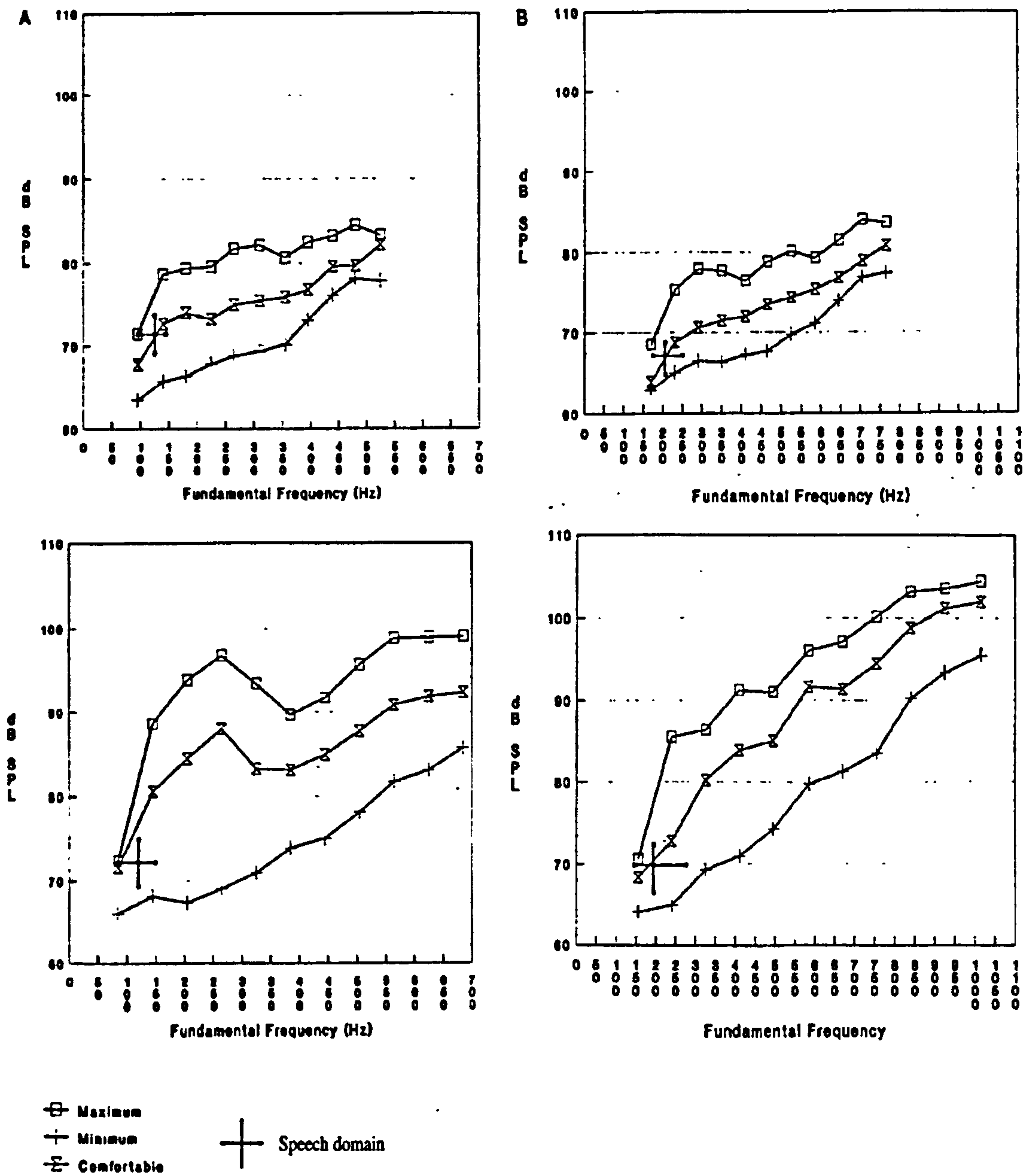


Figure 3.8: Superimposition of speech F0-SPL characteristics and phonetograms for untrained male (top left) and female (top right) and trained male (bottom left) and female (bottom right) vocal groups. Data points correspond to 10% increments of the F0 range (in Hz). (From Awan, 1993)

known a real-time display could be used to encourage an expansion of this range.

3.5.5 The use of fundamental frequency in a real-time display

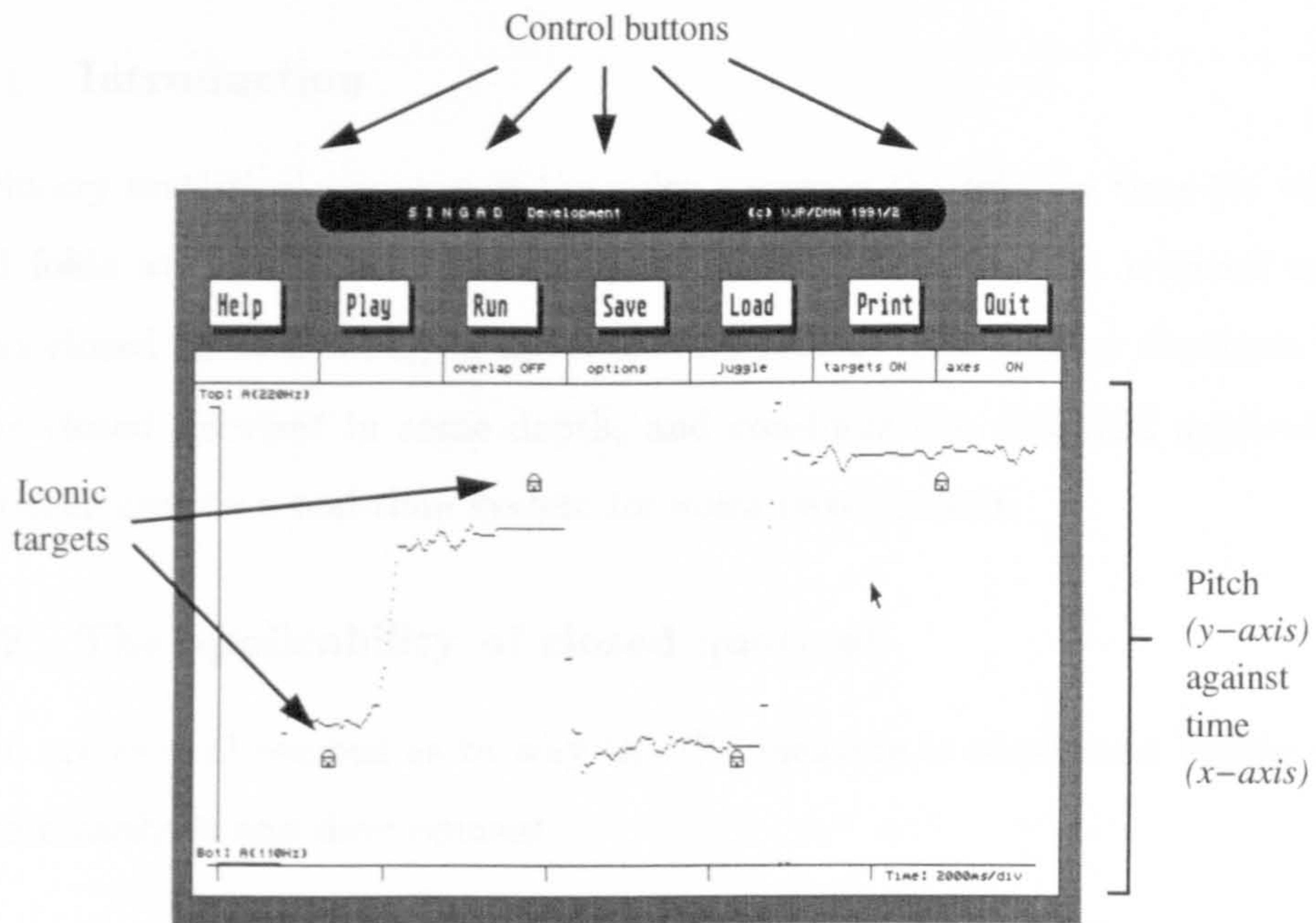
F0 has been widely examined as a parameter for real-time display. A number of systems have been developed for its analysis and display in a real-time domain. One system, discussed in more detail in section 3.6.7, enables real-time analysis of F0 and other voice source parameters, and is able to display statistical information in several of standard analytical modes, including scattergrams and two-dimensional graphs (Howard and Garner, 1992). A very similar system has also been developed by Spaai et al., (1993) with extensions to include the plotting of a target intonation pattern, with word annotation at the correct loci. One system aimed towards developing voice users of the primary school age group is the SINGAD (SINGing Assessment and Development system) system (Welch et al., 1988; Welch et al., 1989a; Howard and Welch, 1993). It employs visual targets in the form of bitmaps of various shapes, such as flags and rockets. An example of the use of bitmaps of houses is shown in figure 3.9(a). The user can attempt to guide the moving line representative of pitch by raising or lowering the pitch of his/her voice and aiming at or around the icons (Rossiter et al., 1993).

Other software products have been produced aimed at adult singing tuition (for example, Smith & Monk, 1994; Morgan, 1994). These use conventional music notation to present pitch stimuli and the response of the user.

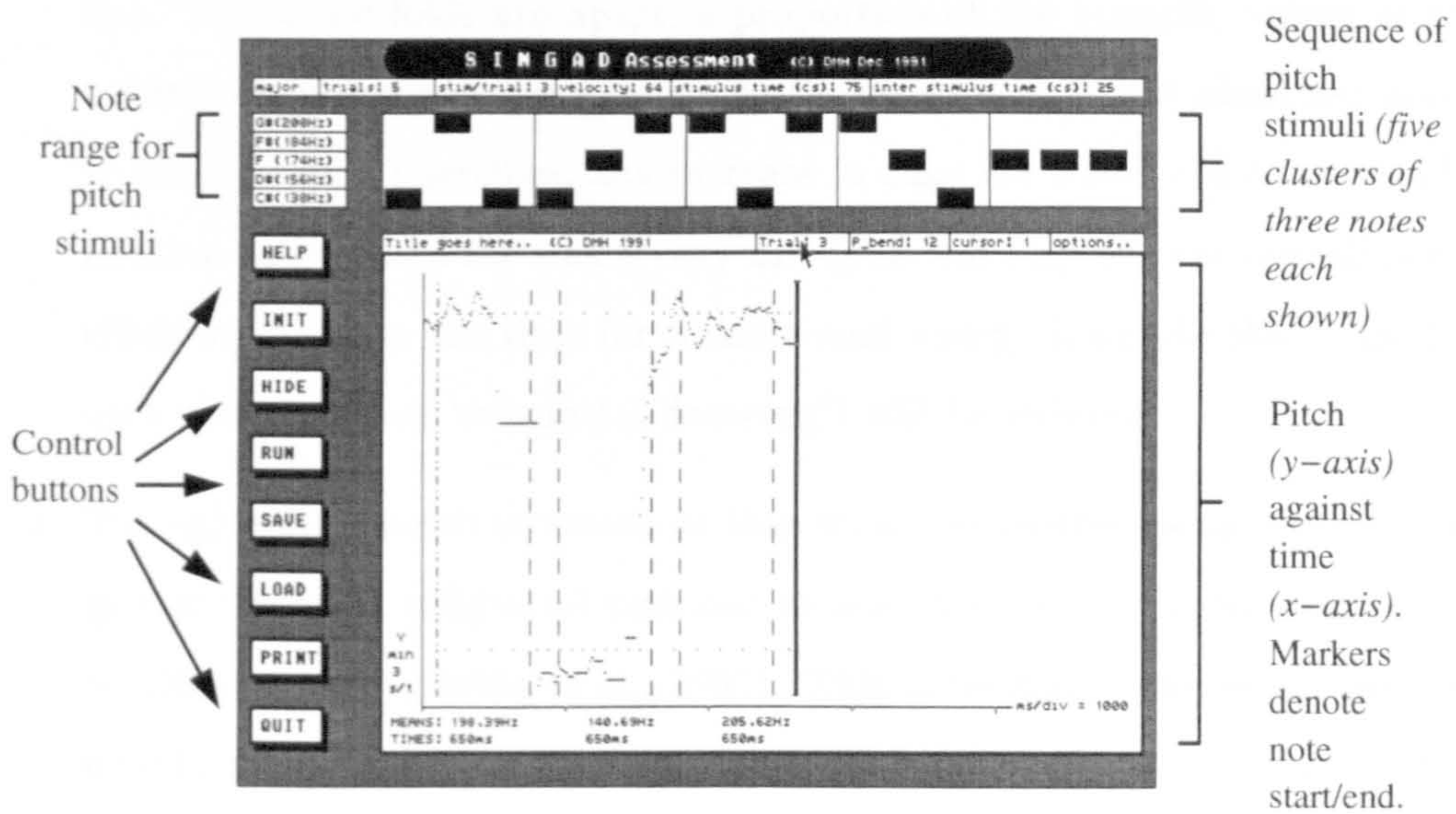
An ability to measure the pattern of F0 is clearly of importance in the design of a real-time tuition analysis and display system. The currently existing systems considered are inflexible in the way they choose to present a visual indication of the status of F0, and are generally unable to jointly consider the performance of other parameters. They are all based around standard graphical techniques, predominantly with the x-axis mapped to time and y-axis mapped to frequency. Very little flexibility is proffered from which an appropriate form of voice analysis and subsequent graphical interaction may be selected.

3.6 Laryngeal closed method

3.6.1 Introduction



(a) SINGAD development screen



(b) SINGAD assessment screen

Figure 3.9: SINGAD screen displays (From Rossiter et al., 1993)

3.6 Laryngeal closed quotient

3.6.1 Introduction

A primary statistical measure of the voice source is the relative time for which the vocal folds are in contact (Ardran et al., 1953). This is often referred to as the larynx closed quotient (CQ) of the vocal fold period. This section discusses the role of the closed quotient in some depth, and considers the potential applicability of this parameter to a real-time system for voice development.

3.6.2 The applicability of closed quotient

There are several reasons as to why the CQ measure is considered highly relevant to voice analysis and development.

1. Howard (1992) suggests the following theoretical base. When the vocal folds are in contact, acoustic energy is transmitted to the listener via the orator's lips. When the folds are apart, a proportion of the acoustic energy is transmitted via the subglottal cavities to the lungs where it is absorbed and lost to the listener. Therefore, any increase in time for which the folds are closed relative to the time for which they are open will improve the overall acoustic efficiency because the time for which sound energy is lost to the lungs during each cycle (termed 'subglottal damping') will be reduced.
2. Transglottal pressure increases so that when the glottis opens, the product of glottal flow and subglottal pressure means that there is more energy in the resulting release (Rubin et al., 1967). This is related to the signal processing waveform and corresponding spectrum discussed previously in section 2.7.2.
3. Prolonged closed quotient helps maximise the mucosal wave, which plays a major role in the dynamics of vocal fold function and the conversion of air flow to sound energy (Watson, 1992).

3.6.3 Changes in closed quotient across different levels of intensity

Stathopoulos & Sapienza (1993) used a flow mask to calculate the level of open quotient (OQ) for men and women at three different levels of vocal intensity. Ex-

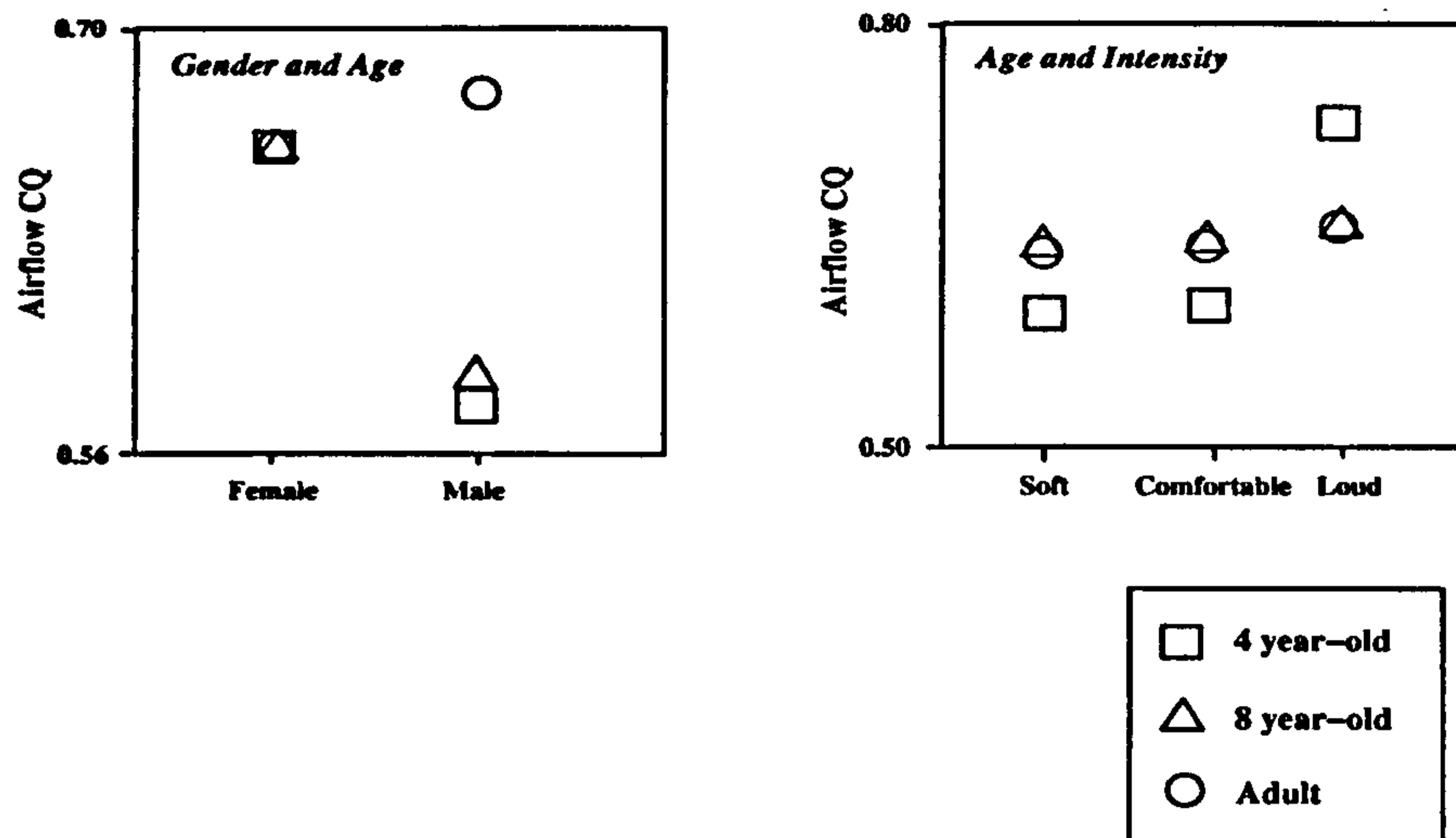


Figure 3.10: Open quotient against age and intensity. (After Stathopoulos & Sapienza, 1993)

pressed in terms of CQ, women were found to have similar levels of CQ for soft and comfortable modes of phonation but a considerably higher level for the loud mode of phonation. For men, there was a progressive increase in the level of CQ across the three levels of intensity. These data are illustrated in figure 3.10, in addition to indications of the pattern of CQ across age and gender. It may be that the potential for a real-time analysis and display system considering CQ in solitude would be considerably enhanced by additionally considering trends in SPL and other parameters.

3.6.4 Changes in closed quotient for different modes of singing

The CQ parameter has also been examined as a measure of different singing styles. Howard (1992) considered the change in CQ for a tenor singing a fragment of an Elizabethan madrigal in different styles. Although following a fairly consistent pattern of CQ measures across fundamental frequency, the average operatic CQ values were consistently higher than the conventional early music values, which were consistently higher than the Elizabethan values.

Evans & Howard (1993) investigated the difference between the quality known as 'belting' and an operatic technique in a female subject. Scattergrams of laryngeal

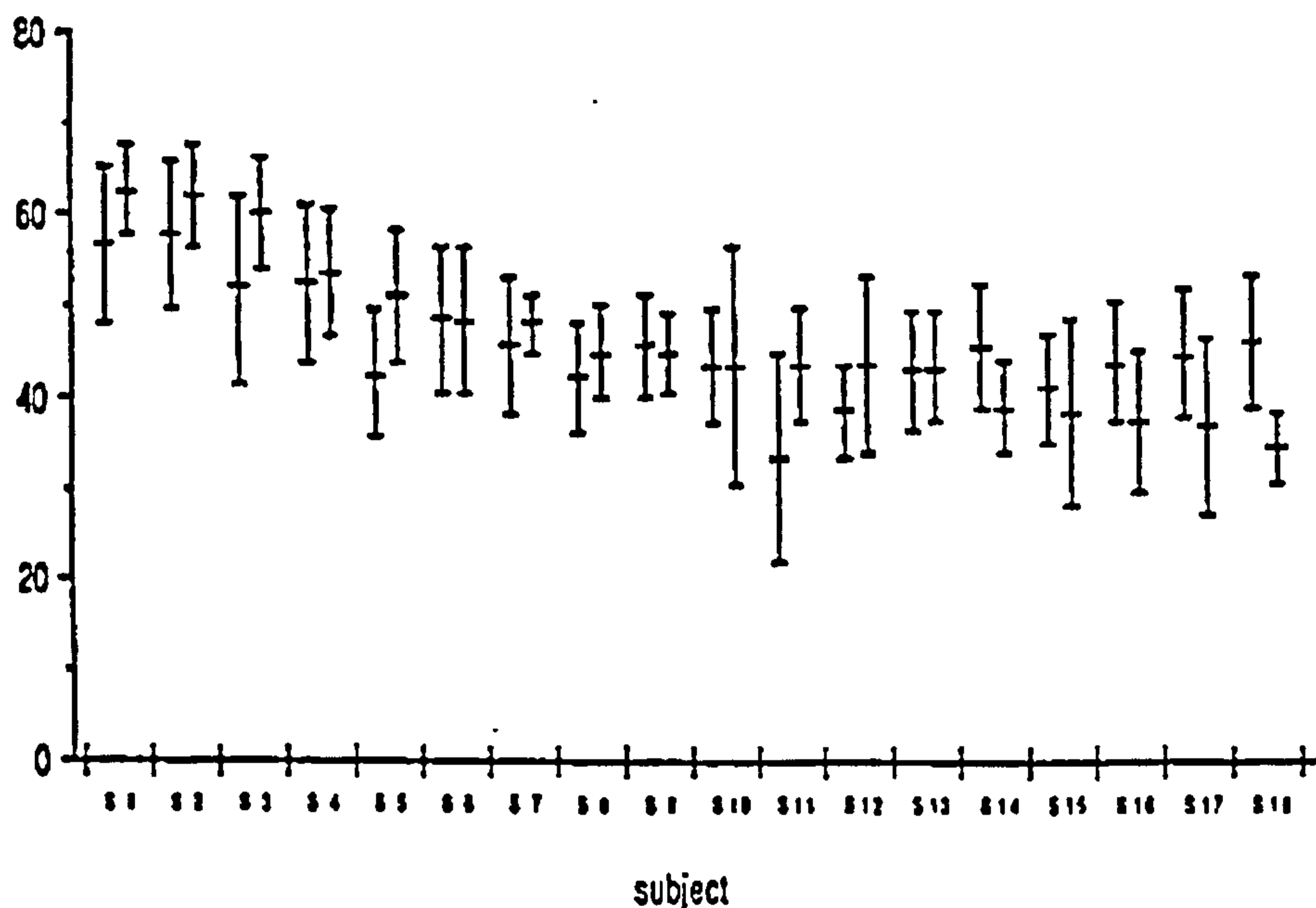


Figure 3.11: Adult male subjects - Mean closed quotient (with standard deviation bar measure) for each subject producing the read passage (left) and the sung scales (right). Subjects are ranked in order of decreasing vocal experience. Mean CQ (%) values are mapped to the y-axis. (From Howard et al., 1990)

CQ against fundamental frequency were made for several vowels sung at different pitches. The level of CQ for the belting quality remained relatively constant across all vowels and pitches at approximately 50% to 60%, whereas the opera quality exhibited a considerably lower CQ for the lower pitches, although it rose with pitch.

3.6.5 The relationship between closed quotient and vocal training

Closed quotient has been demonstrated to be indicative of the level of training in male, non-dysfunctional subjects.

With regards specifically to adult males, Howard et al. (1990a) recorded and analysed 18 subjects whose vocal training ranged from no formal training to a high level of professional training. He found that the level of CQ was in direct proportion to the level of training of the subject. Figure 3.11 shows the concluding graph of this paper, in which the correlation is clearly shown. Gauffin & Sundberg (1989) concluded that the non-singers in their study typically shifted phonation towards

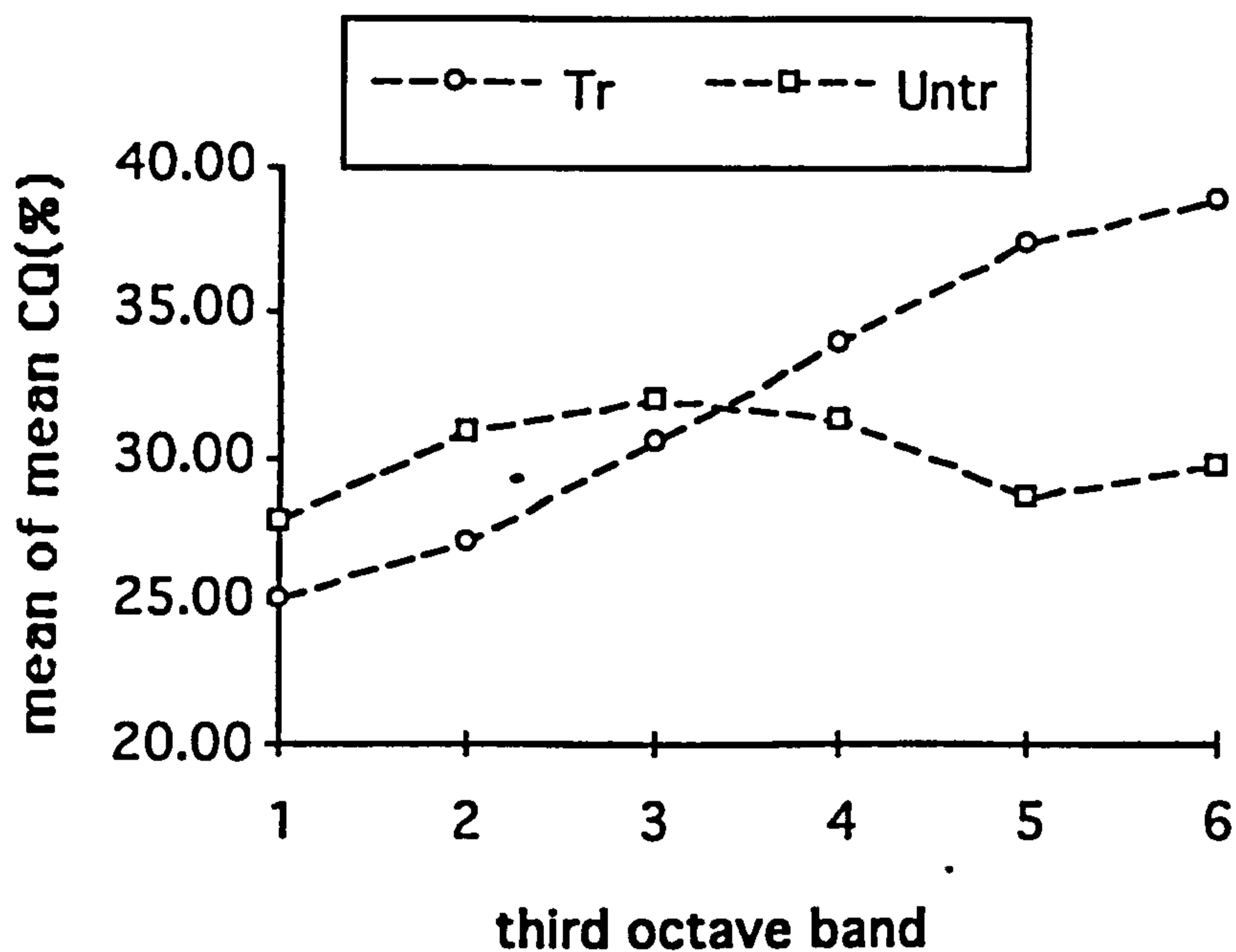


Figure 3.12: Adult female subjects - Mean closed quotient means in each third octave band for the trained and the untrained groups. (From Howard, 1994)

pressed phonation (i.e., unnecessarily increased CQ) when the loudness of phonation was increased, whereas the phonation type of the trained singers changed less with loudness.

With regards to adult females, an initial assessment by Howard (1994) considered the variation of CQ with changing fundamental frequency for a group of adult female singers with varying singing backgrounds. A comprehensive assessment of 26 female subjects of varying degrees of singing experience concluded that the pattern of CQ across F0 was generally constant ($\pm 2\%$) for subjects of little training, but was generally increasing for subjects who had experienced a high level of training. The concluding graph from this paper is shown reproduced in figure 3.12. An earlier study (Howard et al., 1990b) supports the conclusions of this work.

Further evidence of a contrasting performance of CQ across F0 between males and females is provided by Lindsey et al. (1988), who found that in speech the females had decreasing CQ with increasing fundamental frequency while the males did not.

A succinct summary of the performance of CQ across F0 was provided in Howard (1994):

- *For female subjects:*

1. CQ tends to be reduced for pitches below D4 (587.33Hz) and increased for pitches higher than B4 (987.77Hz) with training, and
2. the CQ/F0 gradient within the pitch ranges: G3 (392.0Hz) to G#4 (830.61Hz), and B4 (987.77Hz) to G5 (1568.0Hz) tends to correlate positively with the number of years singing training/experience.

- *For male subjects:*

1. CQ remains essentially constant with F0, and
2. CQ means tend to correlate positively with the number of years singing training/experience.

3.6.6 The suitability of closed quotient as a real-time measure

As Howard (1992) concluded, a measure of CQ would seem to have a considerable degree of applicability to real-time display systems. However, caution is required in interpreting the results of CQ measurement. CQ has been shown to perform in context-dependant ways, and any real-time display would therefore benefit from being able to assess and present information in a form appropriate to the context in which the parameter is being expressed. Watson (1992b) warns

There is... a caveat for laryngologists making laryngograph examinations. A long closed quotient is an attribute of an operatic singer but an untrained voice, without the muscular development achieved during training, will probably be suffering from a pathological voice, if seen to present with extended closed quotients of this order. The diagnosis will be quite different from that for a fully trained singer.

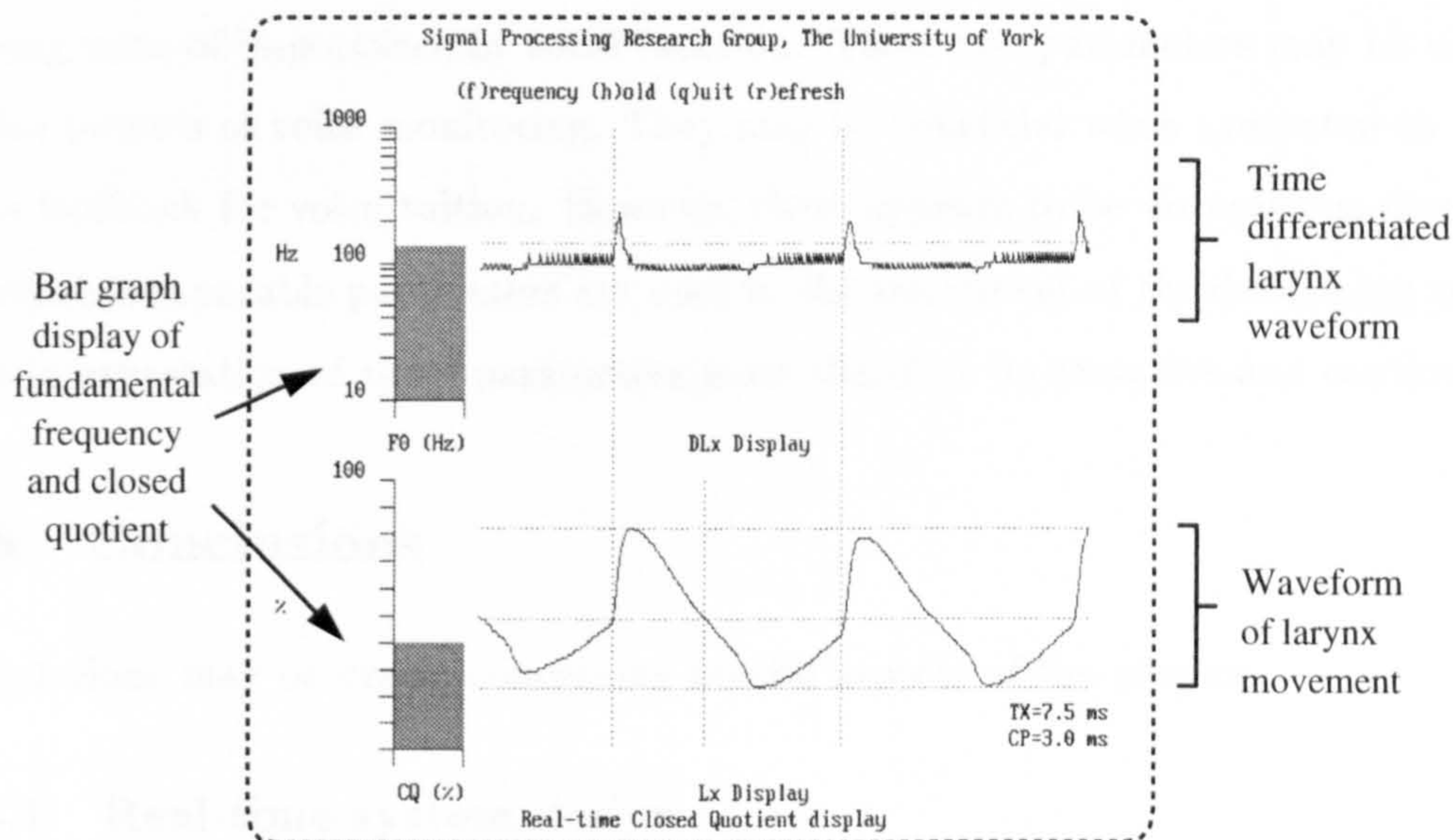
A system intended for use as a tool for vocal development must be able to be configured to draw conclusions about the merits of a change in the level in CQ within

the context of the vocal act. Simply inferring that, for example, the attainment of a relatively higher level of CQ for an adult male singer is an improvement may not induce improved vocal performance, depending on the context of the increase, but instead incite incorrect vocal technique and lead to damage of the vocal mechanism. With regard to problems such as this, it is expected that the real-time feedback tool to be developed would be designed for use *with* a tutor, and not in place of the tutor.

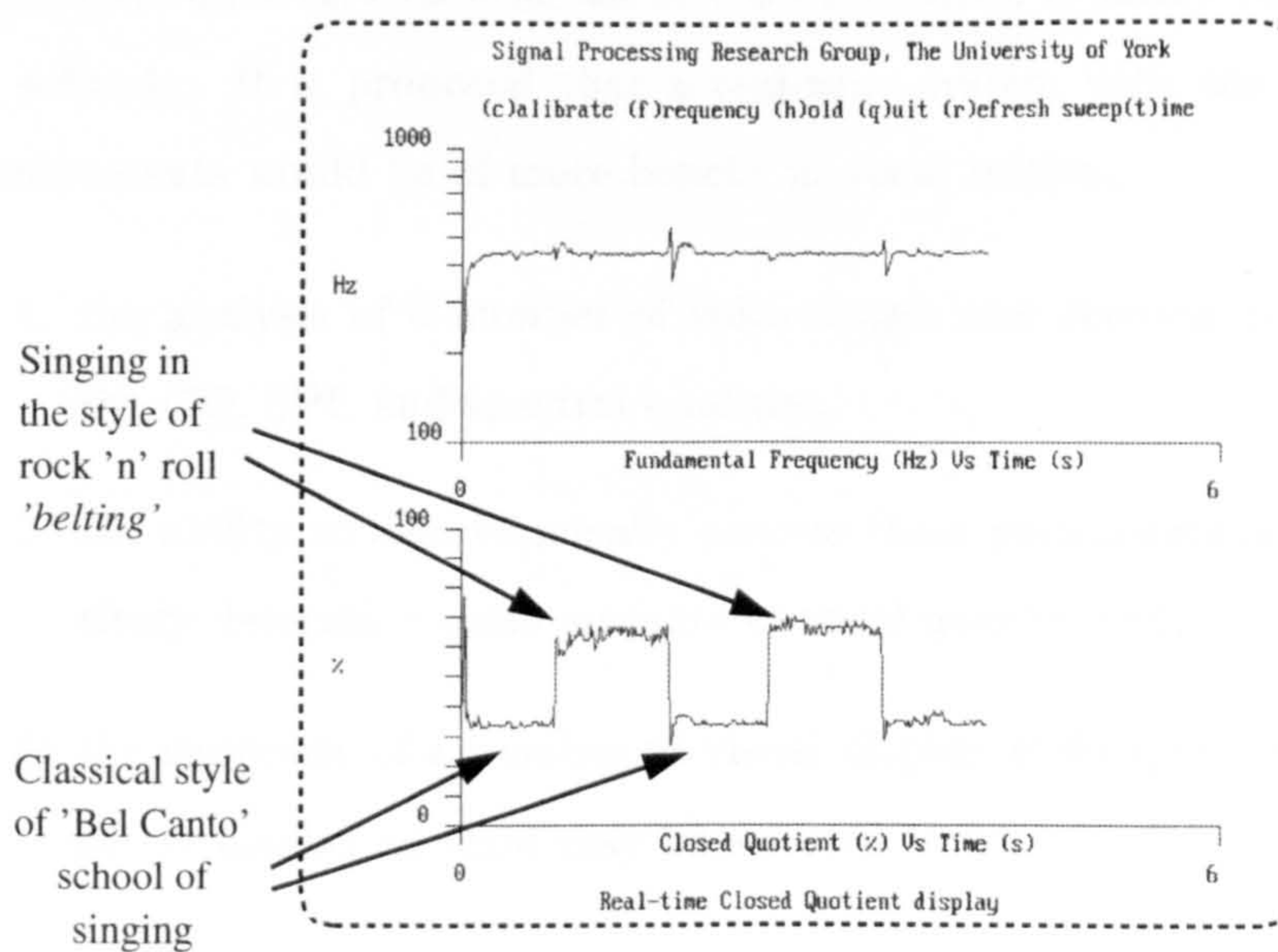
3.6.7 Previous real-time closed quotient systems

In one research project, an analogue circuit was devised and produced (Rothenberg & Mahshie, 1988). This processed the Lx waveform and produced an output level indicative of the level of CQ which could be displayed on an oscilloscope. A more advanced project enabled real-time monitoring of closed quotient in addition to fundamental frequency (Howard & Garner, 1992; Rossiter et al, 1993). Using an Ariel DSP board and an IBM-PC compatible computer, the Lx waveform is processed in real-time on a cycle-by-cycle basis to give F0 and CQ traces which are then displayed graphically against time in real-time. Display options include bar graphs, two-dimensional graphs, and a display of the incoming waveform with the first order differential shown time-synchronised (a calculation which is of prime importance in the algorithmic operation of the system). Example screen displays are shown in figure 3.13.

The latter system is functionally very well suited for stand-alone use as a monitor and analysis system of CQ and F0. However, the system suffers from certain basic constrictions, such as a small selection of fixed axis boundary options, and it is unable to employ other, potentially more appropriate methods of graphical display. It is also unable to consider additional parameters (whether in solitude or combination) such as the level of acoustic output, or the distribution of energy across the acoustic spectrum.



(a) 'Snapshot' of laryngeal activity and associated signal processing



(b) Typical real-time use of system to illustrate the difference in closed quotient between two different singing styles

Figure 3.13: Screen displays of real-time CQ system (From Rossiter et al., 1993)

3.7 Rates of larynx opening and closing

In the previous chapter it was established that the *rates* of larynx opening and closing were of importance in vocal function. These two parameters may be useful in the process of voice monitoring. They may be beneficial when presented as real-time feedback for voice tuition. However, there appears to be no previous research in which comparable parameters are used in the assessment of the developing voice. The interpretation of these parameters must therefore be tentative and cautious.

3.8 Conclusions

Conclusions may be drawn concerning several aspects of the project.

3.8.1 Real-time system design

A number of real-time visual displays have been reviewed. These have been shown to operate on a very reduced set of voice measures, typically one or two parameters in solitude. It is proposed that a real-time system with the following functional requirements would be of more benefit in vocal tuition;

1. the analysis of a number of voice source and acoustic parameters, including F0, CQ, SPL and spectral qualities,
2. the ability to algorithmically process these parameters in order to more effectively determine some measure of vocal quality, and,
3. the support of a number of visual display techniques from which an appropriate display method may be selected to convey visual feedback.

The way in which these points have been addressed is discussed in chapter 5.

3.8.2 Parameters indicative of vocal useage

In this chapter, a number of voice source and acoustic parameters have been identified as potentially useful for real-time visual feedback. These are:

- SPL as an objective measure of loudness

- The ratio between spectral amplitude in the 2 to 4KHz band and the 0 to 5KHz band as an objective measure of the 'singer's formant' phenomenon
- Fundamental frequency
- Closed quotient of the laryngeal period

In addition, there are indications that the following two parameters may reflect vocal development;

- The rate of opening of the vocal folds
- The rate of closing of the vocal folds

3.8.3 Longitudinal assessment

An accurate model of vocal development is a highly important attribute for any system intended for the assessment of vocal function. The studies discussed in this chapter that address measures of the developing voice (i.e., Gramming et al., 1988; Gramming, 1991; Awan, 1993; Howard et al., 1990) draw their conclusions from observations of several different vocal mechanisms at differing levels of vocal development (e.g, a trained group and an untrained group). An alternative approach in which the same set of vocal mechanisms is analysed during the process of vocal development may be a more appropriate approach for the identification of changes that occur as a *function* of development.

This is addressed in the next chapter, in which three studies concerning the identification of longitudinal parameters are presented.

Chapter 4

Longitudinal voice studies

4.1 Introduction

This chapter presents three longitudinal studies that attempt to identify some voice source and acoustic parameters indicative of adult vocal development. Identified trends are likely to be appropriate for consideration in the design of a voice analysis and assessment system able to operate in real-time.

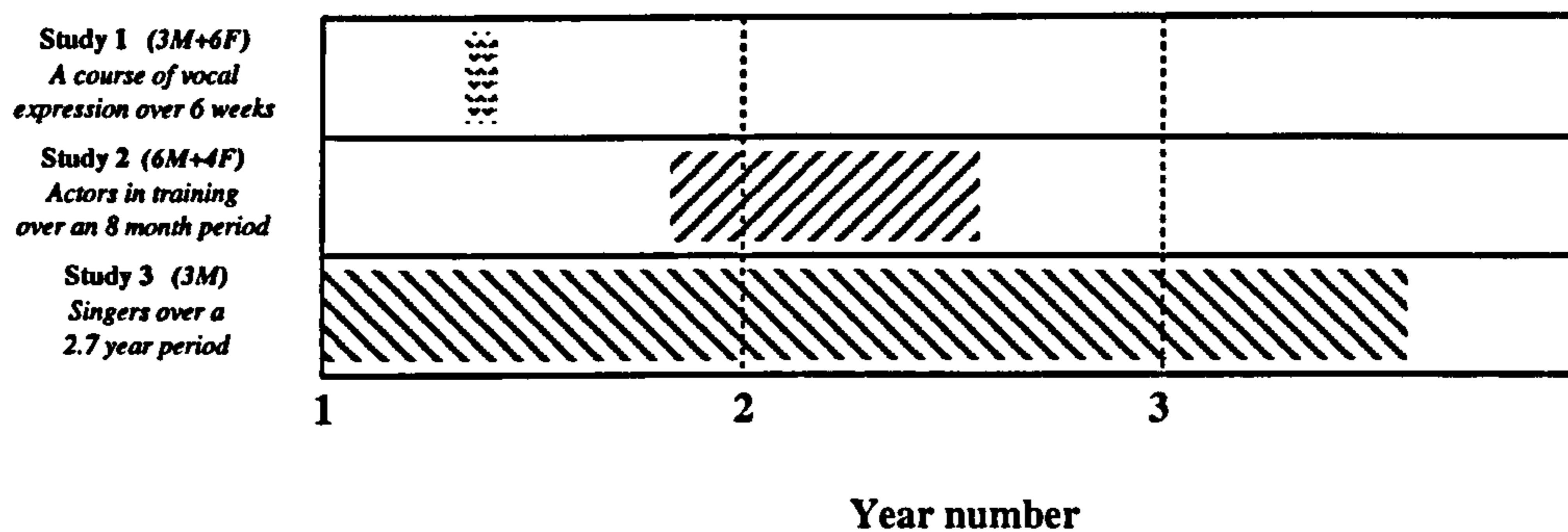


Figure 4.1: Relative length and timing of the three studies. Key: xM=number of male subjects. xF=number of female subjects.

The first study considers objectively measured changes in a group of male and female subjects over a six week part-time course of vocal expression for their speaking voices. The second study considers a group of male and female acting students in full-time training during an eight month period. The third study considers three male subjects who underwent regular singing lessons over differing periods up to 2.7

years.

A visual representation of the relative timing and length of these studies is shown in figure 4.1. The geographical context, the form of vocal tuition received by the subjects, the tutors, and the subjects themselves were different for each study. All subjects were classed as being vocally non-dysfunctional.

The ability to measure the SPL output of subjects was gained in the period prior to the commencement of study 2. For this reason, SPL data is available for the complete length of study 2 only.

4.2 Study 1 *A 6 week course of vocal expression*

4.2.1 Introduction

This study addresses changes in some voice source and acoustic output parameters of male and female subjects as a result of attendance at a short course of vocal expression.

4.2.2 A course for vocal expression

<i>Reference</i>	<i>Sex</i>	<i>Age</i>	<i>Occupation</i>
M1	M	22	Sound engineer
M2	M	25	Voluntary consellor
M3	M	38	Solicitor
F1	F	24	Radio production trainee
F2	F	55	Museum guide
F3	F	50	Counsellor
F4	F	41	Revenue control
F5	F	54	Journalist
F6	F	65	Lay reader
T1	F	64	Voice tutor

Table 4.1: Experimental subjects for study 1

A group of nine people attended a six week part-time practical course in vocal development taught by a professional voice tutor at Warwick University, England. Subjects were classed as not being vocally dysfunctional in any way. One two hour session was held per week, in the early evening. The subjects comprised of three men (age range 22-39; mean=28.33; sd=6.94) and six women (age range 24-65; mean=48.17; sd=12.93). The course tutor (female; age 64) was also recorded as a potential reference against which to consider measured parameters for the nine subjects. A summary is shown in table 4.1. Subjects were treated as a group, with each subject receiving the same treatment.

The overall aim of the course was to increase the ability of the students to use

their vocal mechanism and to increase awareness of the context within which it may be used.

Special guidance was provided in week 2 by a Speech Therapist, and in week 4 by a specialist in the Alexander Technique. Please refer to Appendix A for more detailed notes made by the voice tutor on the schedule undertaken. Further information on the context within which this course was developed may be found in Comins (1992) and Kingston (1994).

4.2.3 Recordings

In addition to the tutor each subject was recorded immediately before and after attending the course. Subjects were given identical instructions during both recording sessions. During the recording procedure subjects were required to remain at an approximate distance of 0.3m from the microphone. Subjects were asked to maintain a fixed distance from the microphone during recordings. They were asked to read the following items (reproduced in Appendix B) aloud in a manner they felt appropriate:

1. an extract from 'The Voyage' by Katherine Mansfield,
2. an extract from 'Androcles and the Lion' by Bernard Shaw, and
3. a sequence of sixteen consonant-vowel-consonant word structures.

The first item is pure narrative, the second requires the subject to express their voice 'in character', and the third item is a word sequence which requires the subject to articulate all the English vowel sounds. In combination, it is considered that these texts provide a fair portfolio of vocal activity across a variety of tasks.

The first two texts were not used as part of the course itself, whilst vowel articulation similar to that employed by the third test item was used in the tuition process.

Recordings were made of the voice source and acoustic output signals. A Sony TCD-D10-PRO DAT recorder was used to record the signals at a sampling frequency of 48KHz. The output from a Sennheiser MKH/40/P48 cardioid microphone was

recorded onto one channel and the output from an electrolaryngograph was recorded on the other channel.

The recorded signals were then transferred into a Silicon Graphics Indigo computer and stored in Speech Filing System (SFS) format (Huckvale et al., 1987; Edginton et al., 1992). Subsequent analysis of the voice source and acoustic output signals was made using the techniques described in chapter 3 using specially developed software.

4.2.4 Data analysis

Several voice source and acoustic output parameters identified in the last chapter were analysed. These are:

1. CQ,
2. Ratio,
3. F0,
4. the rate of opening of the vocal folds, and
5. the rate of closing of the vocal folds.

The results of the analyses are discussed in the following sections.

4.2.5 Results

4.2.5.1 Ratio

In chapter 3 the phenomenon of the singer's formant was discussed. The acoustic signals of the subjects were analysed in order to determine whether a change related to this phenomenon could be objectively measured. A distinctive change would imply that the physiological basis for the phenomenon may have occurred in the subjects as a result of the course of vocal tuition.

A procedure (described in section 5.8.1) was used to evaluate this phenomenon objectively from the digital representation of the acoustic signal. This was carried out for each acoustic recording in order to derive an indication of the spectral balance. The spectral range covering the frequency domain where the phenomenon of

		Ratio			F0		
		Males	Females	All	Males	Females	All
Passage 1	Before	10.5272	7.9608	8.8163	121.16	199.26	173.22
	After	9.6973	8.3381	8.7911	131.64	210.92	184.50
	Change %increase	-0.8299 -7.8836	0.3773 4.7394	-0.0251 -0.2848	10.48 8.65	11.67 5.86	11.27 6.51
Passage 2	Before	10.0544	7.5546	8.3879	134.33	206.87	182.69
	After	10.2420	9.2173	9.5589	161.97	226.91	205.26
	Change %increase	0.1876 1.8655	1.6628 22.0098	1.1710 13.9609	27.63 20.57	20.04 9.69	22.57 12.36
Vowels	Before	5.0068	4.6959	4.7995	127.33	198.25	174.61
	After	8.6077	5.4994	6.5355	140.74	211.32	187.79
	Change %increase	3.6009 71.9195	0.8036 17.1119	1.7360 36.1702	13.41 10.53	13.07 6.59	13.18 7.55
All three	Before	9.6116	7.1959	8.0012	126.77	201.53	176.61
	After	9.6626	8.2965	8.7519	143.81	215.76	191.78
	Change %increase	0.0510 0.5304	1.1006 15.2948	0.7507 9.3828	17.05 13.45	14.23 7.06	15.17 8.59

		CQ			Slc			Slo		
		Males	Females	All	Males	Females	All	Males	Females	All
Passage 1	Before	45.61	43.47	44.19	0.208756	0.190396	0.196516	0.050491	0.067627	0.061915
	After	40.34	43.49	42.44	0.191867	0.317169	0.275402	0.055305	0.065624	0.062185
	Change %increase	-5.28 -11.57	0.02 0.04	-1.75 -3.95	-0.016889 -8.090293	0.126773 66.584032	0.078886 40.142199	0.004814 9.534372	-0.002003 -2.961342	0.000270 0.435364
Passage 2	Before	46.05	46.81	46.55	0.195404	0.266988	0.243127	0.052881	0.072561	0.066001
	After	44.70	46.24	45.73	0.219094	0.415289	0.349891	0.059092	0.072656	0.068135
	Change %increase	-1.35 -2.94	-0.56 -1.20	-0.83 -1.78	0.023690 12.123409	0.148301 55.545845	0.106764 43.912781	0.006211 11.745944	0.000094 0.130235	0.002133 3.232437
Vowels	Before	47.11	45.04	45.73	0.197059	0.289253	0.258521	0.048478	0.065314	0.059702
	After	45.86	45.62	45.70	0.260549	0.301958	0.288155	0.048818	0.066931	0.060893
	Change %increase	-1.25 -2.66	0.57 1.27	-0.04 -0.08	0.063490 32.218832	0.012705 4.392469	0.029634 11.462712	0.000339 0.699969	0.001616 2.474961	0.001191 1.994528
All three	Before	46.00	44.76	45.17	0.201853	0.228282	0.219472	0.050935	0.068919	0.062924
	After	42.97	44.61	44.07	0.209747	0.352234	0.304738	0.055287	0.067949	0.063729
	Change %increase	-3.03 -6.58	-0.14 -0.32	-1.10 -2.45	0.007894 3.910810	0.123952 54.297904	0.085266 38.850572	0.004352 8.544851	-0.000970 -1.407312	0.000804 1.277980

Table 4.2: Summary of mean change for passage 1, passage 2, and vowel sequence for male, female and both genders combined

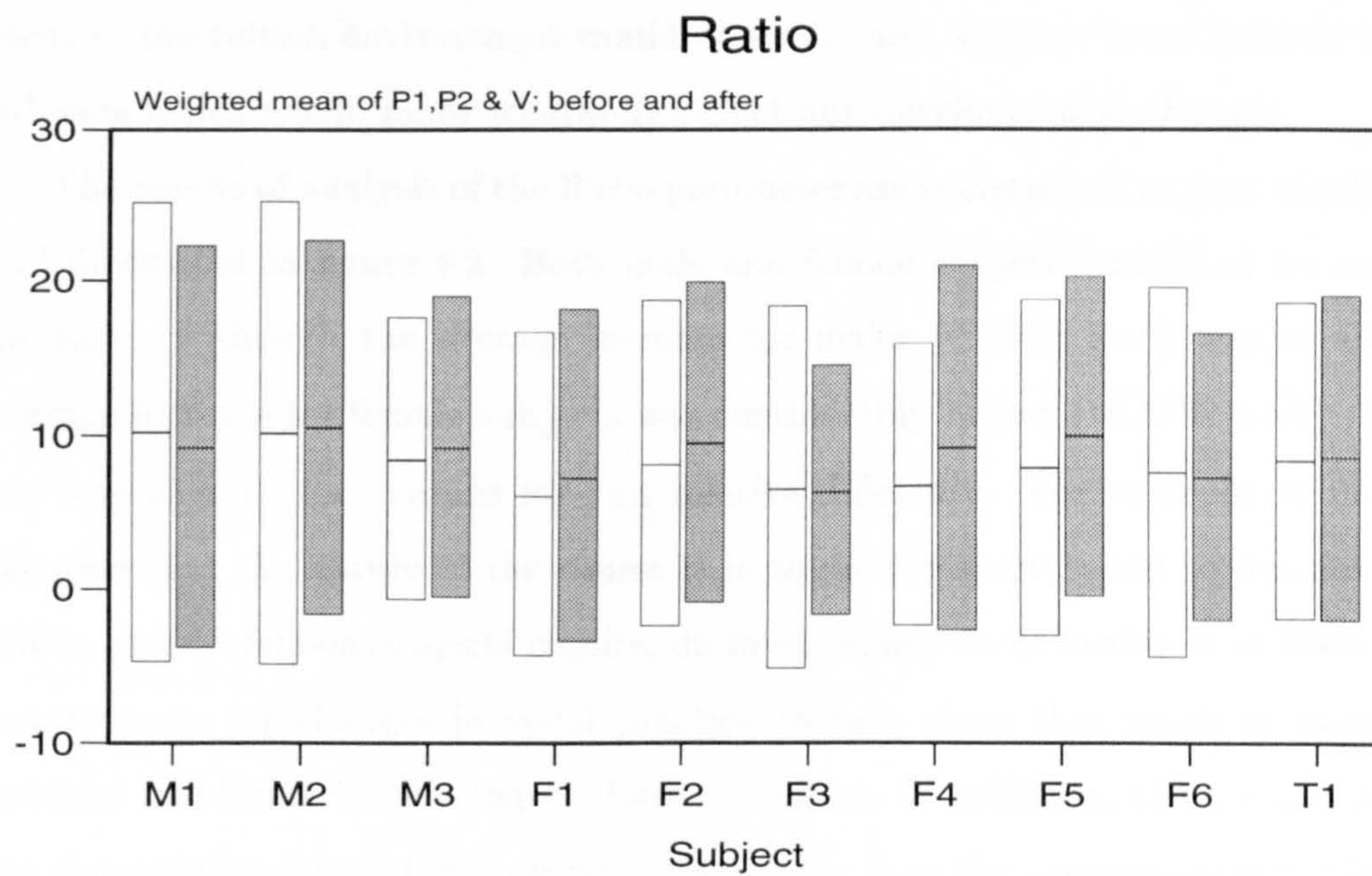


Figure 4.2: Ratio mean and standard deviation measures for all subjects plus tutor. Data assessed prior to training is shown by white bars. Data assessed after training is shown by grey bars. Key: P1=passage 1, P2=passage 2, V=vowel sequence

the 'singer's formant' has been shown to occur was set at approximately 2500Hz-4000Hz, and the spectral range for all dependable acoustic output was set at 0-5KHz. A mean value was obtained for each acoustic recording by summing all the Ratio data produced during the analysis of that recording and dividing by the total number of data.

Both recording sessions took place in the same environment. This was a lecture classroom similar to that used by the class for the duration of the six-week course. The similarity in appearance and acoustic characteristics of the recording environment to the tuition environment would, it was hoped, support vocal delivery in the subjects which would more accurately reflect any developmental changes.

The results of analysis of the Ratio parameter are summarised as part of table 4.2 and illustrated in figure 4.2. Both male and female subjects exhibited an average increase. Although the average increase for males (0.53%; n=3) was small, the average increase for female subjects was considerably higher (15.29%; n=6). There are several potential reasons for this relative difference. For example, it may be the case that the nature of the course is in some way more suited to females than males, or that female subjects require, on average, less vocal tuition than their male counterparts for changes in vocal practice to take place that result in increased spectral amplitude in the singer's formant region. In addition, the average age of the three male subjects (28.3 years) is much lower than the average age of the female subjects (48.17 years) and this may bias the study. It is interesting to note that these increases can be compared to negligible change in the measured parameter for the tutor. This may additionally imply that the measured increase has occurred as a result of the course of vocal expression itself, rather than as some constant change applicable to all measures.

When expressed as a percentage increase, the relative difference in Ratio appears high (for example, there is an increase of 71% for the ratio parameter derived from male subjects reading the sequence of vowels). However, it is important to remember that the actual levels are fairly small, so that any small change will result in a relatively large measure when expressed as a percentage increase. As discussed previously, differences in gender, age or measurement may go some way towards explaining this. One possible explanation for the relatively high increase for the

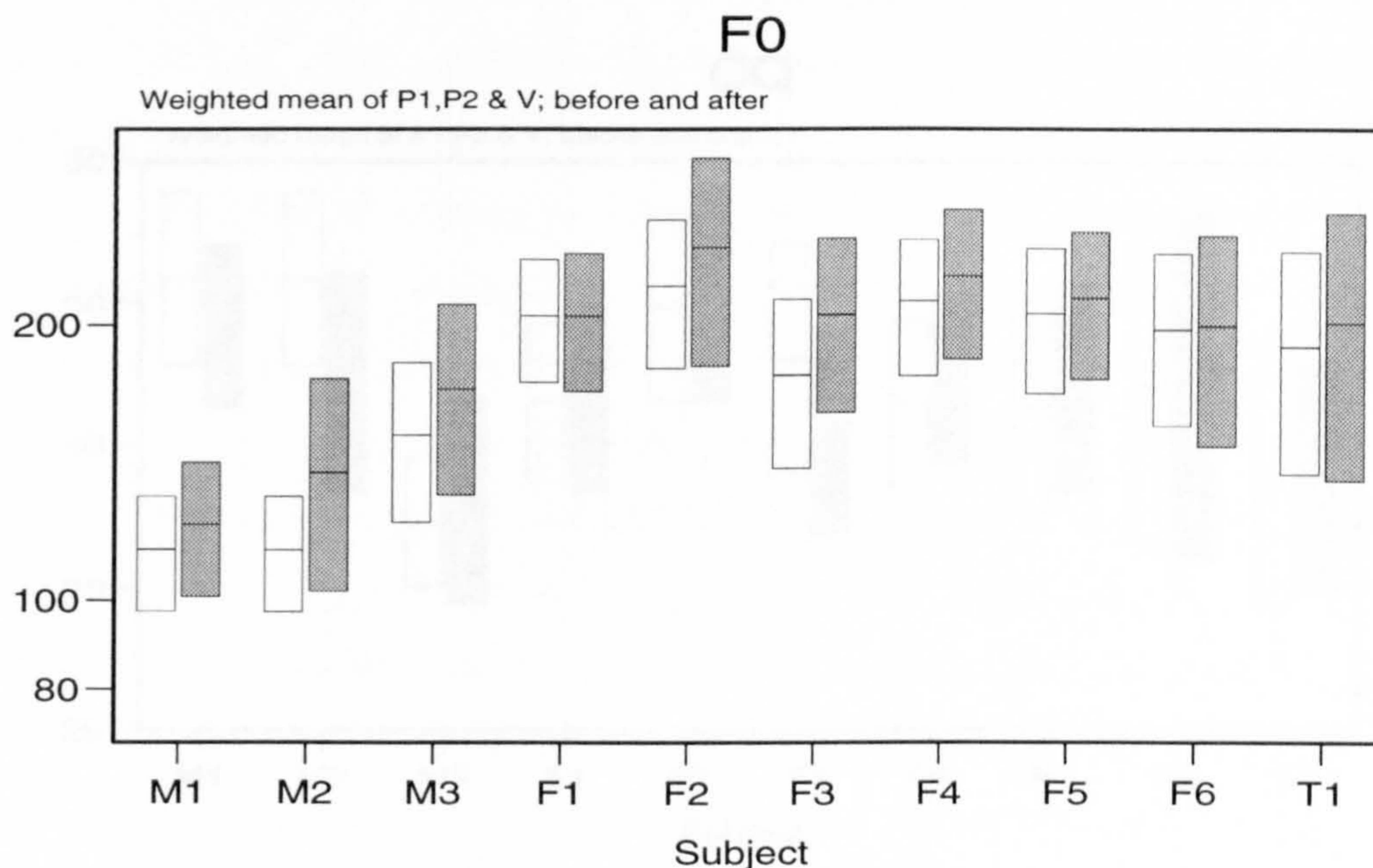


Figure 4.3: Fundamental frequency (F0) mean and standard deviation measures for all subjects plus tutor Key: P1=passage 1, P2=passage 2, V=vowel sequence

vowel sequence compared to the other passages is that the vowel sequence acted as a subconscious signal to the subject to more fully employ the mode of voice articulation taught on the course, during which extended pronunciation of some vowels was used as the basis for several exercises. An alternative hypothesis is that the same vocal mode was consistently employed by subjects, but that the measurement of generally higher level of increase for the vowel passage is largely due to the relatively greater proportion of time for subject vowel phonation relative to non-vowel phonation, such as consonants.

4.2.5.2 Fundamental frequency

The mean fundamental frequency (F0) of the three vocal tasks were calculated from the electrolaryngograph signal using a procedure to be described in section 5.9.2.

The results concerning the average fundamental frequency of the subjects are summarised as part of table 4.2 and illustrated in figure 4.3. The measure of funda-

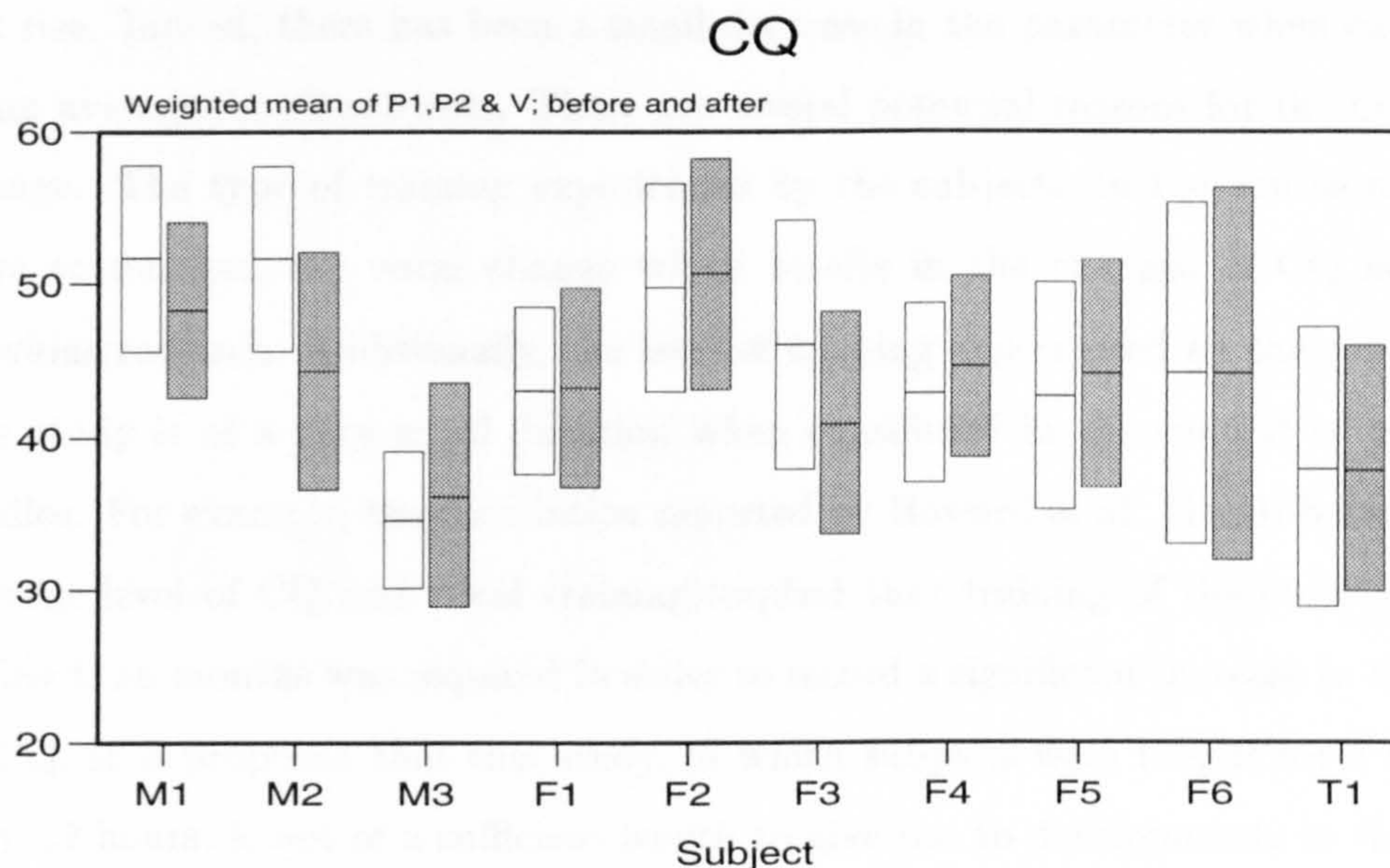


Figure 4.4: Closed quotient (CQ) mean and standard deviation measures for all subjects plus tutor Key: P1=passage 1, P2=passage 2, V=vowel sequence

mental frequency rises for all subjects, with the exception of one woman (labelled F1) for whom it remained constant. It is conjectured that the process of vocal tuition, specifically the work on pitch optimisation (section 4 of Appendix A), has exercised and subsequently tightened the vocal muscles, resulting in a corresponding increase in pitch. It is also interesting to note that this may be applicable to the course tutor, in which a measureable increase in fundamental frequency has also been recorded.

4.2.5.3 Laryngeal closed quotient

As discussed in chapter 3, a measure of the time for which the vocal folds are in contact relative to the time for which they are apart, called the larynx closed quotient (CQ), has been shown to be an important measure in assessing vocal efficiency. This has been derived from the electrolaryngograph signal (using a technique to be described in section 5.9).

Results of the analysis of the level of CQ are summarised as part of table 4.2

and illustrated in figure 4.4. In general, it appears that the average level of CQ does not rise. Indeed, there has been a small decrease in the parameter when expressed as an average for all subjects. There are several potential reasons for the negligible change. The type of training experienced by the subjects on the course may not have encouraged the vocal change which results in the changes in CQ noted in previous research. Additionally, the level of training experienced by the subjects in this study is of a very small duration when considered in the context of previous studies. For example, the correlation reported by Howard et al. (1990) between the average level of CQ and vocal training implied that training of the order of years rather than months was required in order to record a significant increase in the level of CQ. It is proposed that this study, in which subjects were taught for a total of only 12 hours, is not of a sufficient length to give rise to developments in the vocal mechanism which would result in a clear increase in CQ.

Although it has previously been demonstrated that there are effects of changes in CQ on spectral output (section 2.7.2) the results from this study do not appear to support this. For example, a large increase of 71.9% in the measure of spectral output in the Ratio parameter was previously noted for male subjects during phonation of the vowel passage. However, a corresponding decrease in mean CQ of approximately 6.6% was observed. This supports the suggestion that the short duration of the course is insufficient to produce clearly measurable change in some vocal parameters.

4.2.5.4 The rate of closing of the vocal folds

As discussed in chapter 3, a consequence of vocal training may be manifested in the rate of closing of the vocal folds, possibly in a manner which is to some extent independent of the CQ measure. In order to investigate this, the value of the maximum time differential was derived, (using a technique to be discussed in section 5.9.4).

The results are summarised as part of table 4.2 and presented graphically in figure 4.5. In considering the mean analysis of the data it appears that there was an overall increase in this parameter. Expressed as a percentage, the mean increases evaluate at 3.4% for male subjects, and 54.3% for female subjects. However, consid-

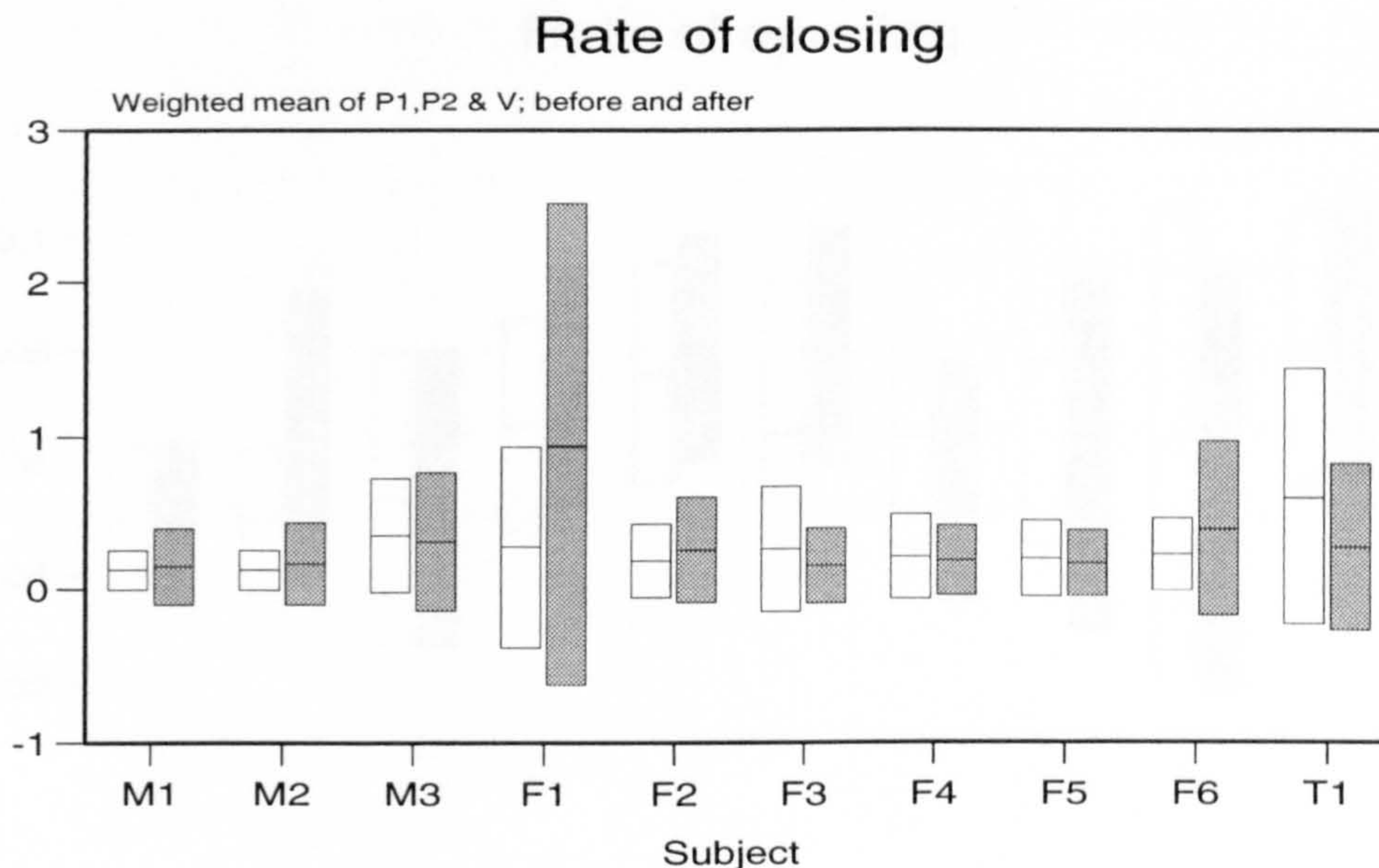


Figure 4.5: Rate of closing mean and standard deviation measures for all subjects plus tutor Key: P1=passage 1, P2=passage 2, V=vowel sequence

erable care needs to be made in interpreting this parameter. Previous research into changes in the objectively measured rates of opening and closing of the vocal folds which is relevant in this context have not been encountered. It cannot be safely concluded that the increase is a product of the training process.

4.2.5.5 The rate of opening of the vocal folds

A measure of the rate of opening of the vocal folds was also derived using the technique to be described in section 5.9.

The results are summarised as part of table 4.2 and presented graphically in figure 4.6. A first observation is that this parameter is considerably lower than that derived for measuring the rate of closure of the larynx. This concurs with previous research in which it has been well established that for the non-dysfunctional voice the rate of closure of the vocal folds is faster than the rate of opening (see section 5.9).

On average there is an increase in the parameter after vocal training for male

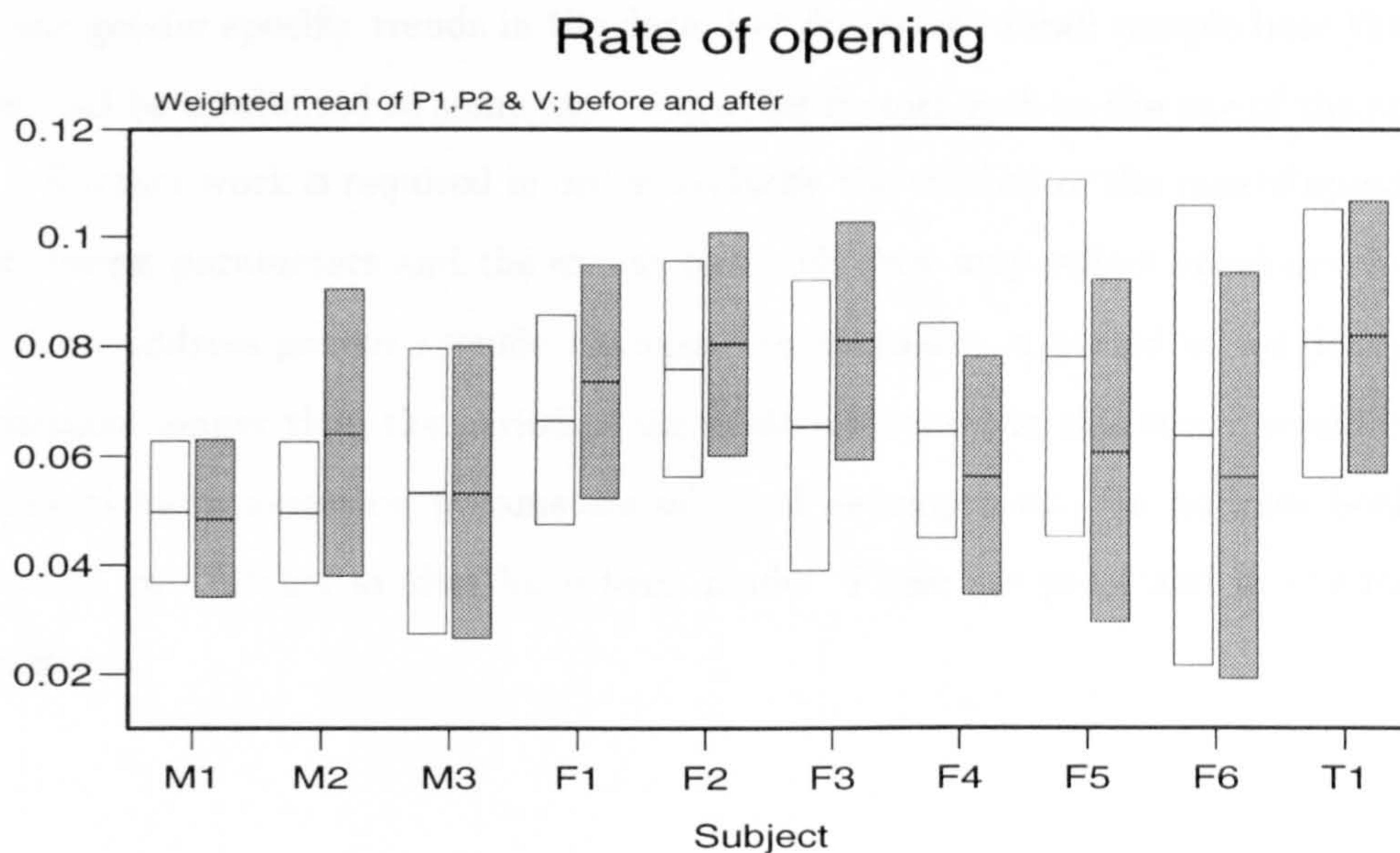


Figure 4.6: Rate of opening mean and standard deviation measures for all subjects plus tutor Key: P1=passage 1, P2=passage 2, V=vowel sequence

subjects. This has been evaluated at approximately 8.5%. However, there is no comparable pattern of increase for female subjects. In terms of electrolaryngographically derived data, differences between the pattern of vocal fold action have previously been determined. For example, in chapter 3 it was shown that the pattern of CQ across F0 is different for males and females. It may be that the measure of the rates of opening and closing is also gender-specific in some way. However, further research would be required to substantiate this.

4.2.6 Conclusions

The voice source and acoustic signals of three male and six female subjects were recorded before and after taking part in a six week course of vocal expression. A number of parameters have been assessed for changes. Clear increases have been observed for the level of spectral amplitude in the spectral region known as the singer's formant and for the average fundamental frequency of the voice. Changes in the average level of CQ have been observed, and increases in the level of opening

and of closing of the vocal folds have been cautiously interpreted. There may be some gender specific trends in the data, but due to the small sample base these may instead be attributed to some extent to other factors such as the age of the subjects.

Further work is required in order to clarify the validity of the rate of opening and of closing parameters and the extent to which they may reflect vocal development, and to address gender specific changes. Furthermore, a period of longitudinal assessment longer than the period of six weeks addressed in this study would be more appropriate in assessing parameters of vocal development. To address both these points, two further studies have been made. These are presented in the following sections.

4.3 An improvement to acoustic processing

Although the method adopted for acoustic analysis in the previous longitudinal study is regarded as sufficient to provide an indication of longitudinal changes in acoustic parameters, an improvement would be to ensure the acoustic signal was processed only when voiced phonation was detected. In this way the acoustic signal would not be processed during voiceless activity or periods of silence.

A refinement to the algorithmic process was therefore devised as follows. Each buffer of electrolaryngograph signal data is analysed for an indication of voiced phonation. If this is detected, the corresponding buffer of acoustic signal data is processed to derive parameters based on acoustic analysis. If voicing activity is not detected, the buffer of acoustic data is not processed. This is illustrated in figure 4.7.

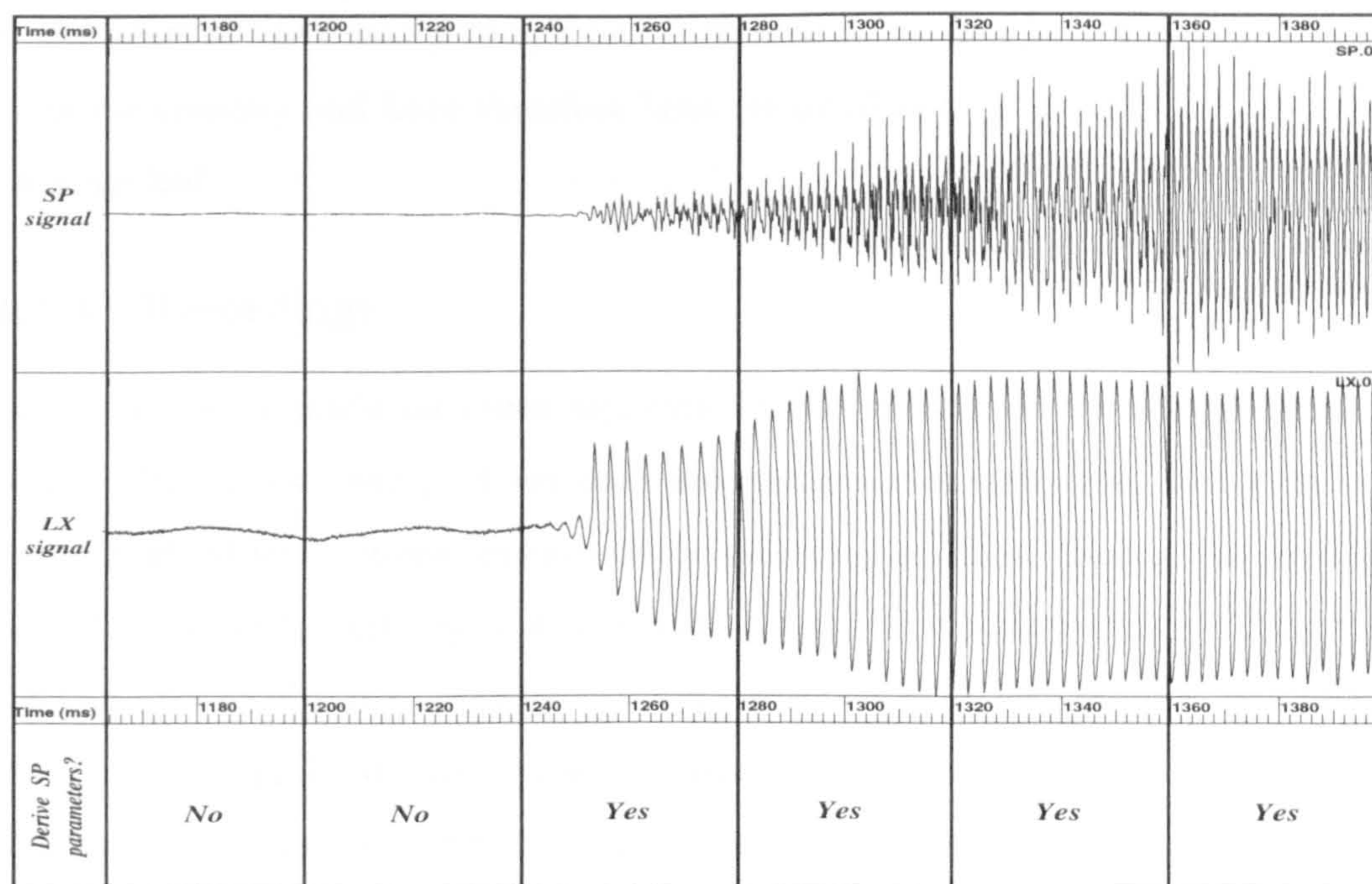


Figure 4.7: This figure shows example time-synchronised acoustic and electrolaryngograph signals (labelled ‘SP’ and ‘LX’ respectively) with vertical bars used to indicate buffer start and end times. Processing of a buffer of acoustic data only takes place if voiced phonation is detected in the corresponding buffer of electrolaryngograph data. In this example a buffer update rate of 25Hz is assumed.

4.4 Study 2 *Actors over an 8 month period*

4.4.1 Introduction

This section presents an assessment of the vocal development of a group of actors in training over an eight month period.

4.4.2 Subjects

The subject set comprised of six men (age range 20-30; mean=22.67; sd=3.54) and four women (age range 20-36; mean=25; sd=6.56). Subjects were recorded on location at the Acting School at which they were working on the second year of their diploma. Tuition is provided in a mixture of practical skills including singing, ballet and several styles of acting performance. Several plays were staged during each term length of approximately 10 weeks. Subjects board at the school, which is situated in the heart of the country away from towns and cities. Subjects undergo the same intensive training and have therefore been regarded as a set to which stimulus has been applied.

4.4.3 Recordings

Recordings were made on seven separate occasions at intervals of approximately four weeks. They were paid for each recording session attended. Subjects were given identical instructions for each of the recording sessions. During the recording procedure subjects were required to remain at an approximate distance of 0.3m from the microphone. Prior to reading the text items, subjects were required to phonate at a constant amplitude for a period of several seconds. During this phonation, an SPL meter was placed next to the microphone and the level was assessed. This test was recorded, and the level noted for future calibration of the subject output level. Subjects were then asked to phonate in five different vocal tasks. Subjects were asked to read the items aloud in a manner they felt appropriate:

1. a sequence of phonetically balanced spoken text (*80 seconds*)
2. a recited passage read in a normally projected voice (*35 seconds*)
3. item 3, read in a voice that is the opposite of maximum projection (*35 seconds*)

4. item 3, read in a voice with maximum projection (*35 seconds*)
5. 'Baa-baa black sheep' nursery rhyme sung verse (*20 seconds*)

Approximate times taken to read the items are given in brackets. This procedure is reproduced in full in Appendix C. The textual content of items 2, 3, and 4 are identical and were selected by the subject prior to the first recording session. They were recited from memory. Subjects were not given a pitch reference for item 5, and were instructed to sing the verse in a comfortable range and manner they felt appropriate. This selection of vocal tasks enables assessment of subjects for their speaking and singing voices, and for recitation at three different modes of projection. In combination, they provide a portfolio of vocal performance which may be assessed to provide an indication of vocal development as a consequence of training.

Stereo recordings were made using a Sony TCD-D10-PRO DAT recorder at a sampling frequency of 48KHz. The acoustic output of the subjects was monitored via a Sennheiser MKH/40/P48 cardioid microphone on one channel, and the output from an electrolaryngograph was recorded on the other channel to monitor the area of vocal fold contact.

Voice source and acoustic output signals were recorded and analysed using an early version of the ALBERT software to be described in chapter 5.

Recording sessions took place in similar, although not identical, environments. These were lecture classrooms similar to those used by the subjects for tuition and performance during their course. The similarity in appearance and acoustic characteristics of the recording environment to the tuition environment would, it was hoped, support vocal delivery in the subjects which would more accurately reflect any developmental changes.

4.4.4 Accumulated data

Some subjects were not present for all recording sessions. Technical failure was also responsible for the absence of SPL measures for one female subject in the sixth recording session, and two female and one male subject in the seventh session. A summary of the accumulated data is shown in figure 4.8. In order to process the recorded data, the difference from one session to the next was first derived for each

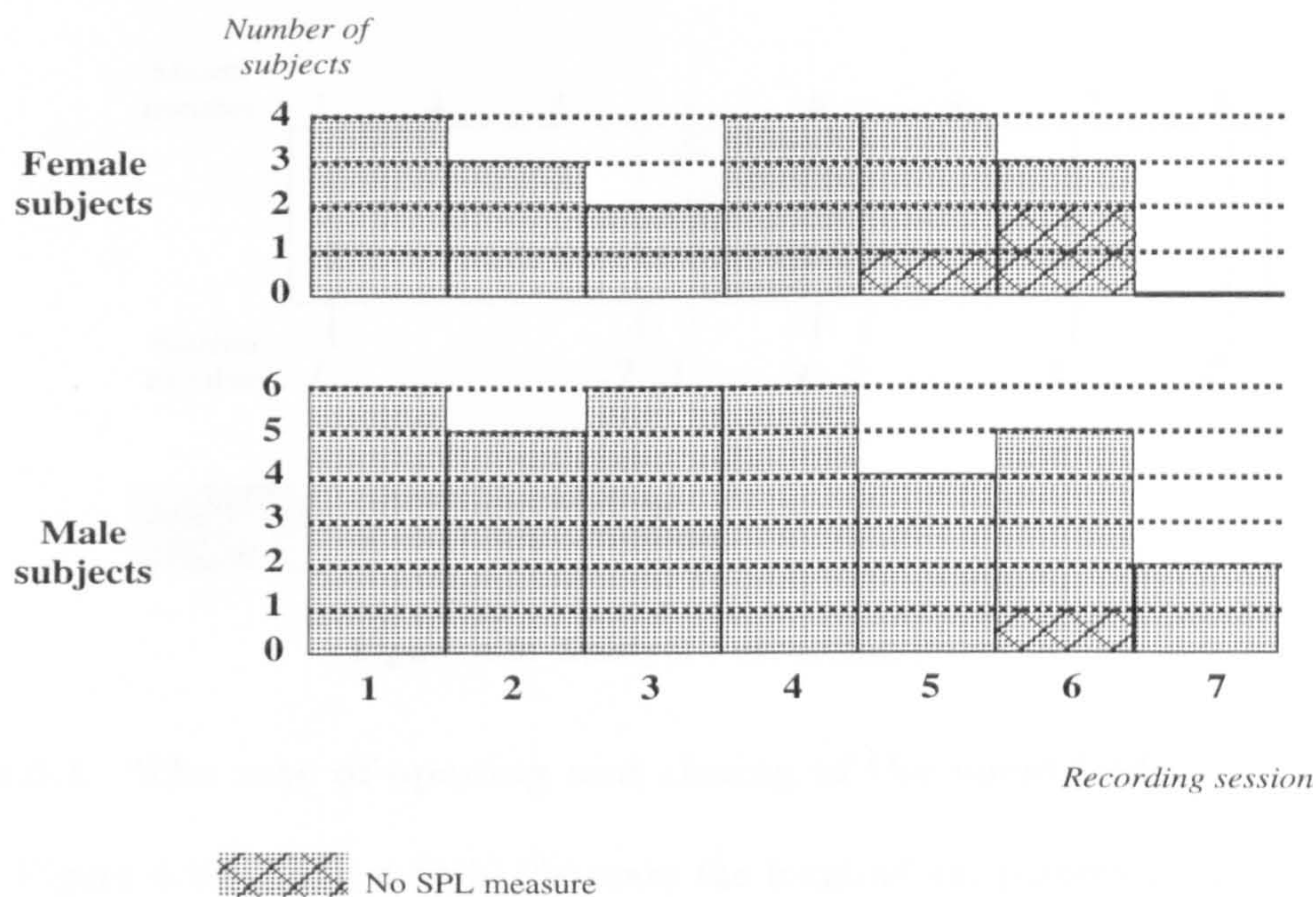


Figure 4.8: Study 2 - data recorded and analysed

parameter. If a gap was present in the data, it was replaced with a value which was a linear interpolation of the value derived from the recording sessions immediately preceding and succeeding the session for which data was absent. Increment values were averaged for all subjects to create two patterns of change for male and female subjects. These were then successively applied to the means from the first recording session, at which all subjects were present, to create a sequence of values representative of mean trends in the data across the sequence of recording sessions.

4.4.5 Results

The results for the five vocal tasks were averaged to produce a measure which in addition to a consideration of the data trends in isolation may help in an assessment of the data. Male and female subjects exhibited several contrasting trends and will be considered firstly in combination, and then in isolation.

Figure 4.9 illustrates the intervals at which recording sessions took place. A period of considerable trauma was experienced at the school which greatly affected all subjects during the period of the third and fourth recording sessions.

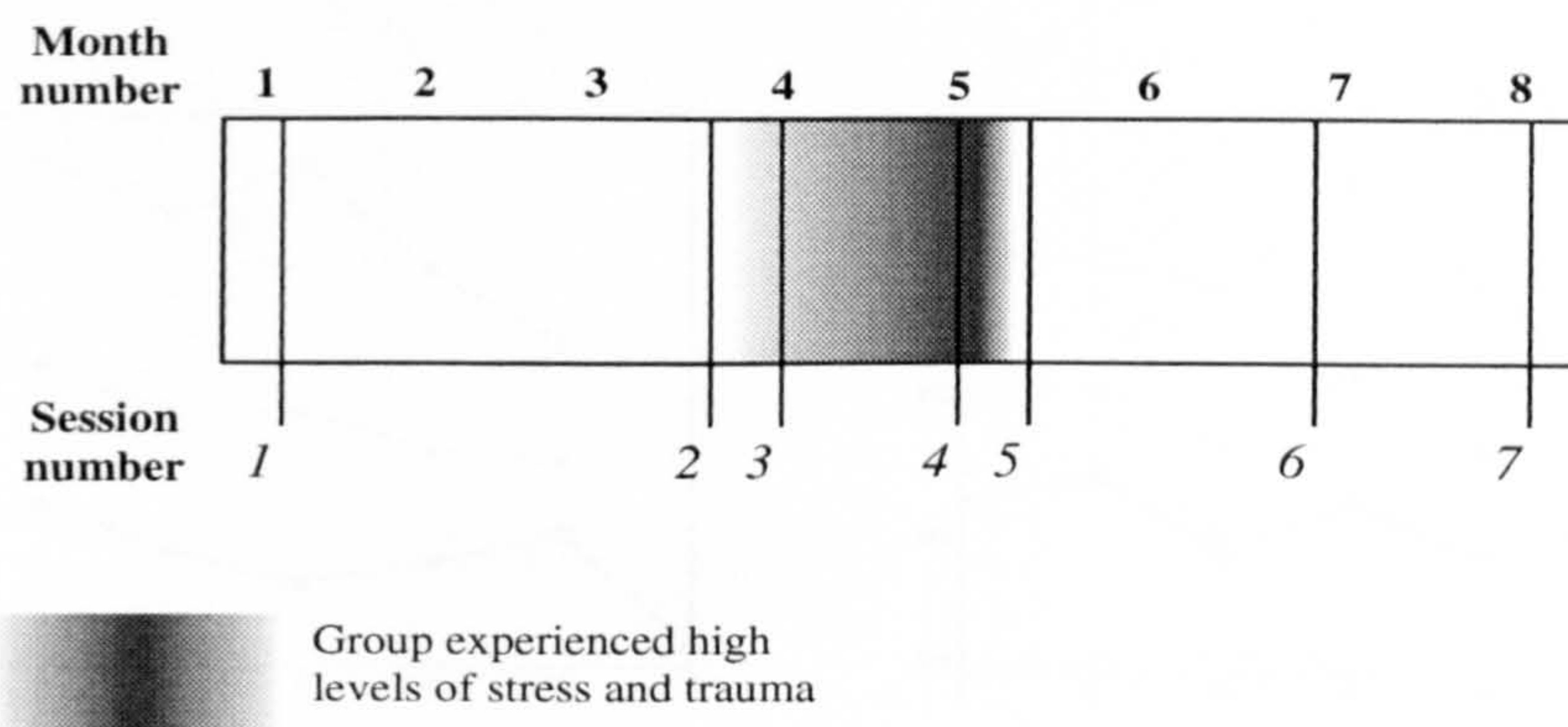


Figure 4.9: Study 2 - recording intervals

4.4.5.1 The rate of opening and closing of the vocal folds

Figure 4.10(a) and 4.10(b) illustrate the longitudinal pattern of change for the rate of closing and opening measures respectively for both genders. There appears to be a consistent general decrease in the rate of closing for female subjects, with a pattern that may be interpreted as a relatively smaller downward trend for male subjects.

For female subjects, the general decrease in rate of closing is clearly visible during examination of the five vocal tasks in isolation (figure 4.10(a)), although there are some positive fluctuations in the measure at sessions 2 and 4. Analysis of the rate of closing for each of the five tasks phonated by males does not produce an indication of clear trends in the data. In this instance it appears that all voice tasks exhibit slightly descending values followed by temporary upwards fluctuations at the third and seventh sessions, but that voice tasks exhibiting relatively higher levels of rates of closing (spoken text, maximum projection and singing) in general exhibit less stability. Both genders exhibit similarities, including a local peak at session 2 and a generally decreasing trend in the data, but the rate of descent across training for females is greater and exhibits local peaks at session 4 which is absent for males.

The average for both genders combined exhibits a general decrease in the measure. However, results for the rate of opening of the vocal folds are not distinct for subjects of either gender. Examination of figure 4.11(a) indicates a generally decreasing pattern for the minimum and maximum projection tasks and spoken text

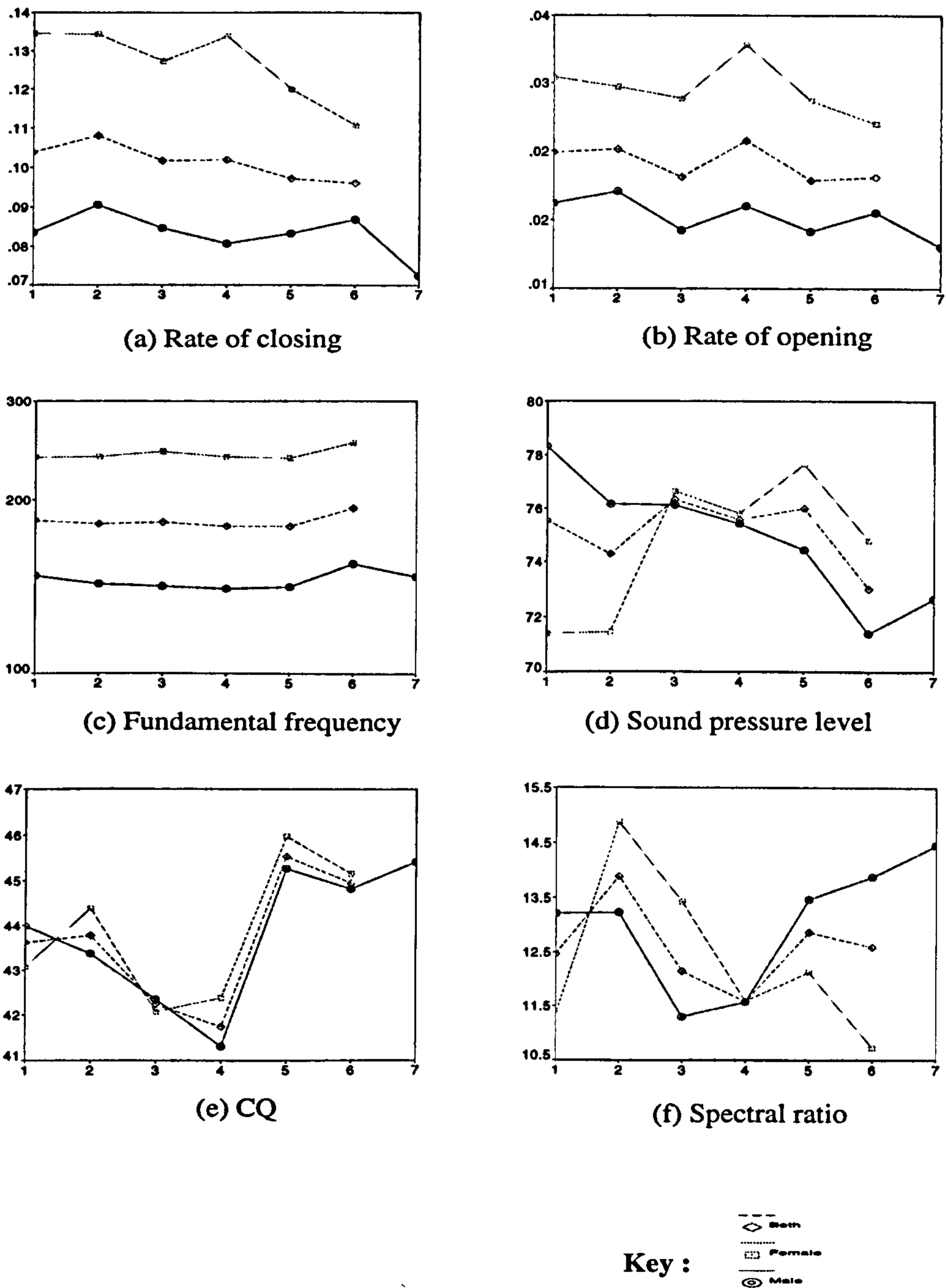


Figure 4.10: Results for each of the six parameters for male, female and both genders combined. The x axis represents the number of the recording and may be regarded as an approximate indicator of time and training.

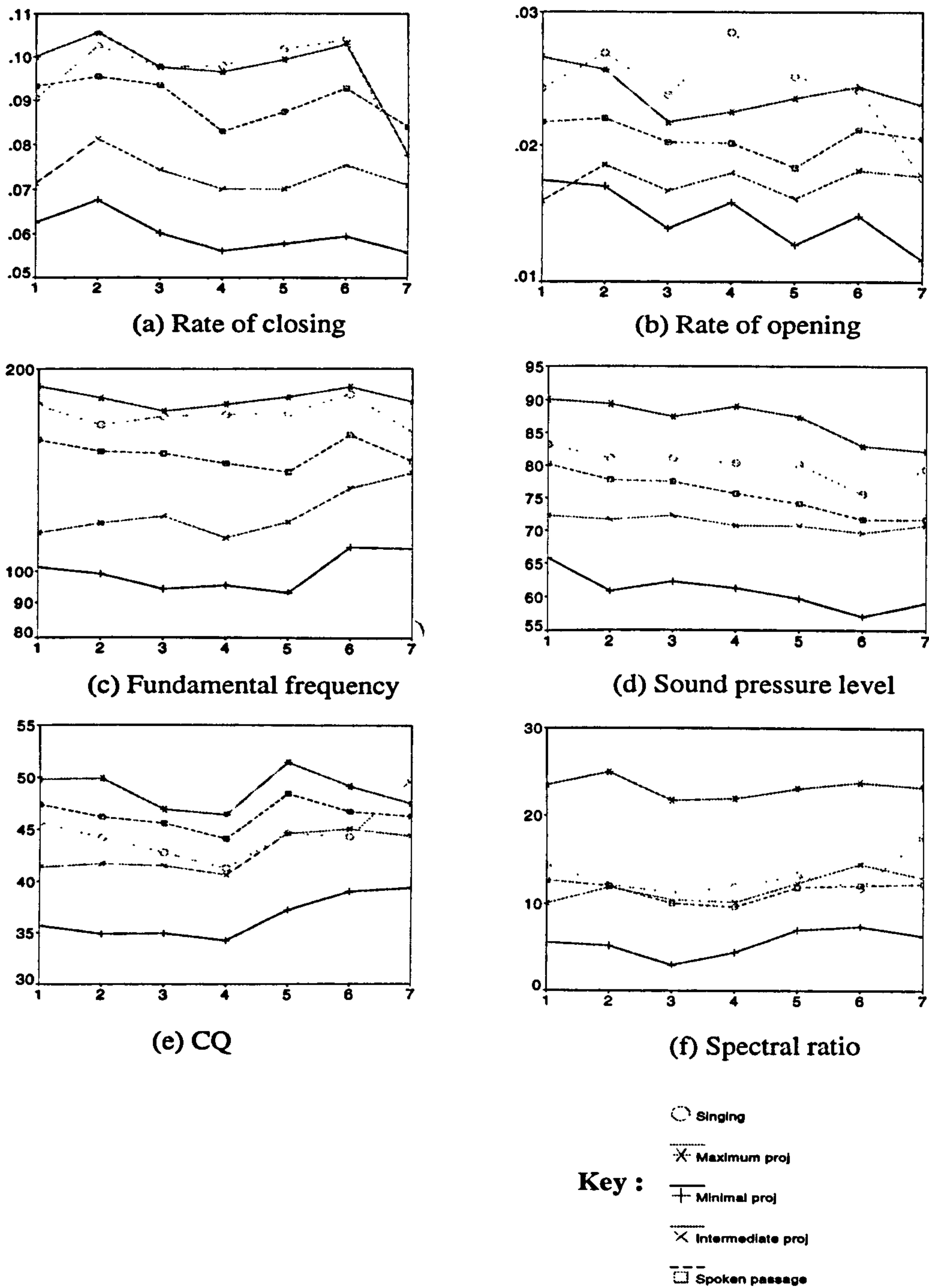


Figure 4.11: Results for each of the six parameters for male subjects. The x axis represents the number of the recording and may be regarded as an approximate indicator of time and training.

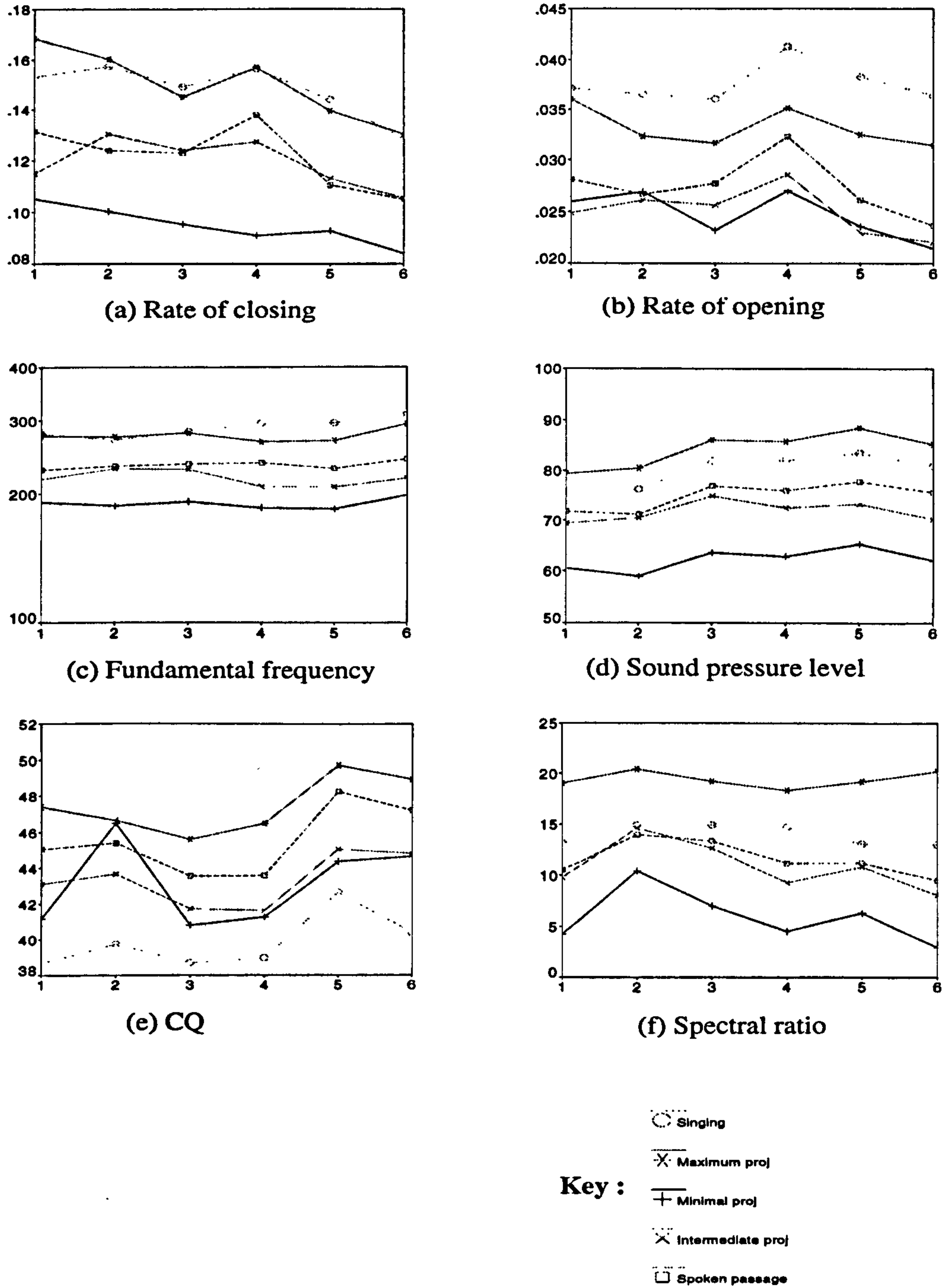


Figure 4.12: Results for each of the six parameters for female subjects. The x axis represents the number of the recording and may be regarded as an approximate indicator of time and training.

of male subjects, with less clear patterns of decrease for the other two assessment tasks. Correlation analysis indicates a small decrease assessed over all sessions for subjects of both genders, but of a small gradient. There are similarities between male and female subjects as for the rate of closing of the vocal folds, but to less of an extent, with the local peak at session 4 of relatively greater magnitude.

Research concerning objectively measured rates of opening and closing of the vocal folds derived from an electrolaryngograph do not appear to be large in quantity. As for study 1, this implies that conclusions concerning these parameters must be carefully drawn. Titze (1990) developed a model of the vocal folds from which he derived a number of physiological changes that would affect the two measurement techniques developed for this study, including increased convergence in the glottis and a change in phase between the upper and lower portions of the vocal folds. However, these were theoretically based and it is considered unlikely that they may be expressed as a consequence of the tuition received by the non-dysfunctional voice users assessed in this study. It has been previously determined that there is considerable variability of the voice source signal across different subjects. For example, the opening phase of the electrolaryngograph signal often exhibits a 'chink' or 'knee' in the signal (i.e. Anastaplo & Karnell, 1988). This apparently occurs less frequently for the male voice than for the female voice, but not to the extent that specific male-female differences can be identified (see Moore's comments in MacCurtain et al., 1981). On the basis of the accumulated data it is not considered that the parameters of the rates of opening and closing adopted for this study can be reliably used as a pointer for the state of vocal development.

4.4.5.2 Fundamental frequency

Results for the measure of fundamental frequency for subjects are shown in figure 4.10(c). In general, it appears that there are gentle fluctuations in the measure, including a distinct rise towards the end of the recording sessions.

Examination of the performance for vocal tasks in solitude indicate that, for female subjects (figure 4.12(c)), a steady climb is discernable for the sung passage (although a period of non-growth is evident for sessions 4 to 5), with all other vocal tasks exhibiting rising tendencies for sessions 1 to 3 and 5 to 6, but with a distinctive

downwards fluctuation at sessions 4 and 5. For male subjects (figure 4.11(c)), vocal tasks exhibiting lower levels of fundamental frequency (minimal and normal levels of projection) may be regarded as exhibiting a clear rise in frequency with downwards fluctuations during the period of trauma (sessions 4 and 5). For the remaining three vocal tasks, the pattern of change is less clear, with some tendencies in common with that of minimum and intermediate projection discussed earlier. Items exhibit a general rise apart from the final session, which it should be remembered is based upon parametric change in a relatively low level of subjects ($n=2$; see figure 4.8).

With regard to the mean values, the pattern of change appears consistent for subjects of both gender. It may be conjectured that the initial drops in fundamental frequency correspond to excitement experienced by the subjects at the prospect of being involved as subjects in a scientific experiment expressed for the first recording session as an artificially inflated F0 value, complimented by the prospect of earning money in an environment which by its solitude does not otherwise provide such opportunity. There appears to be a gentle climb of F0 prior to sessions 3 and 4, during which subjects experienced considerable stress, which commences again during sessions 5 and 6. As discussed in section 3.5, correlations between mood and F0 have been noted in previous research although there appears to be no general agreement on the effect on fundamental frequency as a consequence of depression and related stress (see, for example, the overview in Nilsson & Sundberg, 1985). It may be that the downwards fluctuations in F0 towards sessions 3 and 4 reflect a generally more depressed mood amongst the subjects.

In study 1 a statistically significant increase in F0 was observed in subjects after attendance of a six week course of vocal expression. As considered in the conclusions for that study, such a change would imply that the process of vocal tuition has exercised and subsequently tightened the vocal muscles, resulting in a corresponding increase in pitch. This may be expressed as a function of longitudinal development for many voice subjects, and the results presented here may be very tentatively interpreted as supporting such a proposal.

4.4.5.3 Sound pressure level (SPL)

Results for the mean sound pressure level (SPL) values are shown in figure 4.10(d). This measure appears to exhibit contrasting patterns of change for male and female subjects. Male subjects exhibit a consistent downward trend, whereas female subjects appear to exhibit a general increase, although with less consistency. For both genders, there is a very strong consistency in the pattern of SPL data across the five vocal tasks (figures 4.11(d) and 4.12(d)). As discussed previously (section 3.3), previous research indicates that trained vocalists use significantly greater mean intensity levels than untrained vocalists, which implies that a generally increasing trend in SPL may occur as a function of vocal training. The SPL data for female subjects may be tentatively interpreted as supporting this proposal. However, this clearly does not appear to apply to male subjects. It may be that the contrasts between the two genders indicate differing modes of development for male and female subjects.

4.4.5.4 Laryngeal closed quotient and Ratio

Results for the measure of CQ and Ratio are shown in figures 4.10(e) and 4.10(f) respectively. For CQ, male and female subjects exhibit almost identical patterns of change across time. After small changes during the first two sessions, the pattern of CQ exhibits a relative dip in value for sessions 3 and 4 followed by a large increase. Examination of the data for each of the five tasks individually indicates that for male subjects, the same pattern of change applies to each of the tasks, with a dip noticeable for sessions 3 and 4, followed by a general increase thereafter for each of the five tasks, which is not sustained for the maximum projection and spoken text vocal tasks.

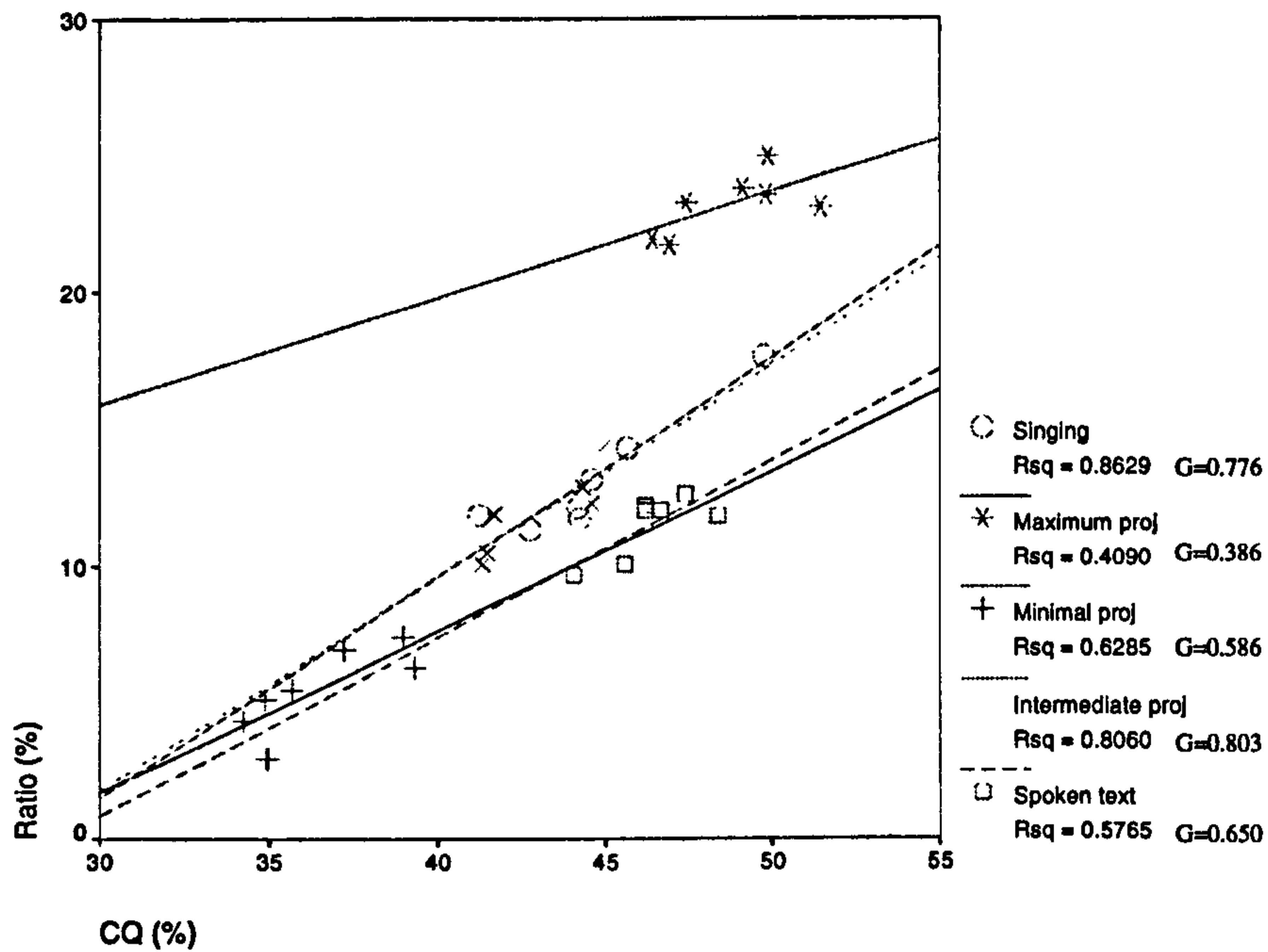
A very similar pattern of behaviour is also evident for the Ratio parameter, as shown in figure 4.11(f), including downwards fluctuations during sessions 3 and 4. The relative dip in CQ evident for speaking voice and maximum projection tasks discussed previously is also expressed in the pattern of Ratio data. Female subjects, however, exhibit a contrasting behaviour of an initial rise in Ratio followed by a decrease in the measure across successive recordings. This performance is generally

consistent across the five vocal tasks, as is evident in figures 4.10(f) and 4.12(f). The downwards fluctuation during recordings 3 and 4 observed for male subjects are also evident. In view of the extended period of psychological trauma experienced by the subjects during the period of time in which sessions 3 and 4 were recorded, it may be conjectured that the downward fluctuation in CQ and Ratio parameters is reflective of stress which has influenced vocal development.

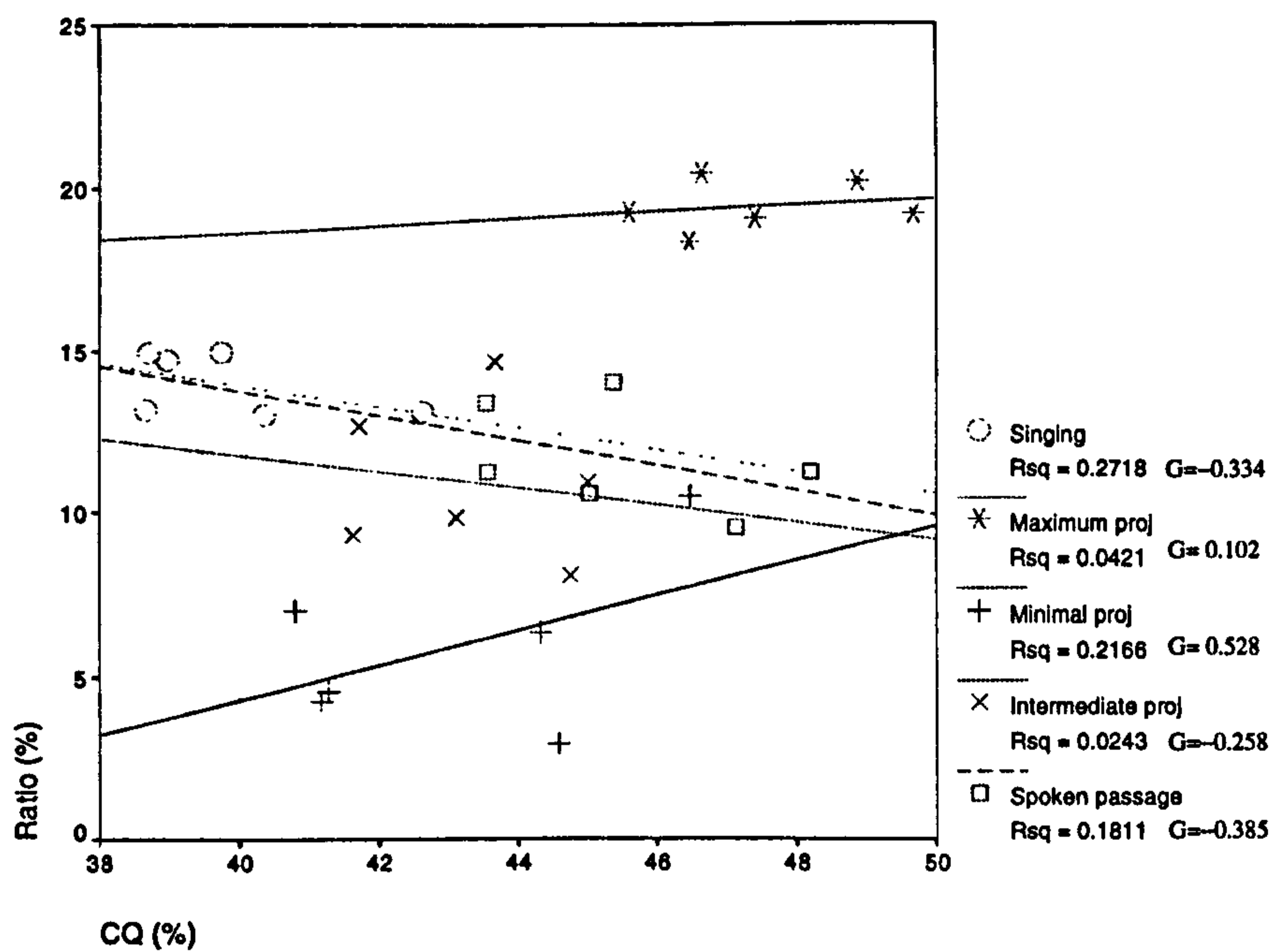
4.4.6 The relationship between laryngeal closed quotient and Ratio

In order to more fully investigate the correlation between CQ and Ratio measures, scatterplots were made of the mean CQ and Ratio values. These are shown for male and female subjects in figures 4.13(a) and 4.13(b) respectively. The linear interpolation values for each of the vocal tasks (labelled G) are listed at the right hand side of figure 4.13. A commonly used measure of the goodness of fit between two variables, the coefficient of determination, (R_{sq}), has been applied to data for each of the five assessment tasks. Male subjects express a clear, strong correlation between CQ and Ratio for each of the five vocal tasks, with the set of large R_{sq} values reflecting a high level of cohesion between the two parameters. In contrast, the data for female subjects indicates that three vocal tasks exhibit a positive correlation, and two exhibit a negative correlation. The R_{sq} values indicate a low level of cohesion between the two parameters, for all five tasks.

The implication from this data is that the acoustic consequences of increased CQ at the voice source level have resulted in increased spectral amplitude in the singer's formant region relative to the full vocal output spectrum measured. The data set for the projected mode of phonation exhibits a correlation which is of a similar gradient to that of the other four vocal tasks, but at a higher level of Ratio constant. It may be that this data is merely the extension of a correlation between the two parameters that is more exponential than linear. However, the distinct location of the data away from all other data implies that some other phenomenon is more likely to be responsible for the shift. The relatively higher volume level of acoustic output demonstrated by the subjects for this mode of phonation (figure 4.11(d)) may involve the use of a physiological mechanism which is not in use for the other four tasks, resulting in a much higher level of spectral amplitude in the



(a) Scatterplot of mean data for all five tasks individually – male subjects



(b) Scatterplot of mean data for all five tasks individually – female subjects

Figure 4.13: Correlation between CQ and spectral ratio from longitudinal data set. The G data represents the increment in Ratio for a 1% increase in CQ.

singer's formant range. For example, the lowering of the larynx and widening of the pharyngeal cavity reported by Sundberg (1974) may be evident in this mode of phonation, but not for the four other vocal tasks. Note that subjects are not trained singers, but are being trained to sing as part of a general regime intended to enhance acting skills.

4.4.7 Contrasts between the two genders

The observed differences between the two genders in SPL, CQ and Ratio parameters may be accounted for according to the following theory. Female subjects have higher fundamental frequencies than their male counterparts (figure 4.10(c)). This results in relatively less partial components in the spectral output. Consequently, the task of 'tuning' resonant frequencies in the vocal tract to match voice source excitation (whether a conscious or subconscious task) must be a harder target for female subjects to achieve than for male subjects. This may be part of the reason for the relatively more indistinct pattern of Ratio (figure 4.10(f)) derived for female subjects over the recording sessions. It is a well-known phenomenon that when loudness is increased, the higher spectrum overtones gain more in amplitude than the lower ones (Sundberg, 1987). This was discussed previously (section 2.7.3; figure 2.16). Gramming (1991) noted that, other things being equal, the higher the SPL, the higher the frequency of the first formant. Female subjects appear to be placing more emphasis on the consequences of increased SPL (figure 4.10(d)) to achieve impact during the period of this study. This is in contrast to male subjects who may be regarded as exhibiting a trend towards increased spectral amplitude in the singer's formant region (figure 4.10(f)) and who therefore may place less emphasis on the use of sound level to provide impact (as indicated in figure 4.10(d)).

Previous research considering the nature of trained singers concluded that for voices below about 300Hz most voices seek to energise formant clusters in the higher partial bandwidth 2200-5000Hz using an extended glottal source bandwidth (Watson, 1992). Furthermore, above this frequency the strategy changes to dynamic tuning of the lower formants (predominantly the first and second) to available low-order harmonics and the employment of a less wide glottal source bandwidth. Although this analysis is based on professional singers rather than developing actors, it is inter-

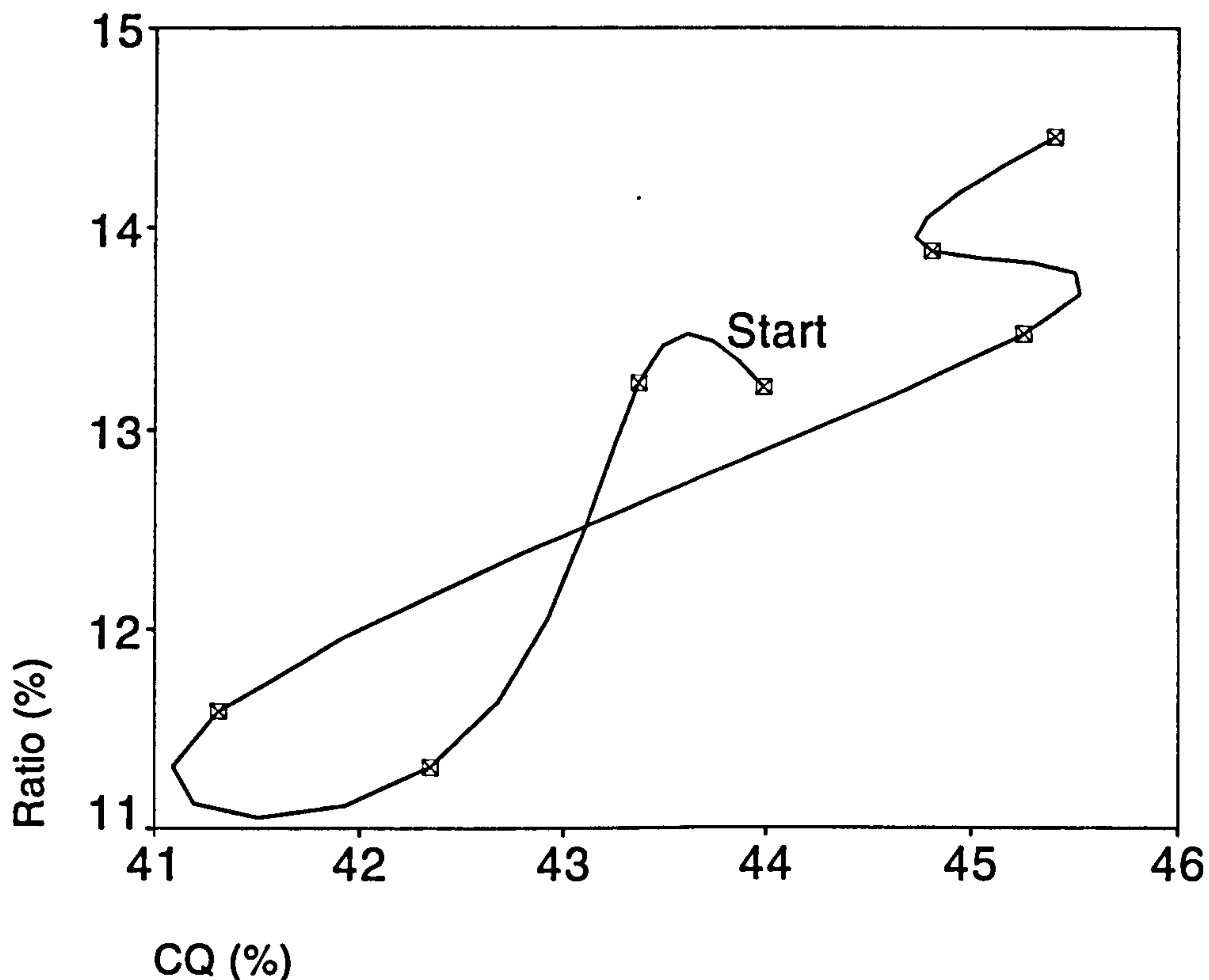


Figure 4.14: Pattern of change of mean CQ/Ratio for all male subjects, for all five assessment tasks. Spline interpolation (Hearn & Baker, 1986) is used to indicate one possible path of development between the data points. The third and fourth data points in the series correspond to a period of high trauma.

esting to note that for the two vocal tasks exhibiting high fundamental frequencies (female subjects; singing and maximum projection assessment tasks, expressing values between 270Hz and 310Hz) there appear to be indications of behaviour which is different from that of the other vocal tasks. This includes a general increase in fundamental frequency for the singing task in contrast to varied performance for the other tasks, and a relatively more stable maintenance of a level of ratio across recording sessions for both singing and maximum projection tasks in contrast to varied fluctuations for the other tasks.

A plot of mean data for all five assessment tasks expressed by male subjects was generated in order to provide a general indication of the pattern of CQ/Ratio

development across the recording sessions. This is shown in figure 4.14. Spline interpolation has been applied to illustrate one way in which the data may have developed between the recording sessions. The data may be interpreted as expressing a generally increasing trend in CQ and Ratio parameters, with a decrease evident for the period of trauma during which recordings 3 and 4 were made. It is suggested that for male subjects one goal of vocal development is to advance along a *developmental continuum*¹, one facet of which is CQ and Ratio parameters, with the caveat that the pattern of movement along this continuum may involve downwards as well as upwards fluctuations. Although female subjects may share a similar goal, this study has not produced evidence to substantiate this proposal.

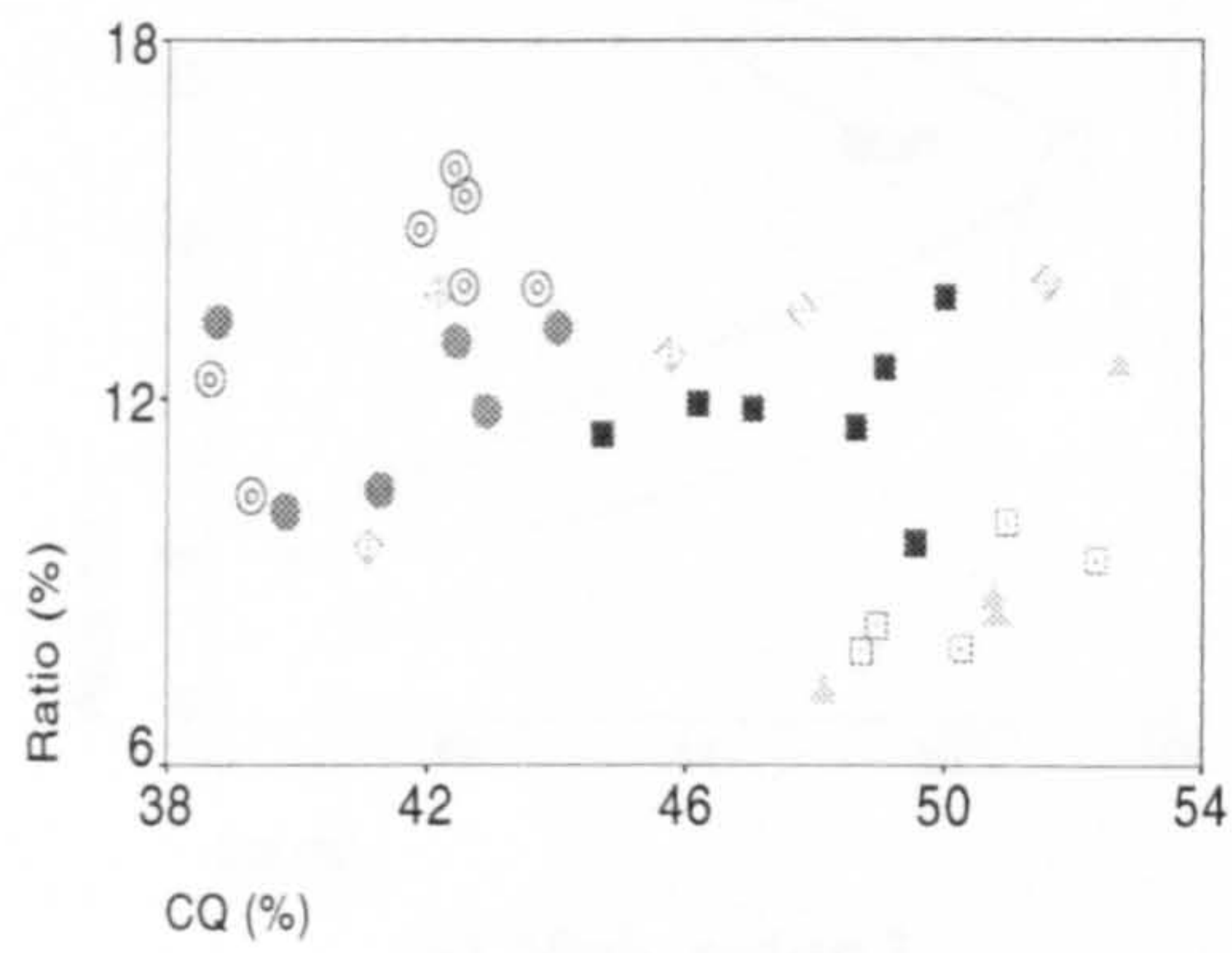
4.4.8 Individual performance

All of the preceding discussion has been based on longitudinal data derived from the means of male and female subjects. In order to more fully consider the CQ/Ratio behaviour in male subjects, profiles of the six male individuals were derived for each of the six vocal tasks (figure 4.15) and of development of CQ/Ratio across the recording sessions for which subjects were present (figure 4.16).

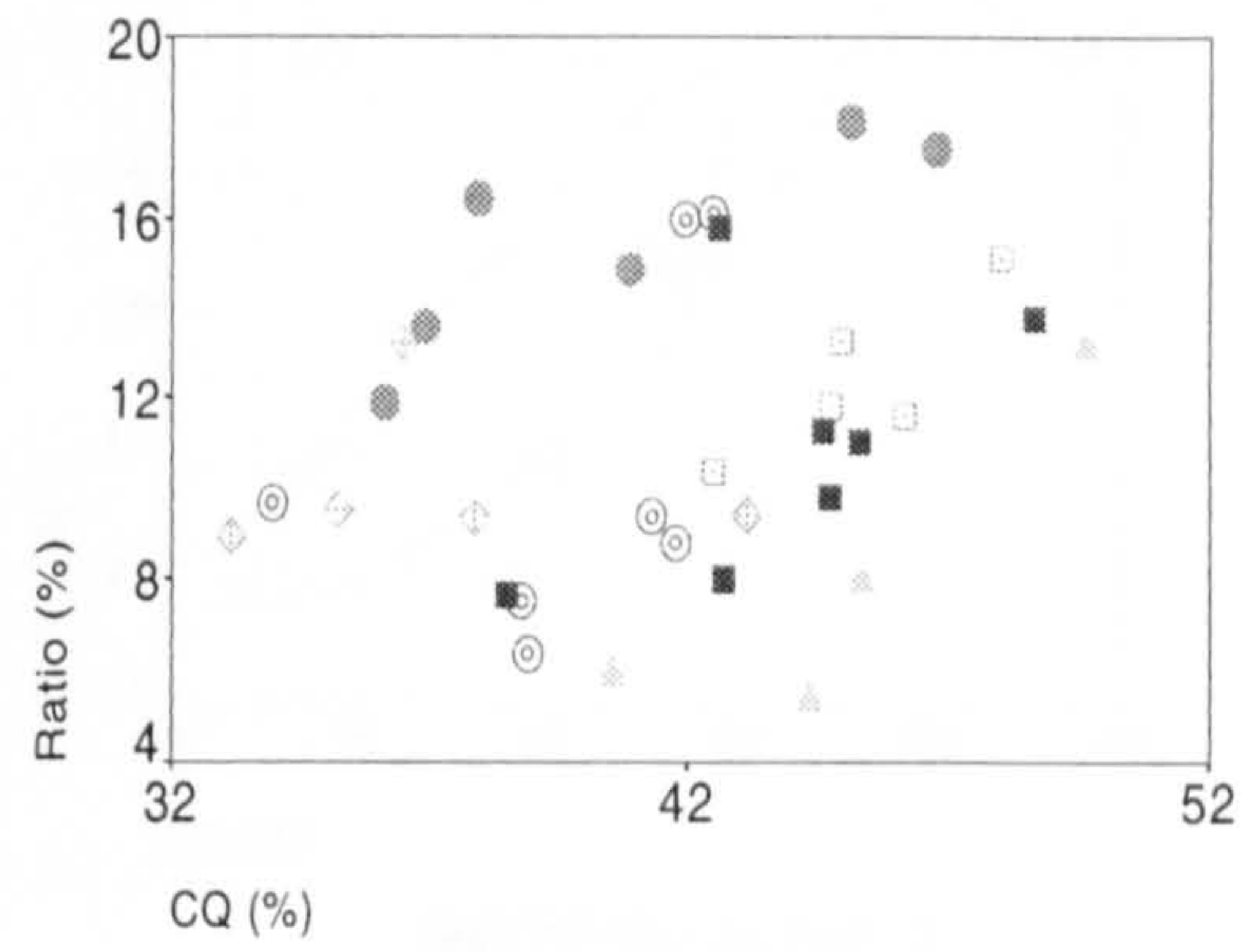
Analysis of figure 4.15 indicates that subjects exhibit individual CQ/Ratio domains. It is interesting to note that clustering is relatively more dispersed for the vocal tasks which by their nature may be likely to exhibit more varied performance. For example, the minimal projection task required voiced phonation. The exclusion of whispered speech from the performance of this task enforced a restriction on the subjects which may be likely to result in a relatively less consistent performance from one recording session to the next. This may help explain the relatively greater dispersion of CQ/Ratio data for this task (figure 4.15(c)) than for the other tasks.

Each individual exhibits a characteristic clustering of CQ/Ratio values for each of the five vocal tasks. Individual CQ and Ratio domains are apparent. For example,

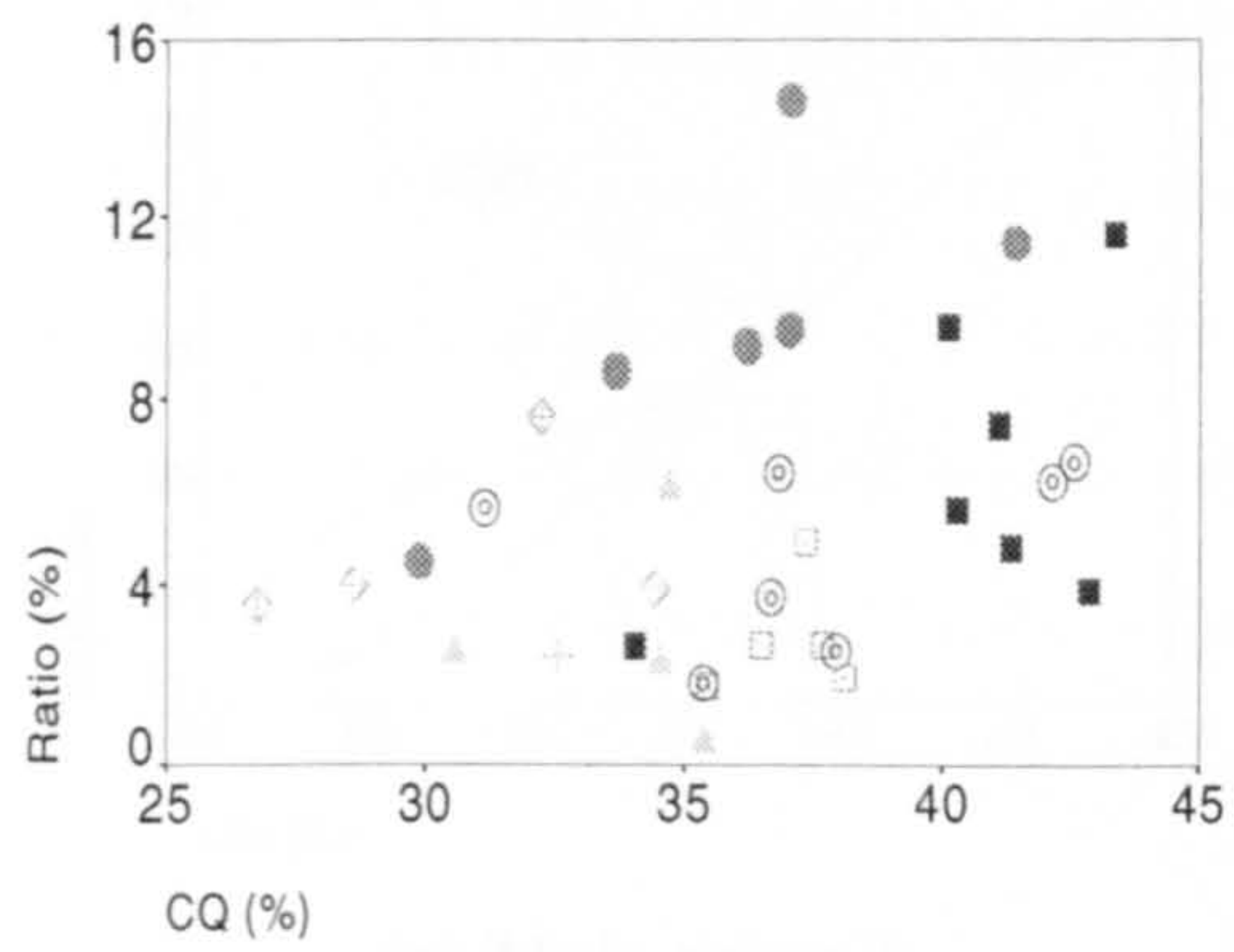
¹The concept of a continuum against which to compare states of some form of development is not in itself a new idea. Other models pertaining to some form of voice development exist, such as those involved in children's singing (Welch, 1986; Welch et al., 1991). The continuum proposed in this thesis is regarded as having application to a wide span of developing adult male voice users, with the considerable benefit that its objective base facilitates incorporation as part of an automated device intended to help the development of a user along the continuum.



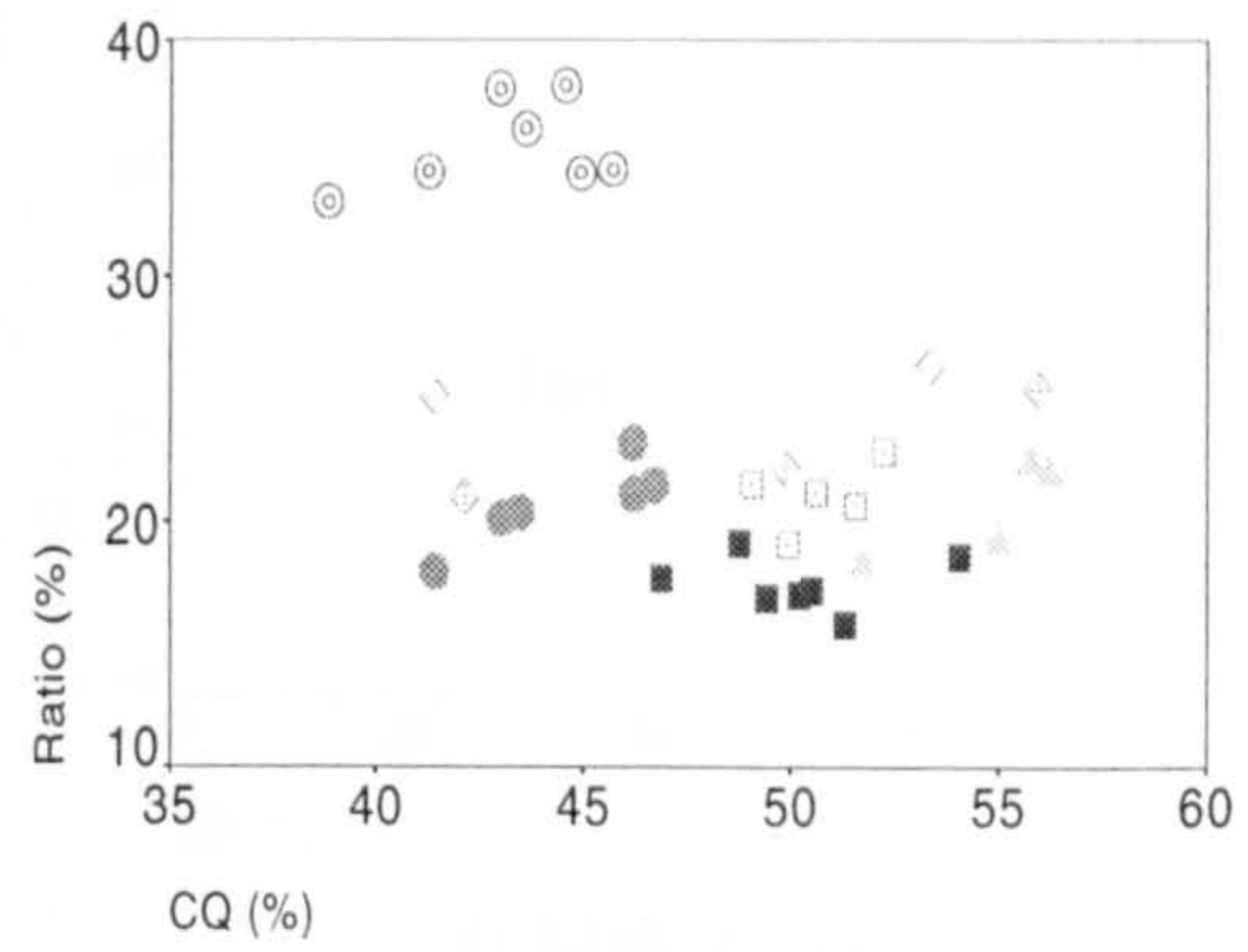
(a) Spoken text



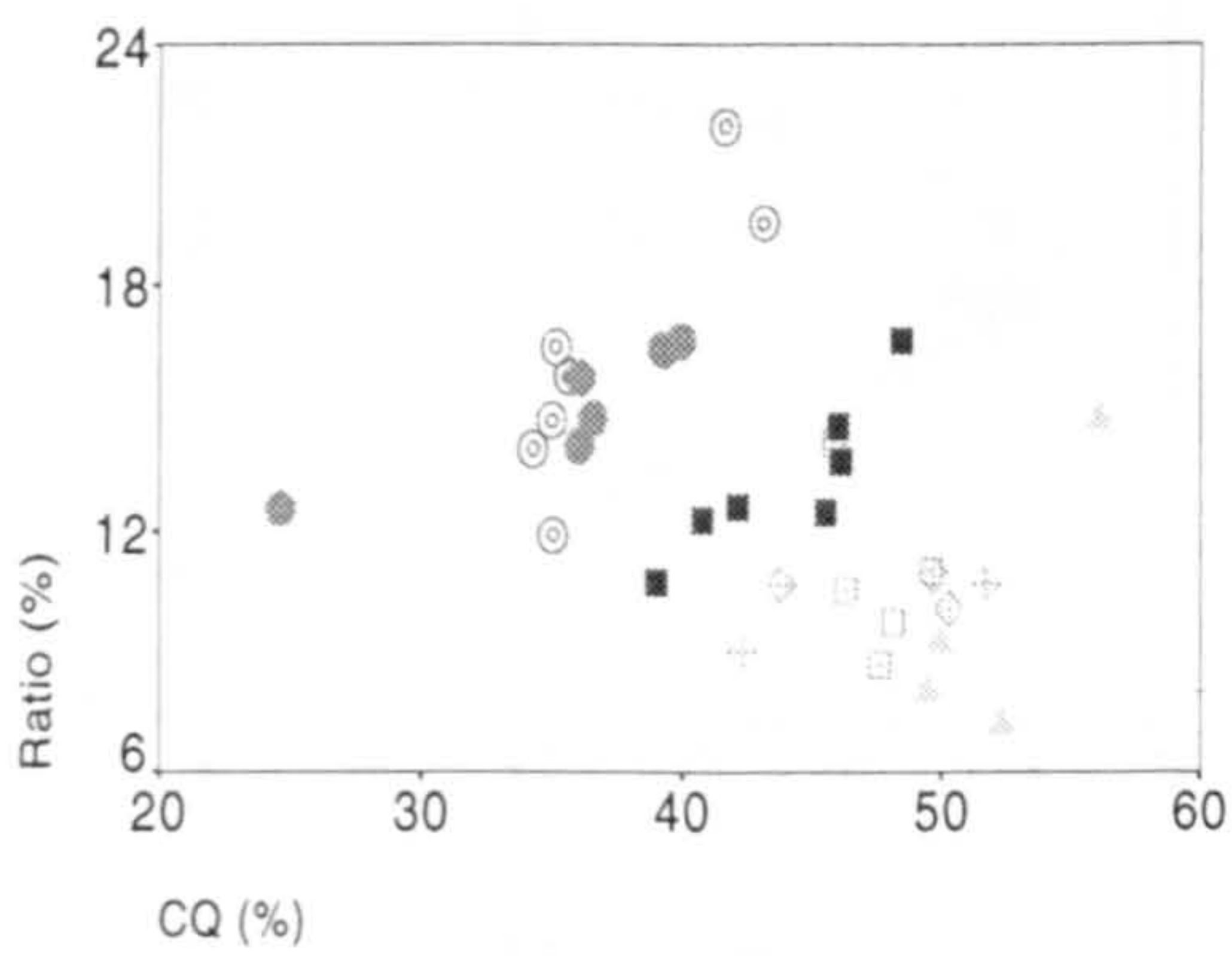
(b) Intermediate projection



(c) Minimal projection



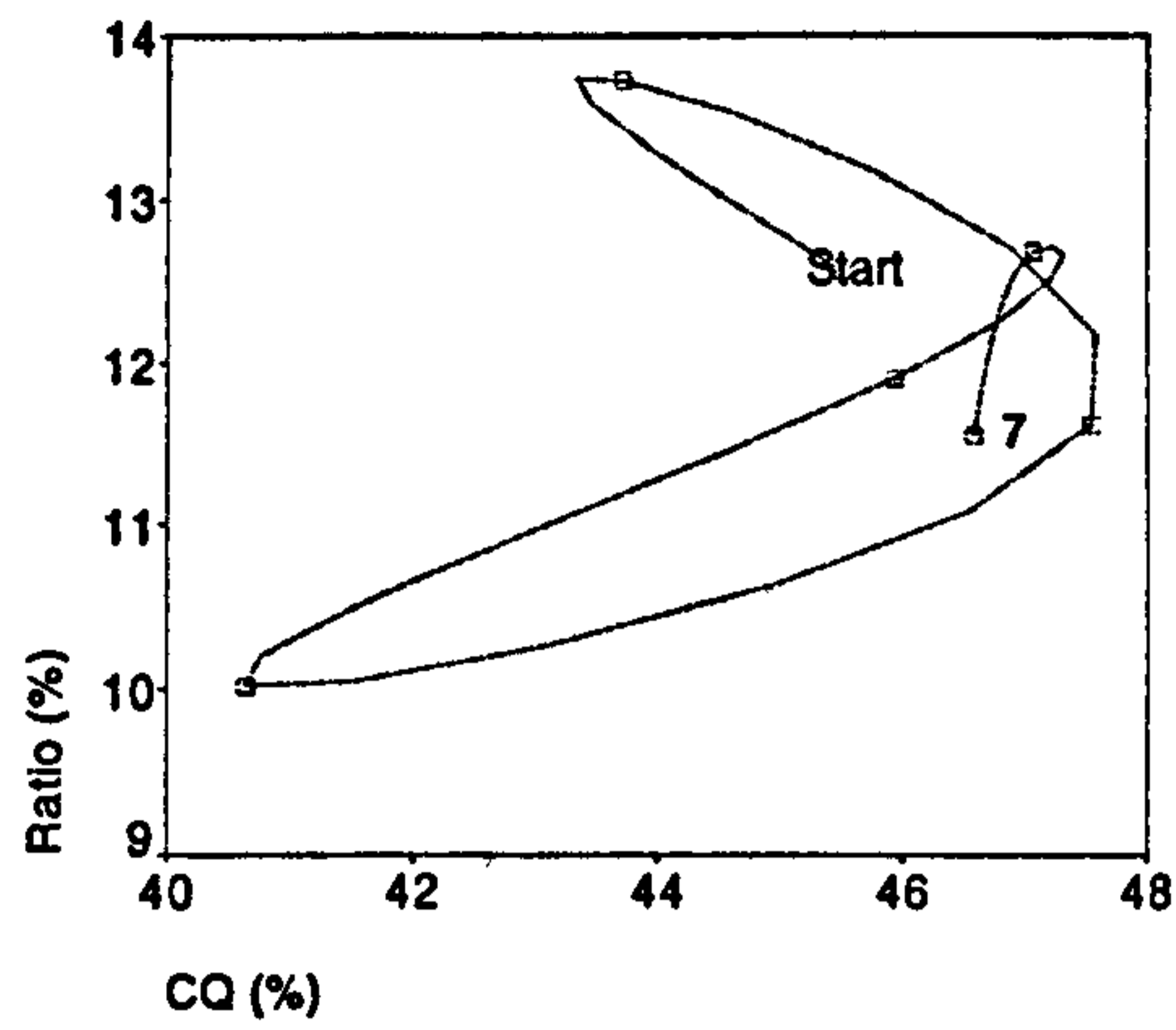
(d) Maximum projection



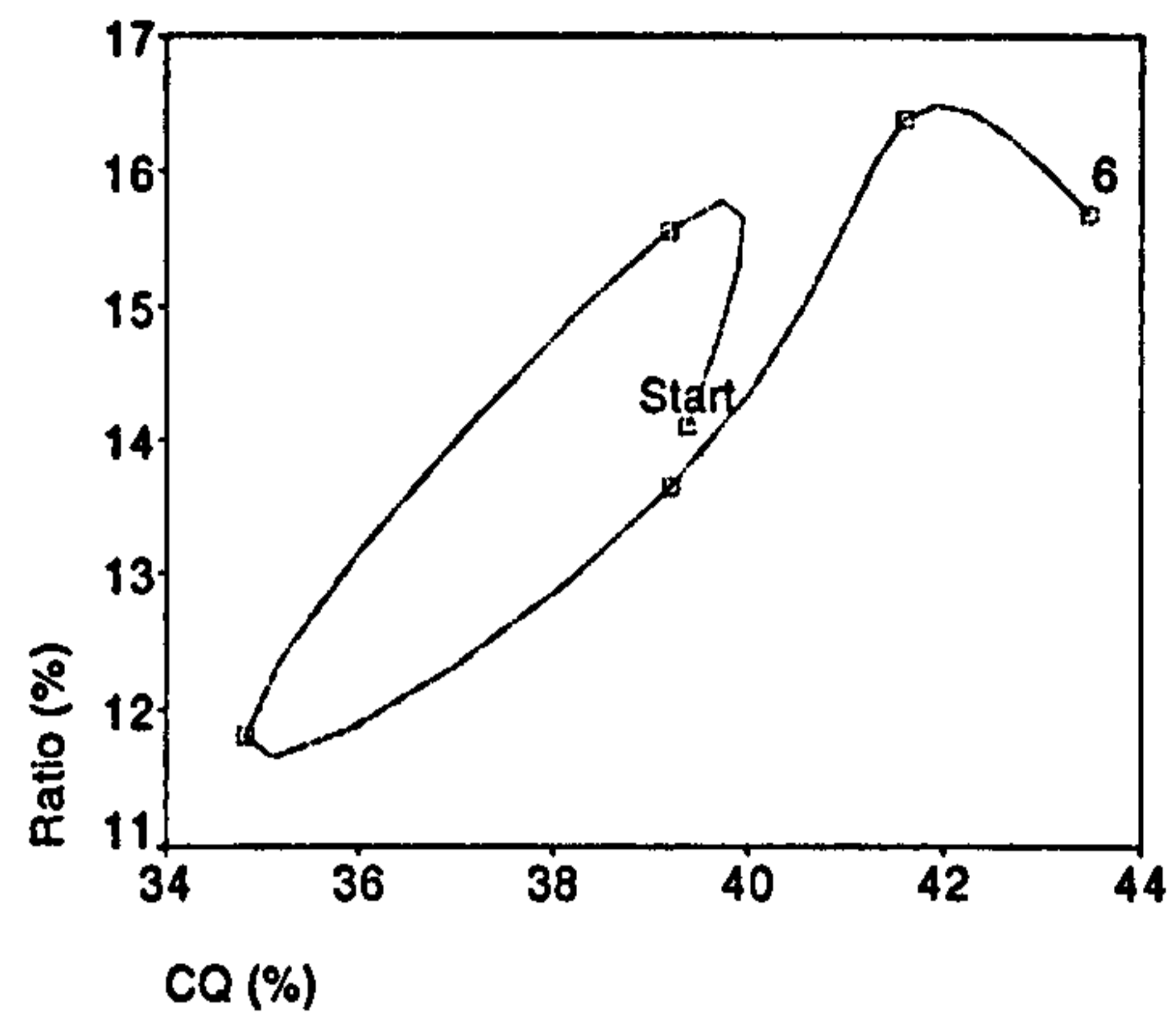
(e) Singing



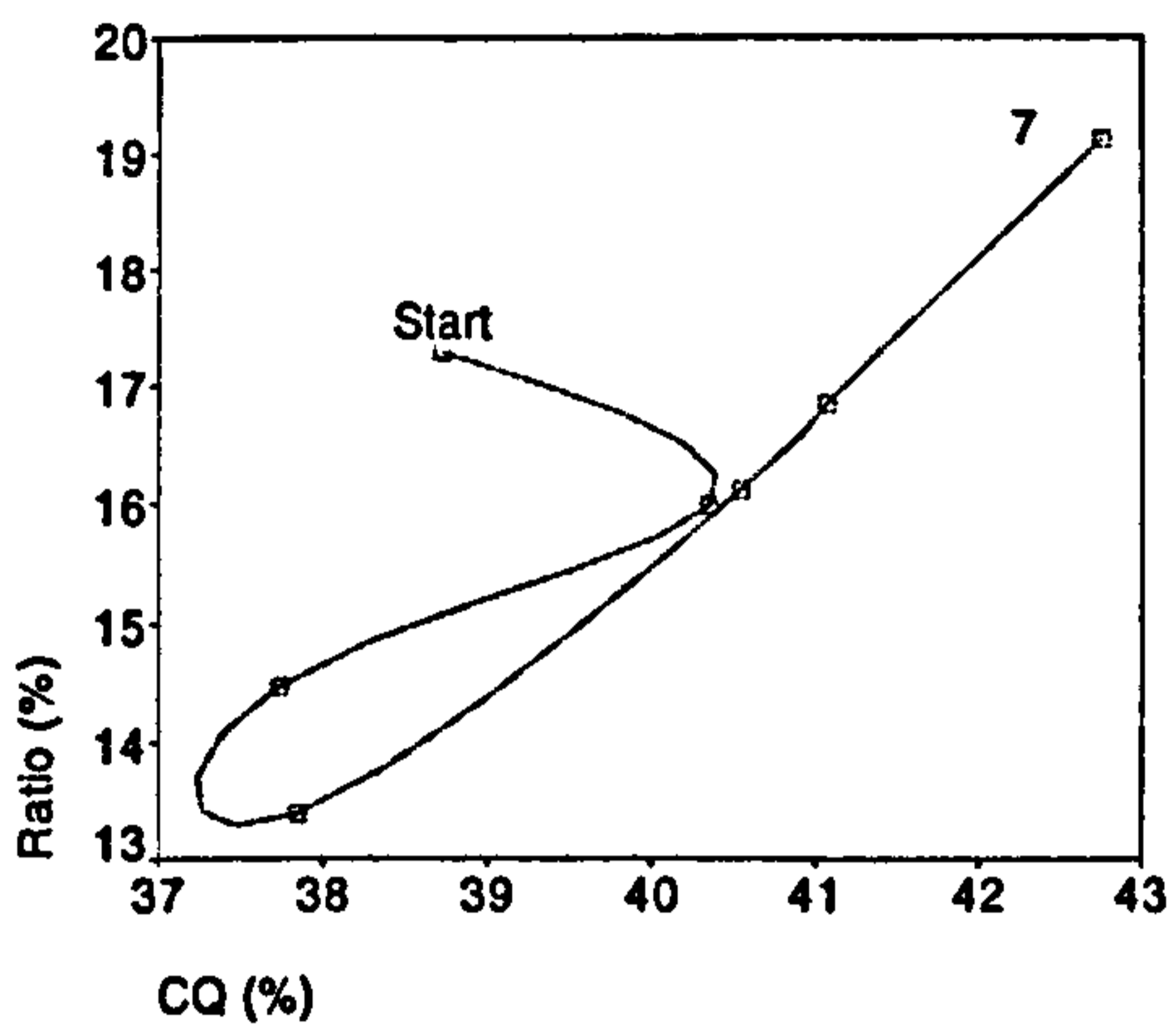
Figure 4.15: Distribution of CQ/Ratio means for each male subject for each vocal task



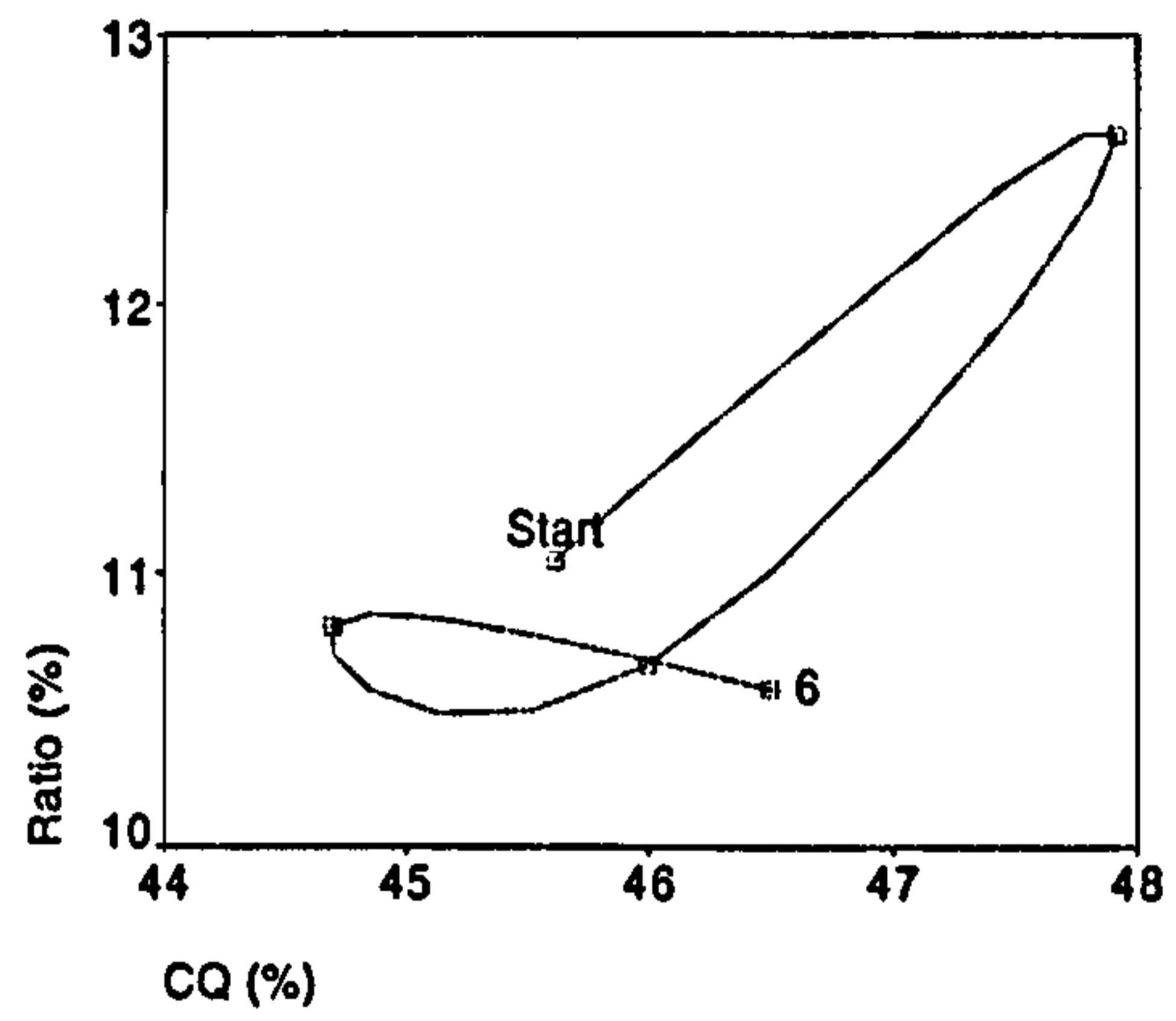
(a) Male actor 1



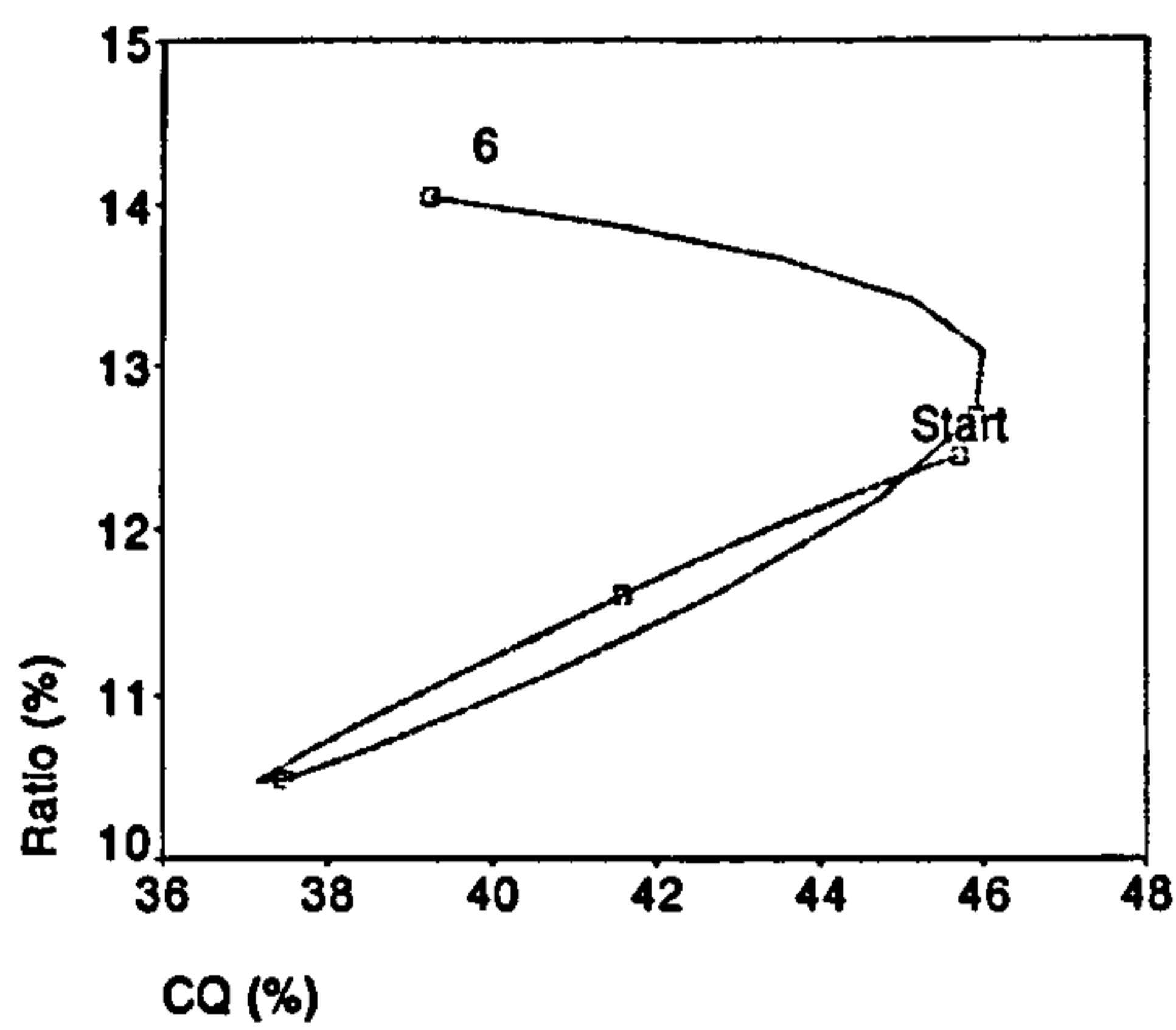
(b) Male actor 2
(absent from session 7)



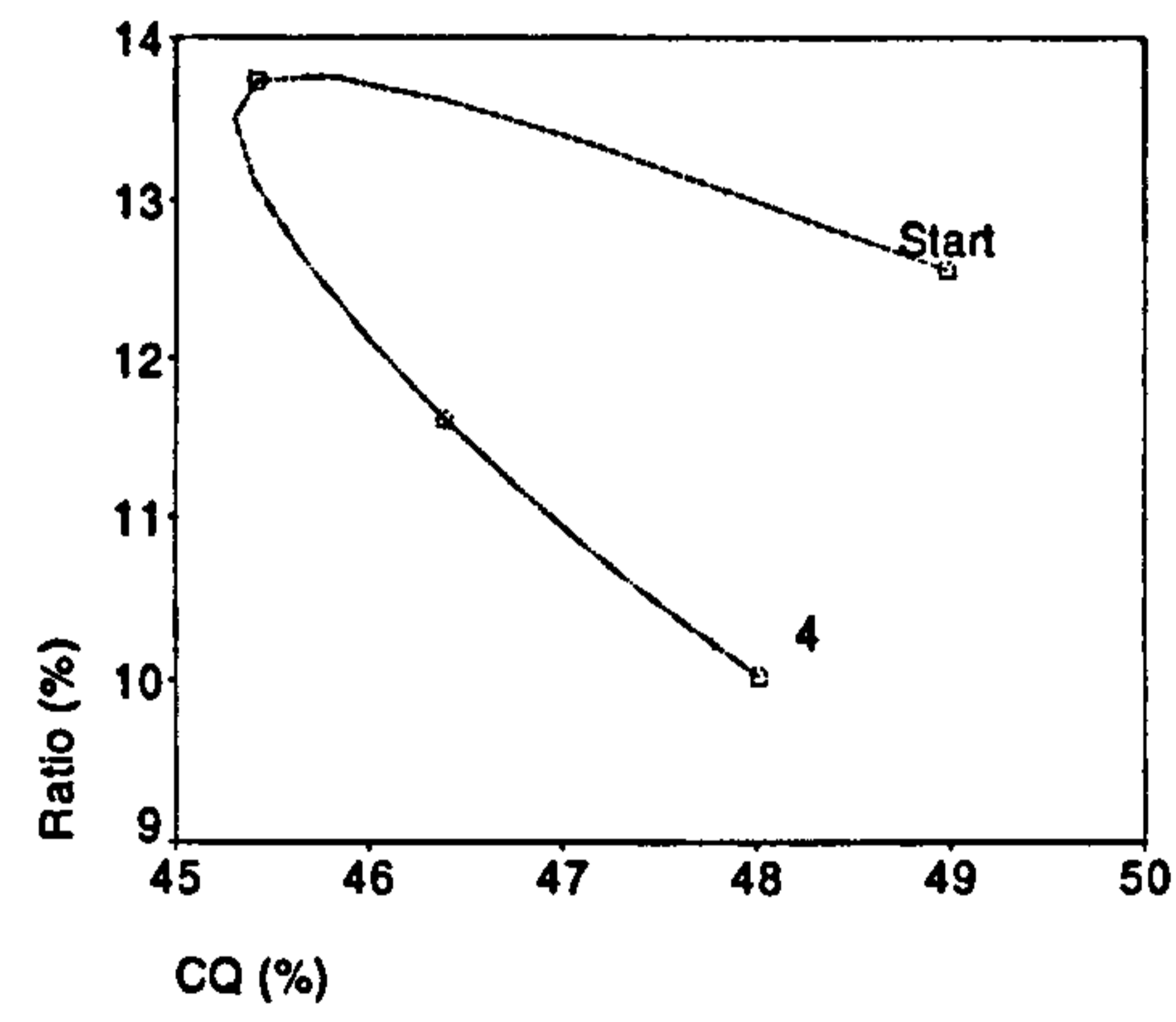
(c) Male actor 3



(d) Male actor 4
(absent from session 5, 7)



(e) Male actor 5
(absent from sessions 2, 7)



(f) Male actor 6
(absent from sessions 5, 6, 7)

Figure 4.16: Individual patterns of change of all five assessment tasks mean CQ/Ratio data. Spline interpolation is used to indicate one possible path of development between the data points. For each subject, the first data point is marked 'start' and the last data point marked with the number of the recording.

relative to the six individuals as a whole subject 6 demonstrates consistently low Ratio values with middle to high range CQ values, and subject 2 demonstrates low CQ values and relatively high Ratio values.

Figure 4.16 provides an indication of individual CQ/Ratio performance across recording sessions. The general pattern of a dip in CQ and Ratio values corresponding to the period of extreme stress noted previously during recording sessions 3 and 4 are clearly exhibited by subjects 2 and 3. A similar pattern is exhibited by subject 1, although with less consistency, especially with regard to initial Ratio values. There are indications that the pattern of CQ/Ratio for subject 4 also follows this pattern, with the second data point possibly indicative of temporary emphasis on a particular vocal technique, or to ill health. Subject 6 was only recorded for the first four sessions, and appears to exhibit signs of an initial decrease in CQ/Ratio performance followed by a gradual increase. However, this is hard to quantify in view of the relatively small number of data elements. Subject 5 exhibits the same dip in CQ/Ratio value shown by all male subjects concluding an errant data point which may have arisen from causes similar to that proposed for the subject 4.

4.4.9 Further correlations

Further investigation has been made concerning the relationship between the different voice parameters for the male and female subjects in this study. With regard to the application for which this study has been made, measures of normal correlations can be used in the configuration of real-time displays to appropriately encourage the enhancement of improved modes of phonation. For example, a real-time display of spectral amplitude in the singer's formant band may be of benefit in the process of singing tuition. However, if this measure is shown to increase with the level of acoustic output, a subject may, for example, be fooled into believing he has improved his resonant output when in fact this has only been expressed as a consequence of increased output level. If the exact relationship between the singer's formant and SPL can be expressed mathematically, a real-time computer feedback system can be configured to take this into account.

Previous research has established that subjects phonating at successively higher levels of sound level exhibit increased levels of fundamental frequency. For example,

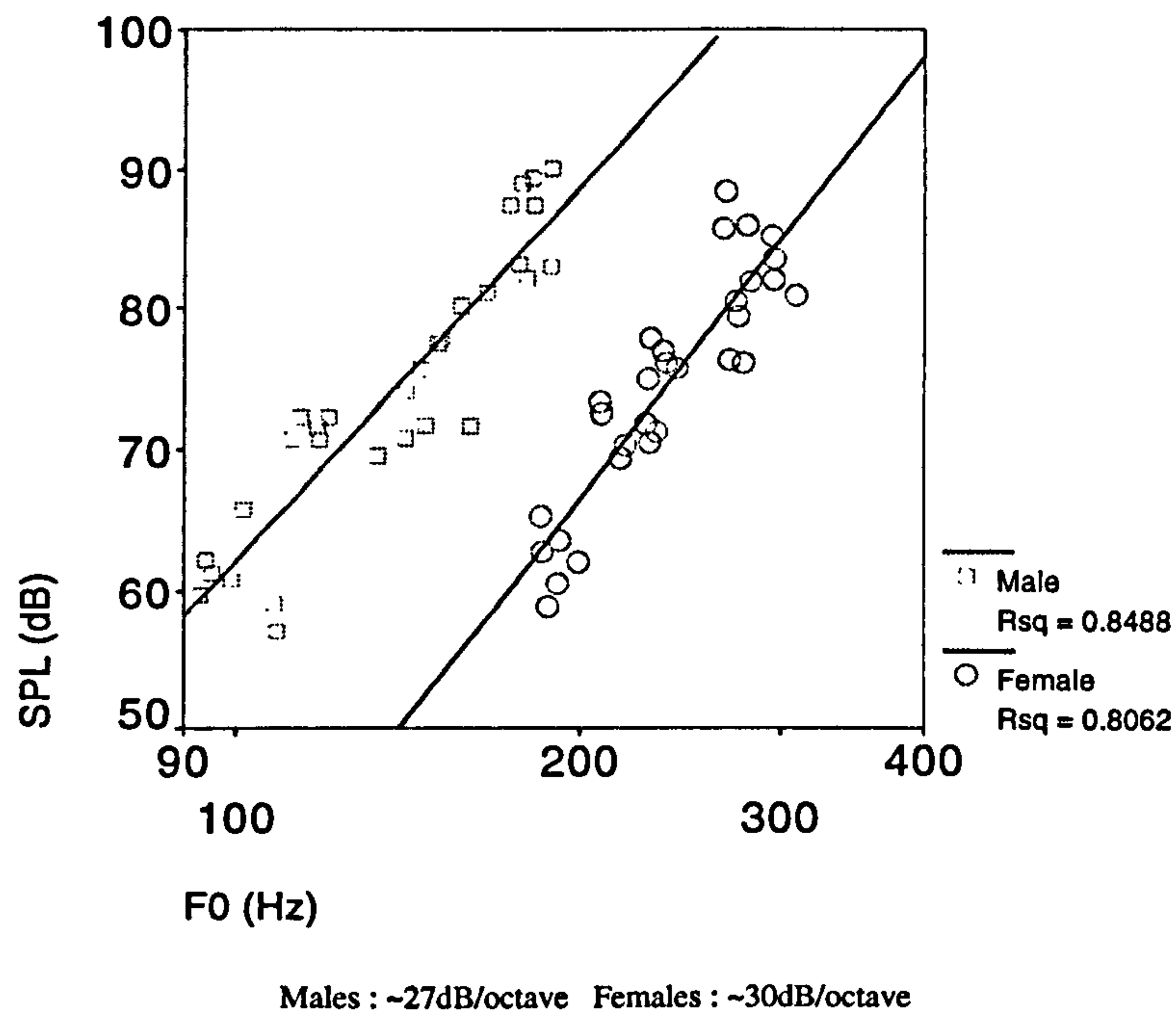


Figure 4.17: Scatterplot and linear correlation of F0 and SPL

Gramming et al. (1988) determined that an increase of 1dB results in an increase of approximately half a semitone. This correlation was later refined by the author to 0.3 to 0.5 semitones per dB (Gramming, 1991). A plot of mean data for F0 and SPL values was generated (see figure 4.17), from which it was evaluated that male subjects exhibit an increase of approximately 27dB per octave, which corresponds to 0.44dB/semitone. Female subjects exhibited a similar correlation of 30dB/octave, which corresponds to 0.4dB/semitone. Both values are therefore within the refined range of increment proposed by Gramming. The increased level of data are likely to be of benefit in further refining the nature of changes in F0 and SPL.

As discussed in chapter 3, Stathopoulos & Sapienza, (1993) required their subjects to phonate at three different levels of loudness (soft, comfortable, loud) and recorded decreased OQ (increased CQ) across these modes. However, a correlation statistic was not proposed. A plot of CQ against SPL (see figure 4.18) from the data derived for this study indicates that an increase in output level of 1dB is associated with an increase in CQ of 0.4% for male subjects. The same statistic was evaluated for female subjects as 0.08% per dB, although the correlation between the two parameters is much weaker than for male subjects. This is reflected in the relatively lower Rsq value shown.

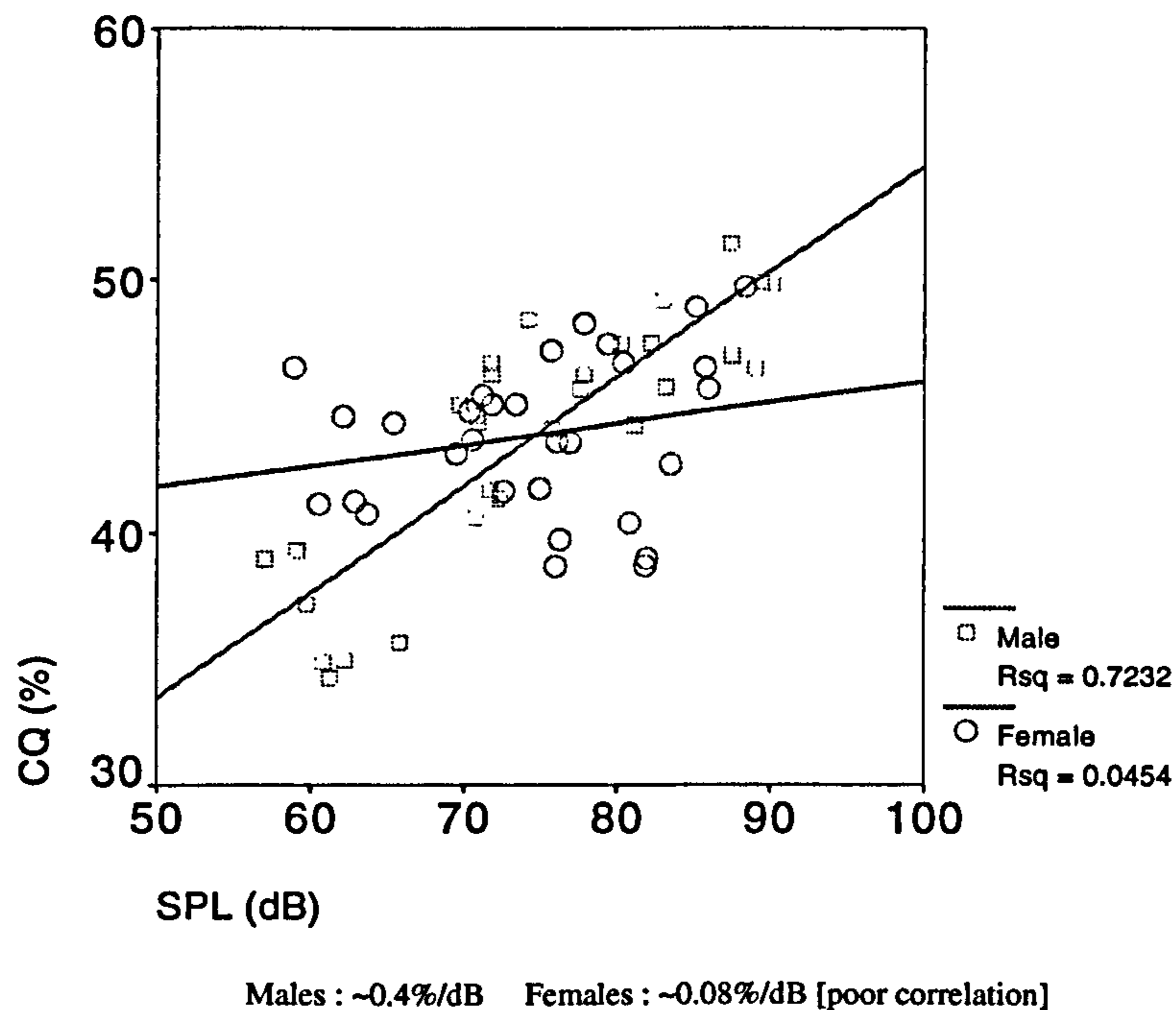
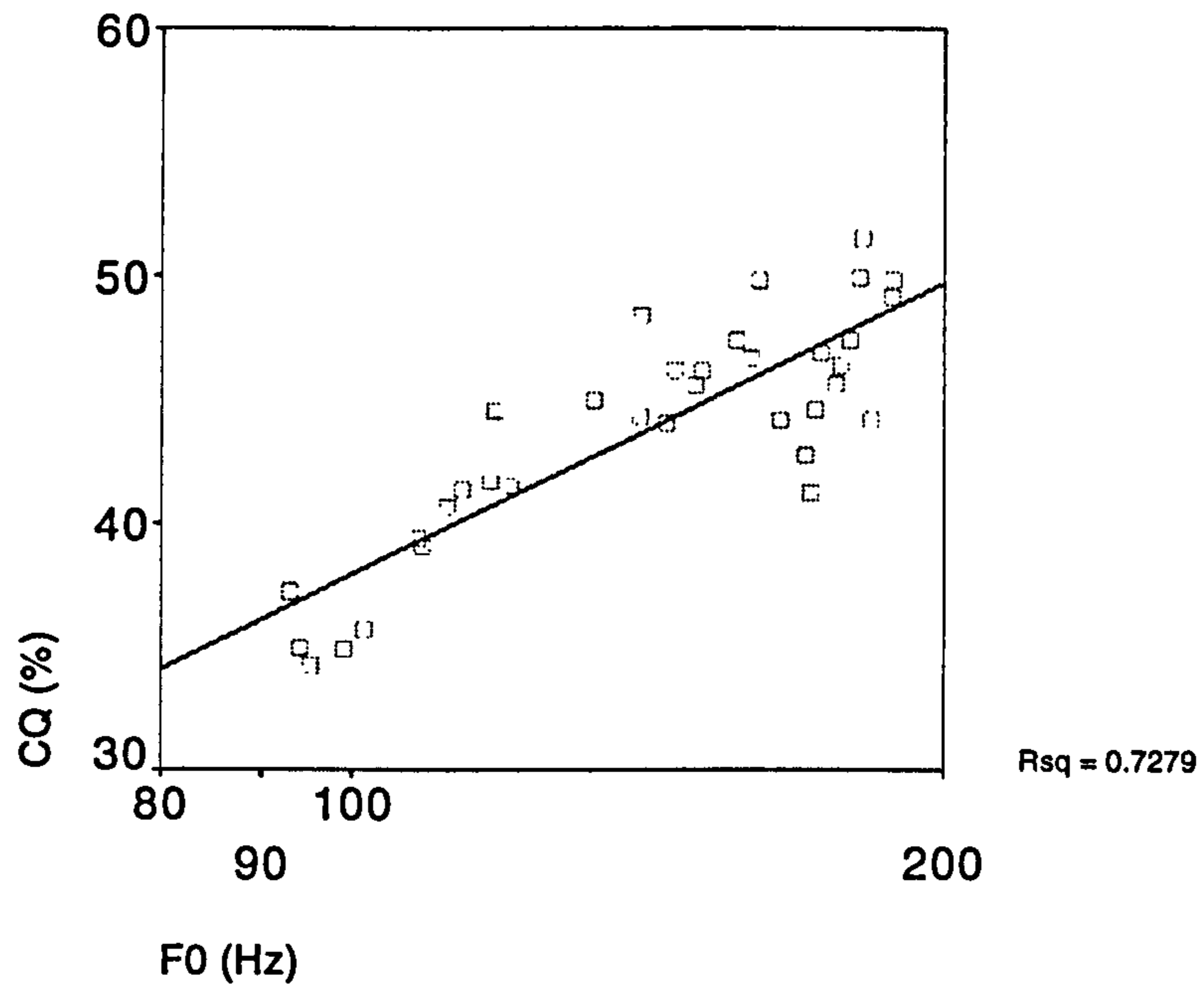


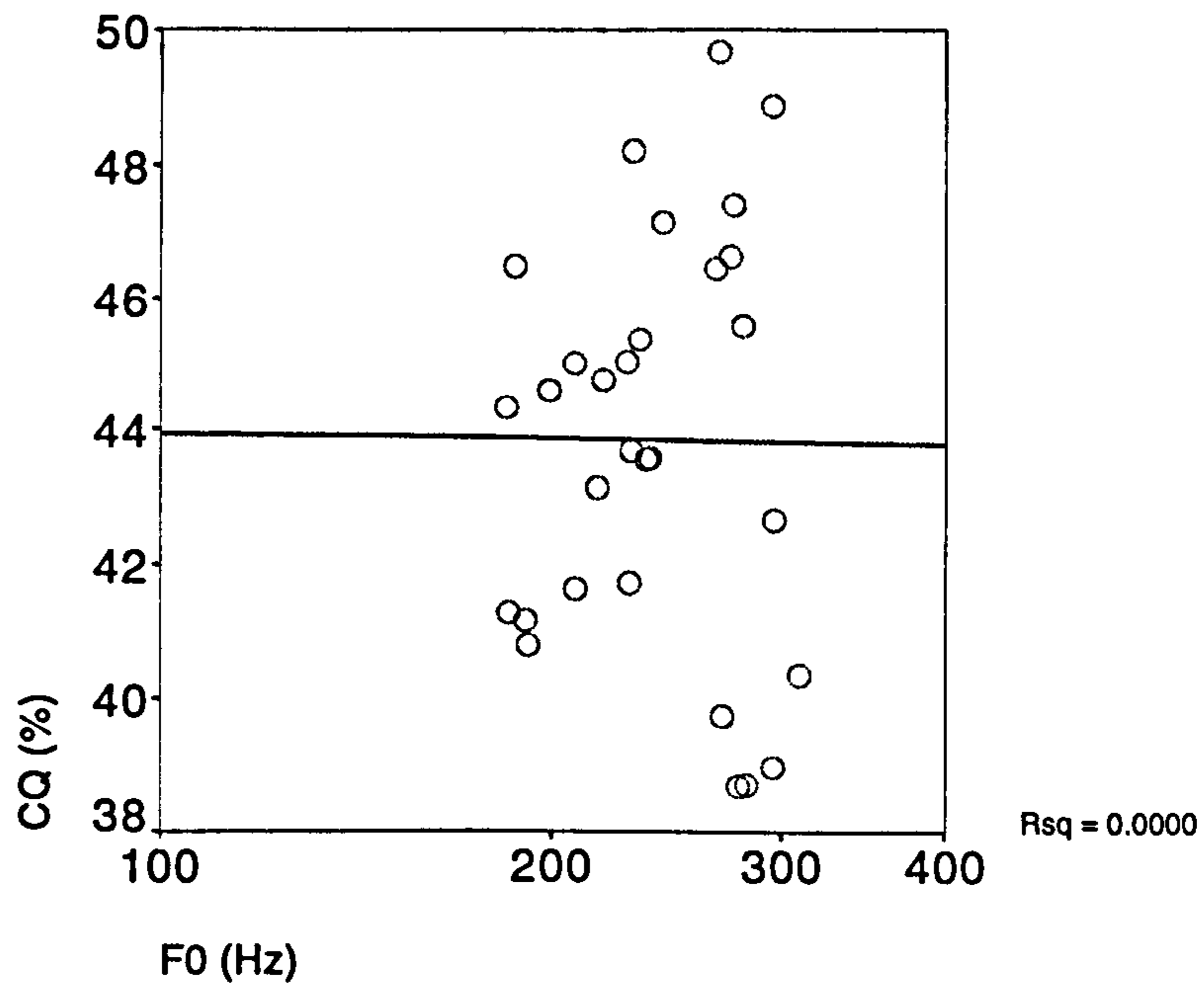
Figure 4.18: Scatterplot and linear correlation of SPL and CQ

It is logical to infer that because a positive correlation has been established between SPL and CQ, and between F0 and SPL for male subjects, it is likely that there will also be a positive correlation between F0 and CQ. Figure 4.19 illustrates the results of plotting CQ against F0. Male subjects (figure 4.19(a)) exhibit a strong positive correlation, with correlation analysis indicating that an increase in fundamental frequency of one octave will result in an increase in CQ of approximately 12%. The data for female subjects (figure 4.19(b)) exhibit no correlation, which concords with the weak relationship between CQ and SPL (figure 4.18).

The graph shown in figure 4.20 demonstrates that the level of spectral amplitude in the singer's formant region increases with the level of sound output. A previous study investigated two male and two female subjects phonated at three different levels of intensity (low, medium, high) (Hollien, 1993). Subjects exhibited increased spectral amplitude in the 2700-3400Hz band expressed as a ratio relative to the SPL, although no correlation statistic was given. In another study a high correlation was observed for each of three professional singers between the two parameters (Cleveland & Sundberg, 1983). For example, for the baritone subject analysed in the study an increase of 1dB in overall SPL resulted in an increase of approximately 1.7dB of the singer's formant. From the data derived for this study and presented



(a) Male subjects : ~12%/octave



(b) Female subjects : no clear correlation

Figure 4.19: Scatterplot and linear correlation of CQ and F0

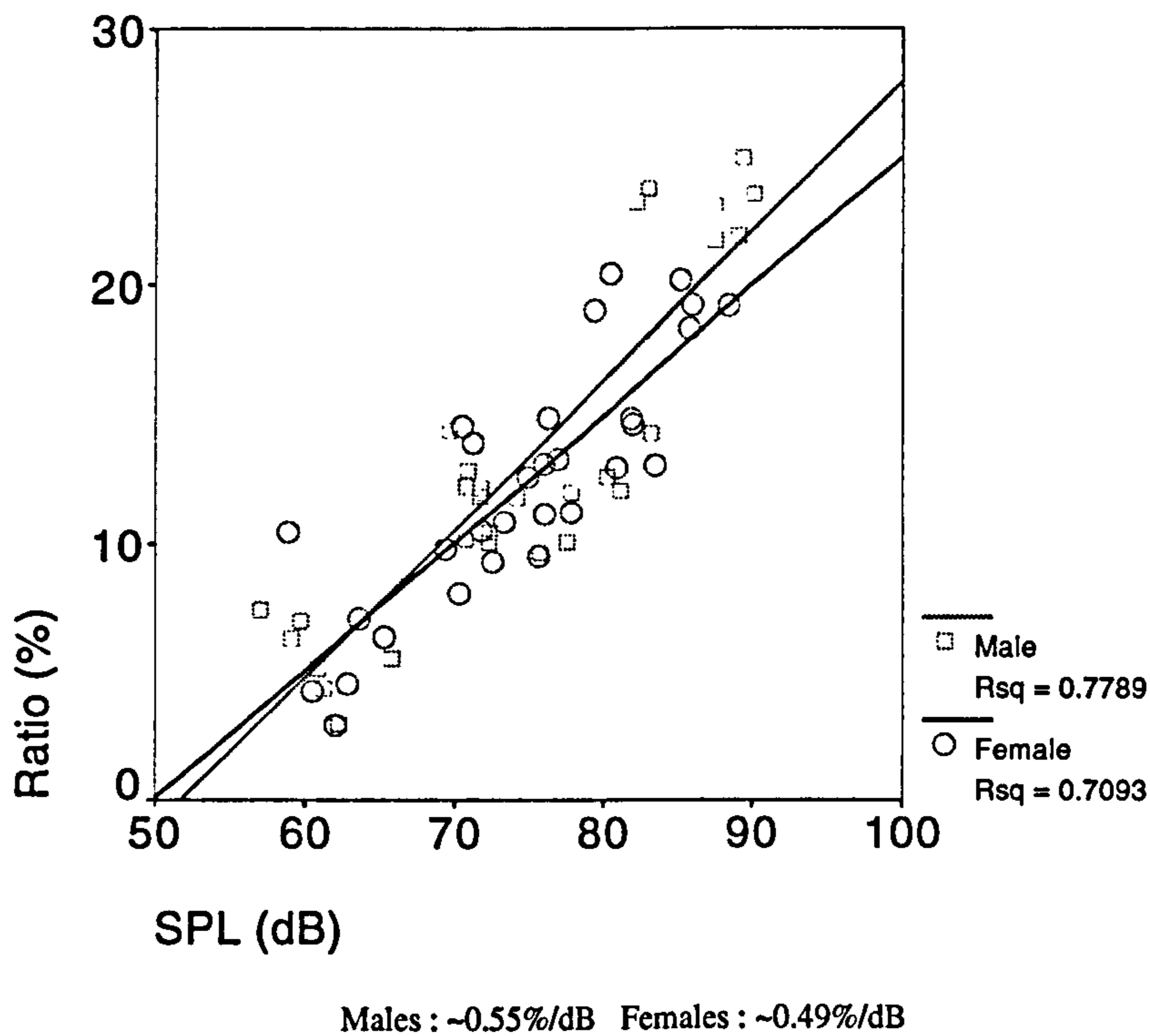


Figure 4.20: Scatterplot and linear correlation of SPL and Ratio

in figure 4.20 it has been determined that for an increase in acoustic output of 1dB there is a relative increase in spectral amplitude in the singer's formant band relative to the rest of the acoustic output spectrum of 0.55% for male subjects, and 0.49% for female subjects.

A logical inference from the combination of SPL and F0 (figure 4.17) and SPL and Ratio (figure 4.20) positive correlations is that the Ratio parameter also exhibits a positive correlation with F0. A plot of these two parameters is shown in figure 4.21 for subjects of both genders. Measurements from the graph indicate that, for both genders, an increase in frequency of one octave will result in an increase of spectral amplitude in the singer's formant band of 16%.

These data considered in aggregate provide strong indications that male and female subjects employ different modes of voice usage. Female subjects are shown to exhibit very different patterns of CQ across both SPL (figure 4.18) and across F0 (figure 4.19). However, they exhibit very similar performances of Ratio across SPL (figure 4.20) and across F0 (figure 4.21). The implication is that female subjects do not employ the same voice source mechanism as that employed by male subjects,

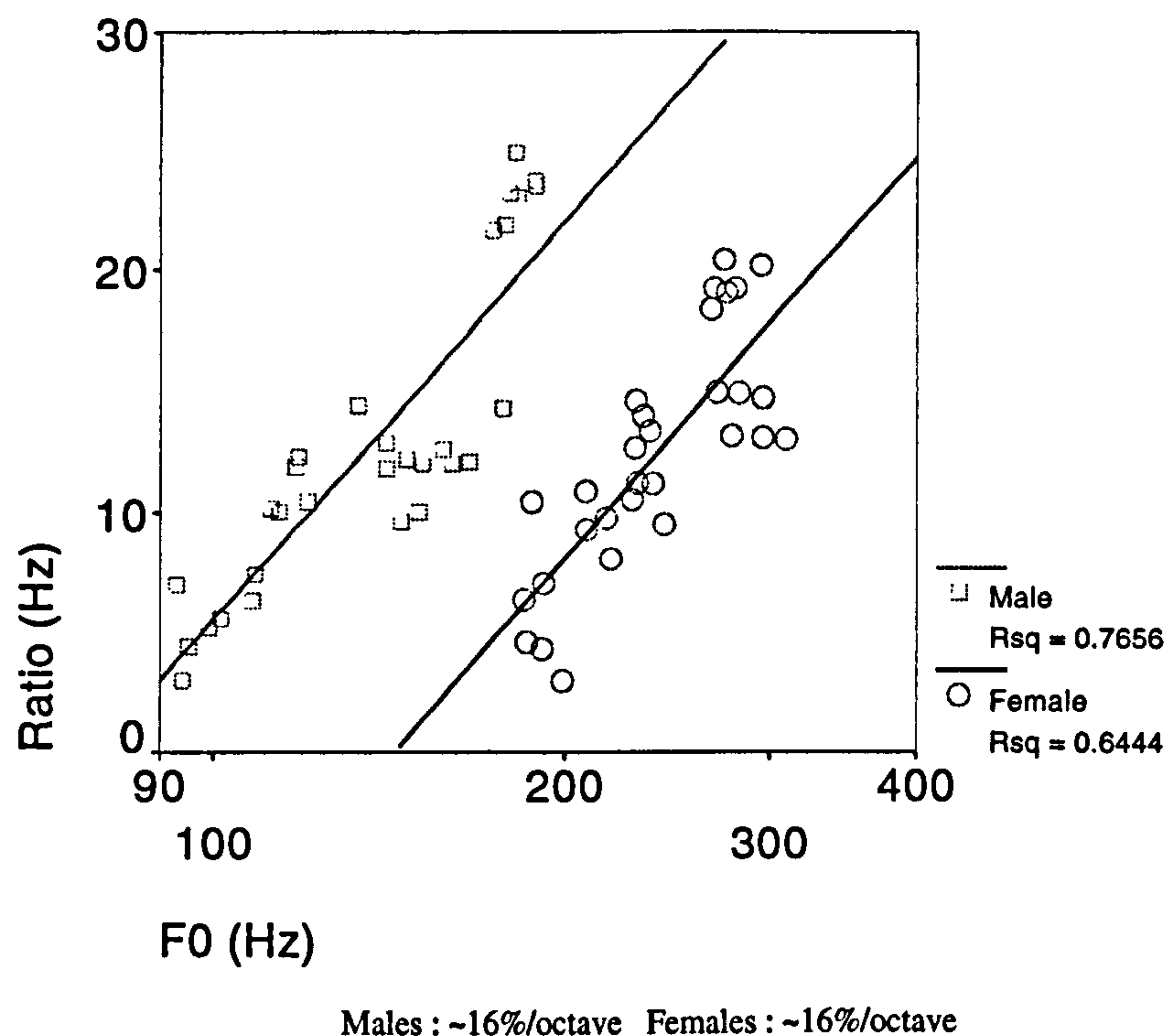


Figure 4.21: Scatterplot and linear correlation of F0 and Ratio

but are nonetheless able to achieve similar performance in terms of acoustic output.

4.4.10 Conclusions

This study has investigated a number of voice source and acoustic output parameters in a group of male and female actors under training. Results indicate that male and female subjects exhibit similar patterns of CQ with training. Male subjects express generally increased levels of spectral amplitude in the region of the singer's formant with training. This is not the case for female subjects, who appear to rely on increased levels of SPL output to achieve impact. There are further indications that male and female subjects do not employ the same voice source mechanism. For male subjects, a strong correlation is observed for the level of closed quotient and spectral amplitude in the singer's formant band. A developmental continuum has been proposed which is reflected in the relationship between CQ and Ratio parameters.

There is additionally some indication that trauma experienced by the subjects during the course of the study resulted in temporary changes to their voice usage.

4.5 The discard of two parameters

With regard to the two parameters: the rate of opening and closing of the vocal folds, neither study 1 nor study 2 provided any clear indication of the effectiveness of the parameters as indicative of vocal progress (section 4.2.6 and 4.4.5.1 respectively). Although this does not mean that the parameters are not useful in some voice assessment contexts, these parameters (as defined in section 5.9.4) have been discarded as monitors of long-term vocal development.

4.6 Study 3 *Singers over a 2.6 year period*

4.6.1 Introduction

This chapter presents a long-term study over a maximum period of 2.7 years of three male subjects undergoing singing tuition.

4.6.2 Subjects

Three male subject were recorded. They are referred to as subjects M1, M2 and M3. They were recorded over a period of 141, 120 and 69 weeks respectively. This equates to approximately 2.7, 2.2, and 1.3 years. All subjects were in their early twenties when recording began.

Subject M1 had several years of vocal training experience and chose singing performance as his main study interest as part of a first degree in Music. The other two male subjects had no previous vocal tuition of any form. They attended singing lessons for personal interest. Each of the three subjects received tuition by different tutors.

A summary of details concerning the subjects is shown in table 4.3.

<i>Subject key</i>	<i>Age at start</i>	<i>Recording span (weeks)</i>	<i>Number of recordings</i>	<i>Experience prior to recording</i>	<i>Experience/ motivation</i>
M1	20	141	10	2 years pre-degree study	Main instrument of study for first degree
M2	23	120	9	None	No public performances, for personal interest
M3	20	69	12	None	Amateur performances, for personal interest

Table 4.3: Experimental subjects for study 3

4.6.3 Recording

In order to assess their speaking voice, subjects were required to read aloud a sequence of phonetically balanced text frequently used for speaking voice assessment. In order to assess their singing voice, subjects were required to sing a two-octave scale up and down, with inhalation between each octave. The procedure used in the assessment of these tasks is reproduced in Appendix D. Singers were not specifically asked to warm up prior to recording. In terms of frequency range, the same allocated two octave scale was sung for each recording session. The times taken by the subjects to read the spoken passage and sing the two octave scale were typically 80 seconds and 25 seconds respectively. For the recordings, a cardioid Sennheiser MKH/40/P48 microphone was used. Subjects were required to maintain a constant distance of approximately 0.3m from the microphone. Prior to recording, subjects were required to provide an indication of the loudest level at which they were to phonate and appropriate recording levels were set to ensure clipping would not occur. These remained unaltered for the duration of the session.

The measure of the goodness of fit between two variables, the coefficient of determination (R^2), was applied to the data derived.

4.6.4 Results

4.6.4.1 Laryngeal closed quotient

Figure 4.22 illustrates the pattern of change for the singing voice assessment of the CQ parameter across time. Subject M1 exhibits a strong increase during the first 20 weeks of analysis, with a general increase from the first assessed CQ of 37% across the first 18 weeks to a CQ of 47%. Data from the remaining three assessments indicate a generally stable CQ value in the mid 40% region, with a final CQ value of 49%. The general pattern of increase in CQ is substantiated by the positive linear and rank correlations shown. The difference between initial and final CQ for this subject is approximately 12%. It is interesting to compare this span with that presented by Howard et al. (1990) in their assessment of 18 male singers who had differing levels of vocal tuition. This study listed extreme values of 35% to 63%, a span of 28%. The smallest value in this study was the assessed mean CQ of

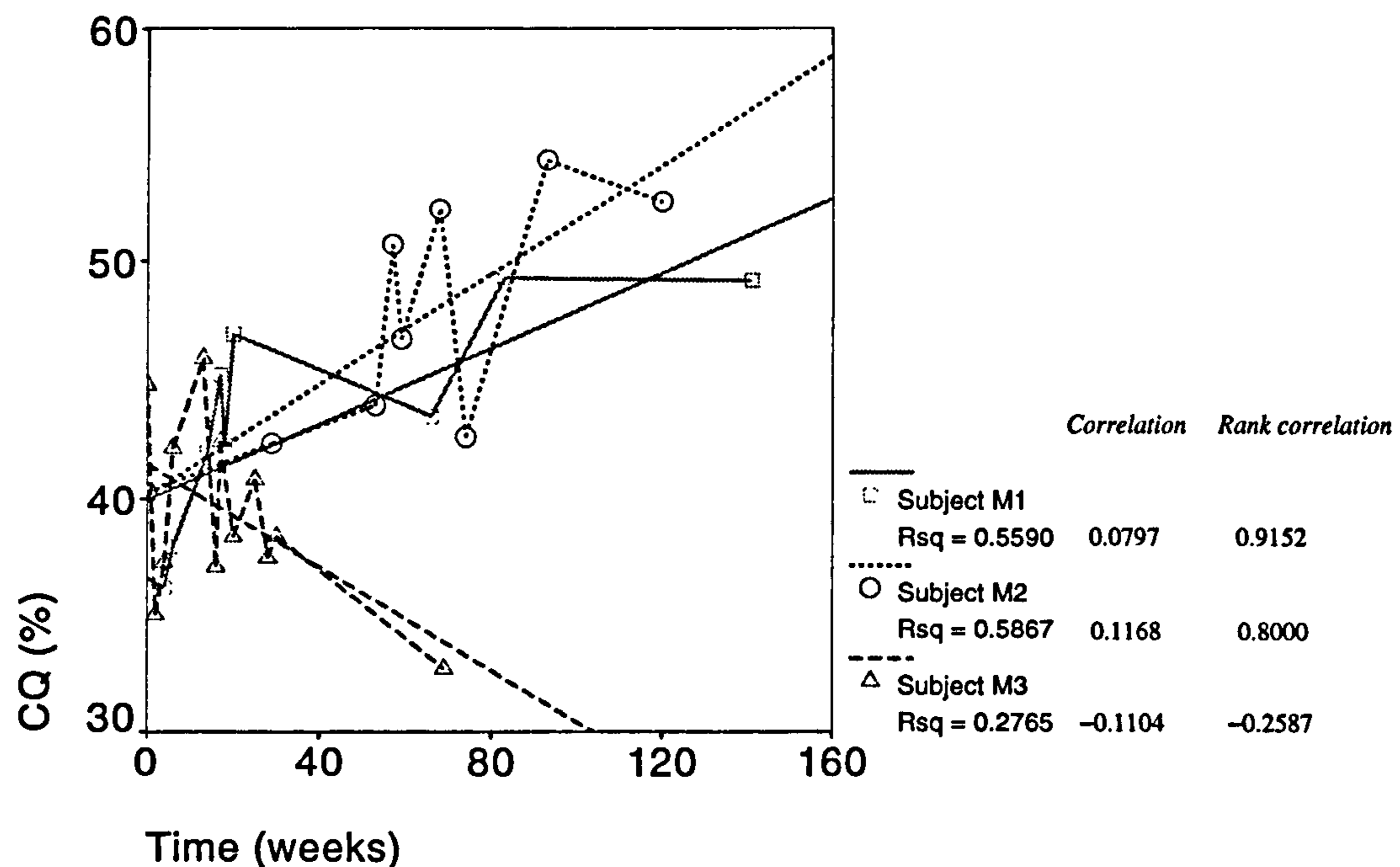


Figure 4.22: Measures of CQ across the recording sessions for the singing voice assessment

a male subject who had no singing training or experience. The largest value was the assessed mean CQ for a subject with 6 years singing training and 30 years singing experience. The increase of 12% expressed by subject M1 over a period of 2.7 years therefore indicates a comparable rate of growth.

The initial assessment for the singing voice of subject M2 is 40%. There then follows a consistent pattern of increase during the first 40 weeks, followed by a period of fluctuation over the next 40 weeks. The final two data points indicate a sustained level of CQ in the region of 50% to 55%. The span from initial to final CQ value for this subject is approximately 13%. This figure is similar to that noted previously for subject M1, and it also compares favourably with data from the Howard et al. (1990) study. Positive linear and rank correlations reflect the observed pattern of general increase.

The increasing trend demonstrated by subjects M1 and M2 is not expressed by subject M3. The assessment for this subject demonstrates a highly fluctuating pattern. Linear and rank correlations indicate a generally decreasing trend. In the assessment of the contrasts presented by this subject, it is important to remember

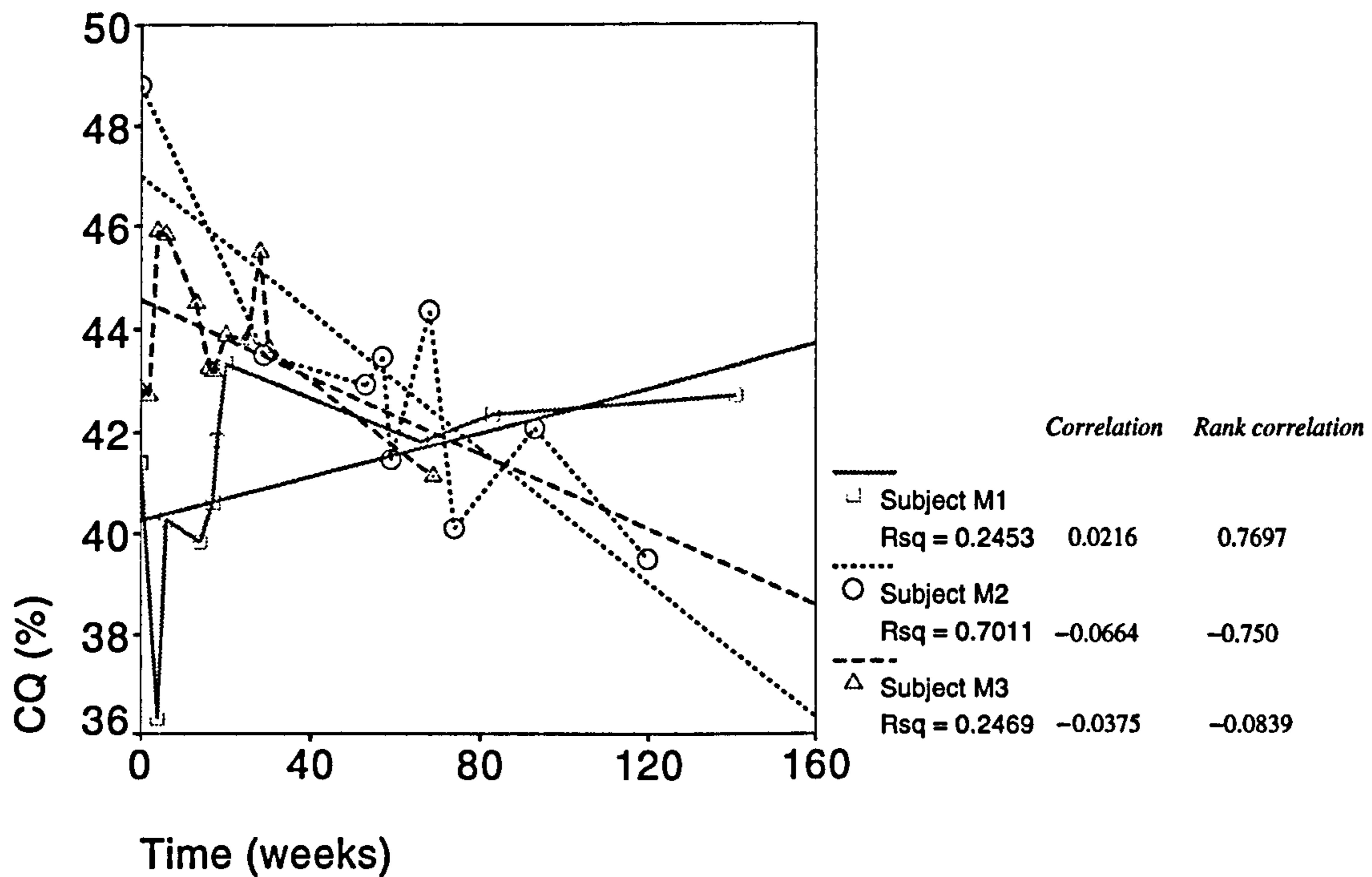


Figure 4.23: Measures of CQ across the recording sessions for the speaking voice assessment

that the data for this subject is available for only half the time period of the other subjects. It may be that although the subject underwent singing lessons his appreciation of the instruction he received was limited. An objective measure of the level of absorption of advice received during singing tuition is very hard to derive, and this has not been attempted. Another possible reason for the contrasting trend may be that the tuition received by the subject was not appropriate for a developing singer. This might not be identified in the amateur context within which the subject infrequently performed. The tutor may have regarded subject M3 as too restricted in his vocal delivery and prescribed exercises for increased breathiness, which might result in a decrease in the level of CQ. Alternatively, it may be that for this individual an increase in mean CQ is not expressed until after a relatively much greater period of vocal tuition than that received by subjects M1 and M2. If further assessment had been possible a subsequent shift in direction towards increasing CQ may have been recorded.

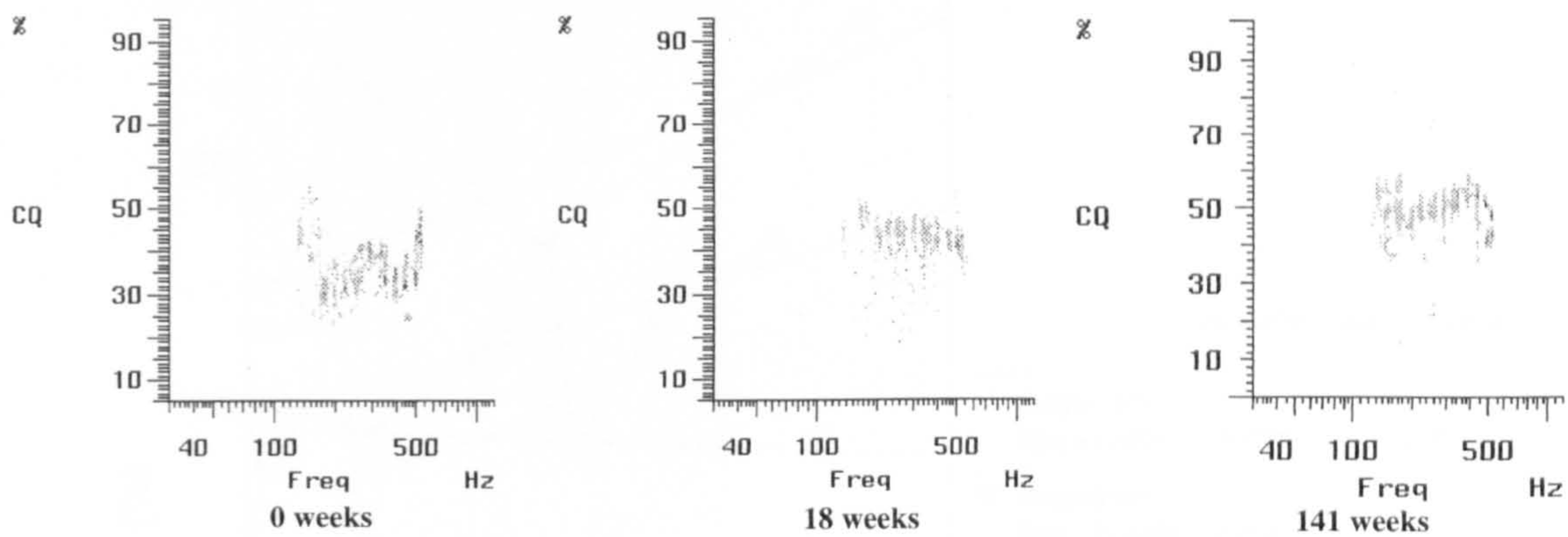


Figure 4.24: Scattergrams of CQ against F0 for subject M1. The number underneath each figure indicates the number of weeks that have elapsed since the first recording.

Figure 4.23 illustrates the pattern of change in mean CQ for the speaking voice assessment of the three subjects. The pattern of CQ generally fluctuates more than for the corresponding singing voice data, although the magnitude of change in CQ is typically half that of the singing data. Subjects M1 and M2 exhibit decreasing trends, whilst subject M1 exhibits a positive correlation, although this is of a much smaller magnitude than that for his singing voice assessment (0.0216 compared to 0.0797). The contrast between the singing and speaking voice data trends may be a clear reflection of the form of vocal tuition experienced by the subjects; tuition was given in the art of the singing voice only. The data also concurs with an observation of the Howard et al. (1990) study discussed previously in which the high correlation between mean CQ and the level of singing tuition for singing voice was shown to be considerably smaller for the speaking voice.

Figure 4.24 presents three graphs of CQ against F0 for the singing voice assessment of subject M1. The subject exhibits an essentially constant, linear pattern of CQ across F0, with a clear tendency for the general level of CQ to rise across time. Again, this concurs with the Howard et al. (1990) study in which a generally level response of CQ across F0 was observed.

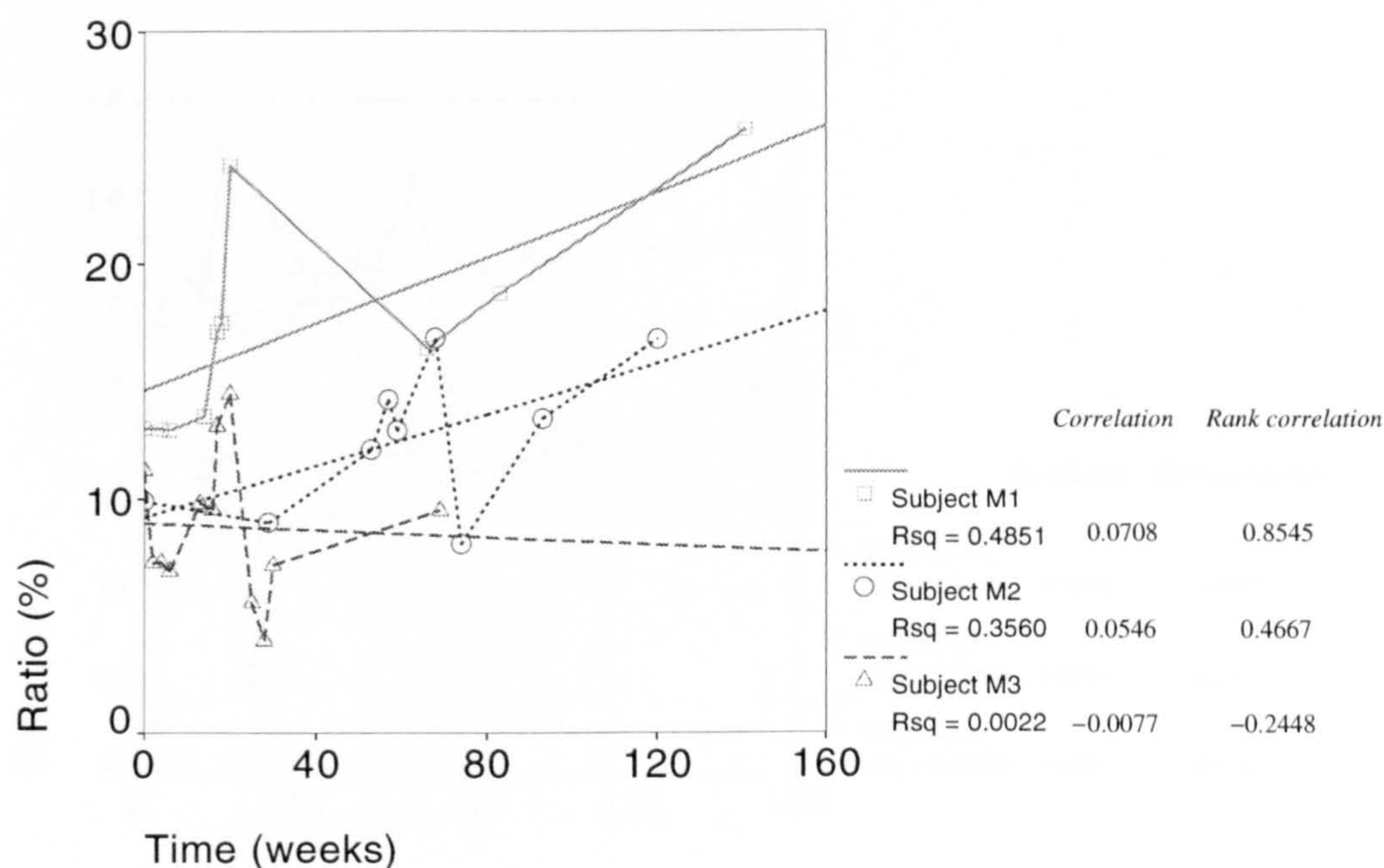


Figure 4.25: Measures of Ratio across the recording sessions for singing voice assessment

4.6.4.2 Ratio

Figure 4.25 illustrates the pattern of change for the Ratio parameter across time for the singing voice assessment. For subject M1, the pattern of change in Ratio is similar to that of CQ. The subject exhibits a strong increase during the first 20 weeks of analysis, with an increase from the first assessed mean Ratio value of 12% to a value of 25% across the first 18 weeks. The following three data points indicate a dip in value to 16%, followed by a linear rise to the final Ratio value of 26%. The difference between initial and final Ratio values is 13%. The general pattern of increase in Ratio is confirmed by the positive linear and rank correlations shown.

The initial assessment for the singing voice of subject M2 is 10%. There then follows a pattern of general increase during the first 60 weeks to 17%. This is followed by a downwards fluctuation over the next 40 weeks, resulting in a final Ratio value of 17%. The span from initial to final Ratio value for this subject is approximately 7%. Positive linear and rank correlations support the identification of a generally increasing trend, although of smaller magnitude than that for subject M1.

The increasing trend demonstrated by subjects M1 and M2 is not exhibited by

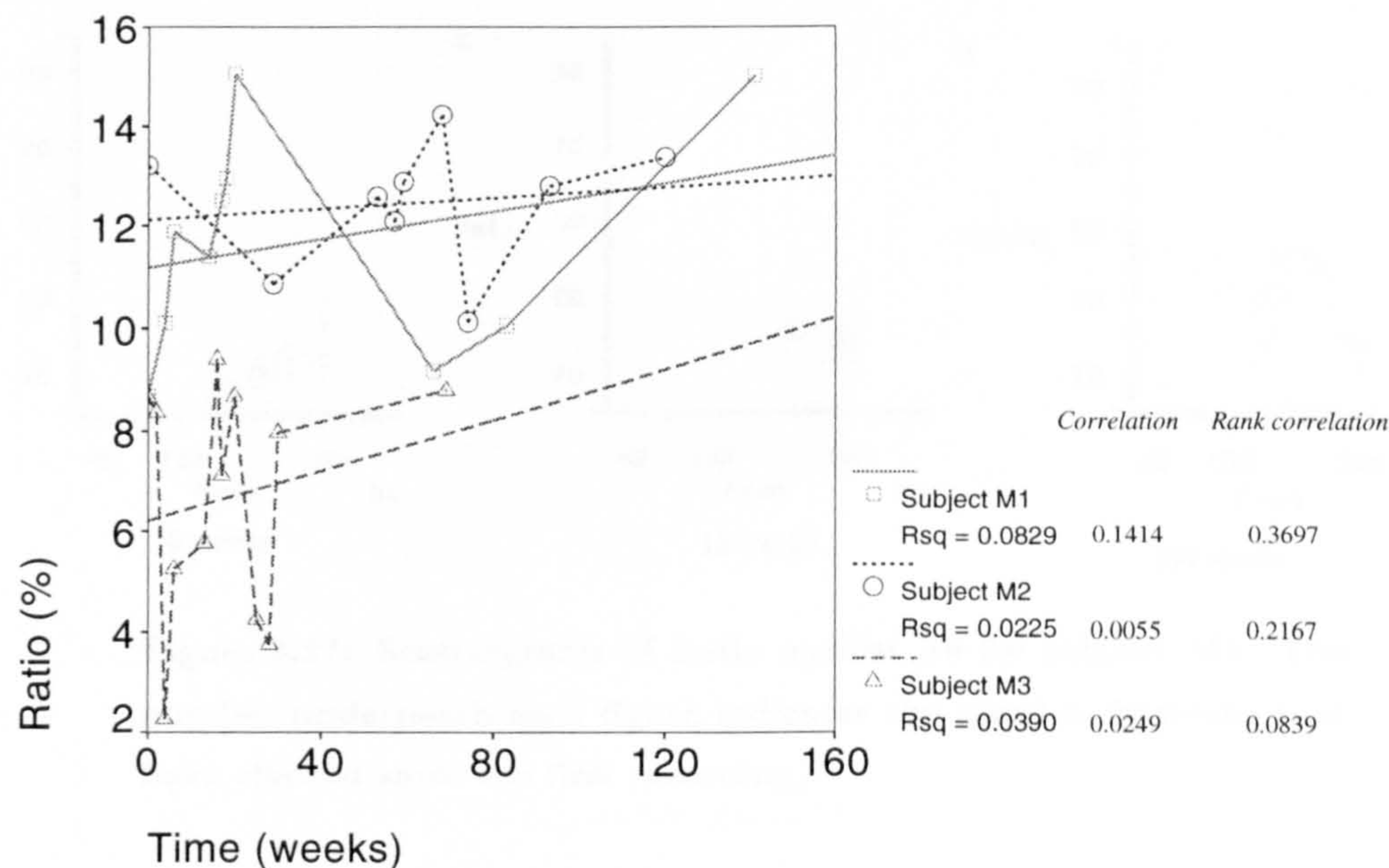


Figure 4.26: Measures of Ratio across the recording sessions for speaking voice assessment

subject M3. The assessment for this subject demonstrates a highly fluctuating pattern with a slightly decreasing trend over the 69 weeks. The reasons proposed previously for the contrasting pattern of CQ expressed by this subject apply equally to the observed difference in trend for the Ratio parameter. These include a naturally slower rate of voice development, consequences of the relatively smaller time span, and potentially different or incorrect singing tuition that may have been received by the subject.

Figure 4.26 illustrates the pattern of change in mean Ratio for the speaking voice assessment data for all three subjects. The pattern of Ratio for speech fluctuates more than for the corresponding singing voice data, and this is reflected in the relatively lower rank correlation values. However, all subjects exhibit a generally increasing trend in the parameter over time. The level of increase in Ratio over the time period of the assessment is much smaller than that of the singing voice data for subjects M1 and M2. As for the CQ parameter, the contrast between singing and speaking voice data trends is regarded as a consequence of tuition in singing voice performance only. Subject M3 exhibits an increase over the assessment period for

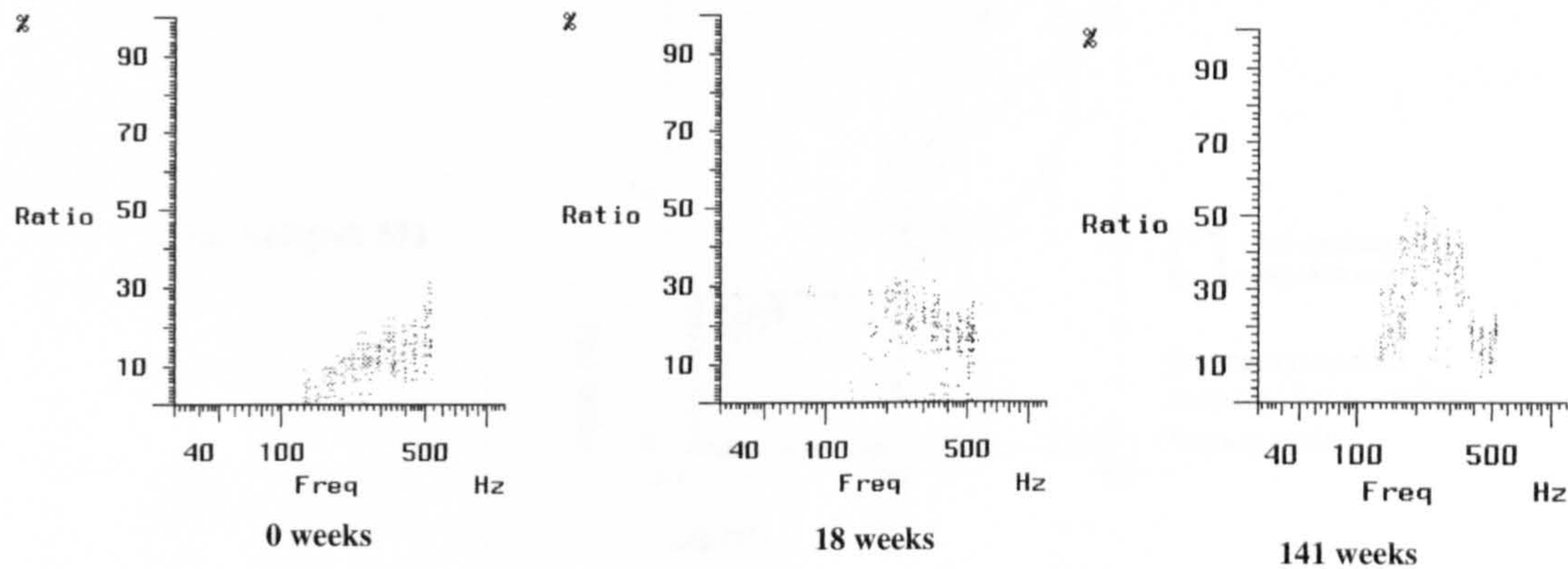


Figure 4.27: Scattergrams of Ratio against F0 for subject M1. The number underneath each figure indicates the number of weeks that have elapsed since the first recording.

his speaking voice which contrasts with the decrease for his singing voice, but linear and rank correlations for both sets of data indicate very low levels of correlation.

Figure 4.27 presents three graphs of the Ratio parameter across F0 for the singing voice assessment of subject M1. This figure corresponds to the CQ data shown in figure 4.24. Although the pattern of CQ across F0 is not as stable as that illustrated previously for CQ, the general level of the Ratio parameter is also shown to rise across time. The less consistent response across F0 may be a reflection of the relatively higher level of variability in the acoustic domain compared to that of the voice source domain.

4.6.5 The correlation between laryngeal closed quotient and Ratio

In previous research, a continuum was proposed based upon a linear correlation between the CQ and Ratio parameters (Rossiter & Howard, 1994a). The data from the three male subjects in the current study were assessed for an indication of this trend. A plot of CQ against Ratio is shown for the singing assessment data derived for each subject in figure 4.28. The range of CQ and Ratio data derived from the speaking voice assessment are also shown superimposed. This facilitates a comparison between the ranges of the two contrasting assessment tasks. Scatterplots and lines of best fit between the two variables for the speaking voice assessment data

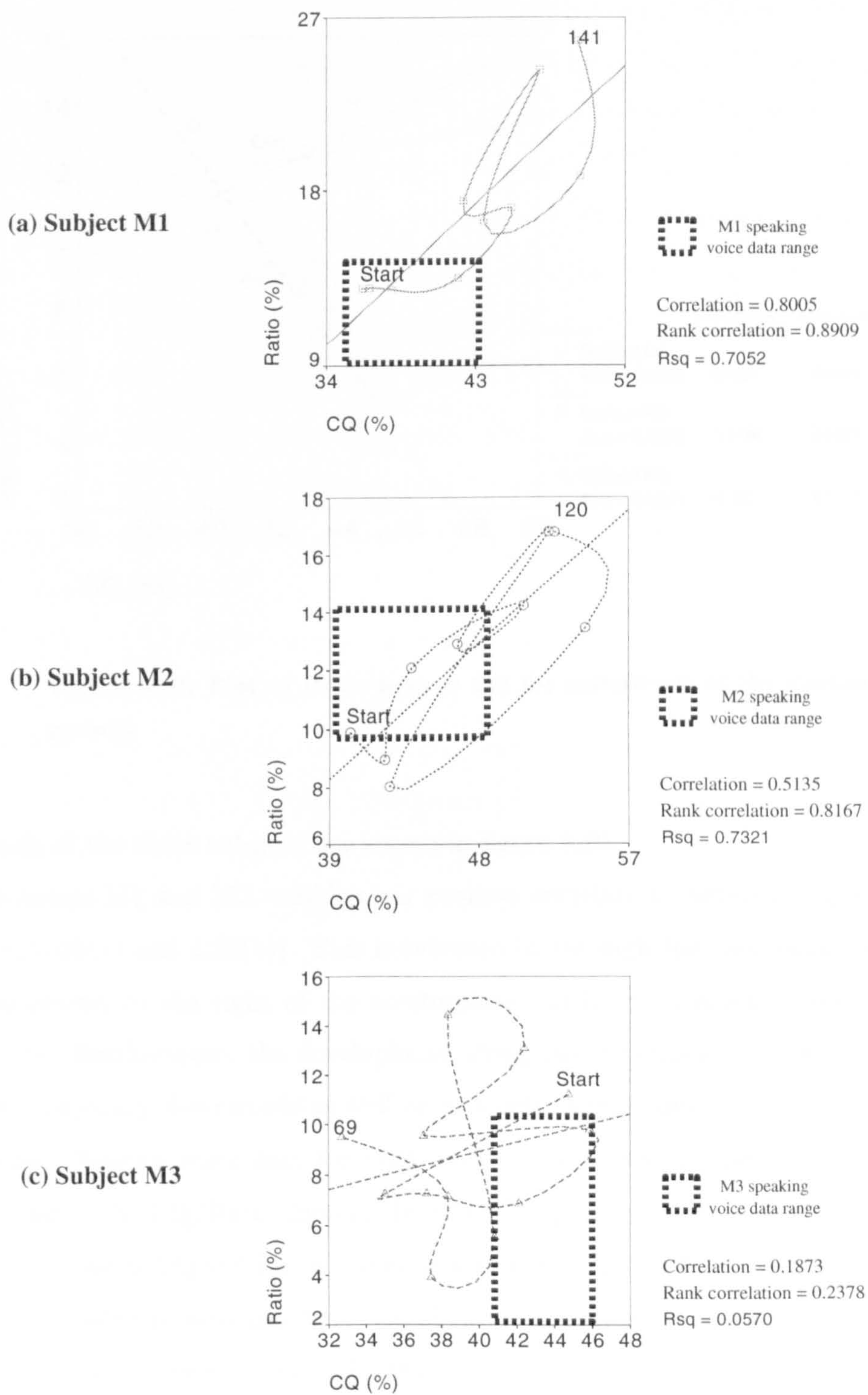


Figure 4.28: Individual patterns of singing voice CQ/Ratio development, with superimposed data ranges for speaking voice. The first recorded assessment is marked 'Start'. The last recorded assessment is marked with the number of weeks since recording began. Spline interpolation has been used to indicate one possible path of progress between the data points.

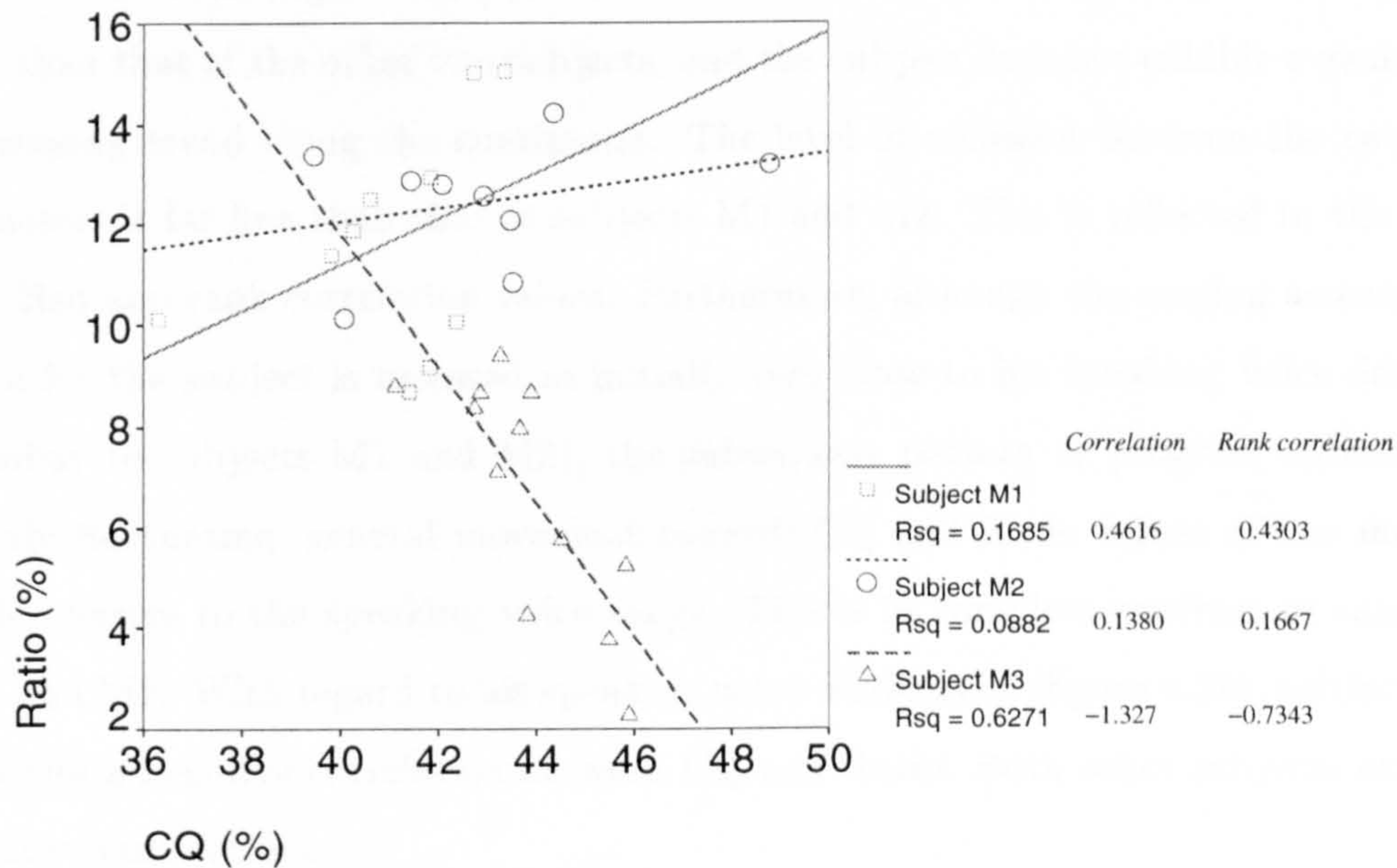


Figure 4.29: Plot of Ratio against CQ for assessment of the spoken passage

for each of the three subjects are shown in figure 4.29.

Subjects M1 and M2 exhibit clear positive correlations between CQ and Ratio (figure 4.28(a) and 4.28(b)). This is reflected in the high Rsq and rank correlation values shown to the right of the scatterplots which are of similar value for both subjects. Furthermore, the development along the continuum is clearly expressed (with temporary downwards as well as upwards fluctuations) in the data for both subjects. Singing voice data for both subjects is shown to generally start at the lower end of the CQ/Ratio domain identified for speaking voice and subsequently to develop towards CQ and Ratio values at a relatively higher level (figure 4.28(a) and 4.28(b)). Both subjects may be regarded as reinforcing the proposal made previously that one goal of voice tuition is to advance voice users along a continuum between the two parameters. These subjects also exhibit a positive correlation between CQ and Ratio for their speaking voice (see figure 4.29) although the strength of these correlations (indicated by the Rsq and rank correlation values shown) is considerably less than that for their singing voice, for which they have received tuition.

Subject M3 exhibits several contrasting properties. With regard to his singing

voice assessment (figure 4.28(c)), the line of best fit is of considerably smaller gradient than that of the other two subjects, and the subject does not exhibit a generally increasing trend along the continuum. The level of cohesion between the two parameters is far less than that of subjects M1 and M2. This is reflected in the very low R_{sq} and rank correlation values. Furthermore, although the singing assessment data for the subject is assessed as initially very close to his speaking voice domain (similar to subjects M1 and M2), the subsequent pattern of progress indicates a highly fluctuating, general movement towards CQ and Ratio values of less magnitude relative to the speaking voice range. This is in complete contrast to subjects M1 and M2. With regard to his speaking voice assessment (figure 4.29), subject M3 exhibits a negative correlation between CQ and Ratio. Both other subjects exhibit positive correlations.

Clearly, the CQ/Ratio performance of subject M3 is very different to that of the other two subjects. No specific reasons for this suggest themselves. If the subject received singing tuition of a different, possibly inappropriate, form relative to that experienced by subjects M1 and M2, this might account for the difference. However, the exact nature of the tuition received by any subject over such a long period of time would be impossible to monitor objectively.

4.6.6 Conclusions

The singing and speaking voice performance of three male subjects has been assessed over periods of time ranging up to two and a half years during which the subjects received singing tuition. Trends found for two of the subjects indicate increased levels of laryngeal closed quotient and spectral amplitude in the singer's formant region as a function of training. A third subject, observed for half the time of the other two subjects, was not shown to exhibit any clear trend with respect to these parameters over the assessment period.

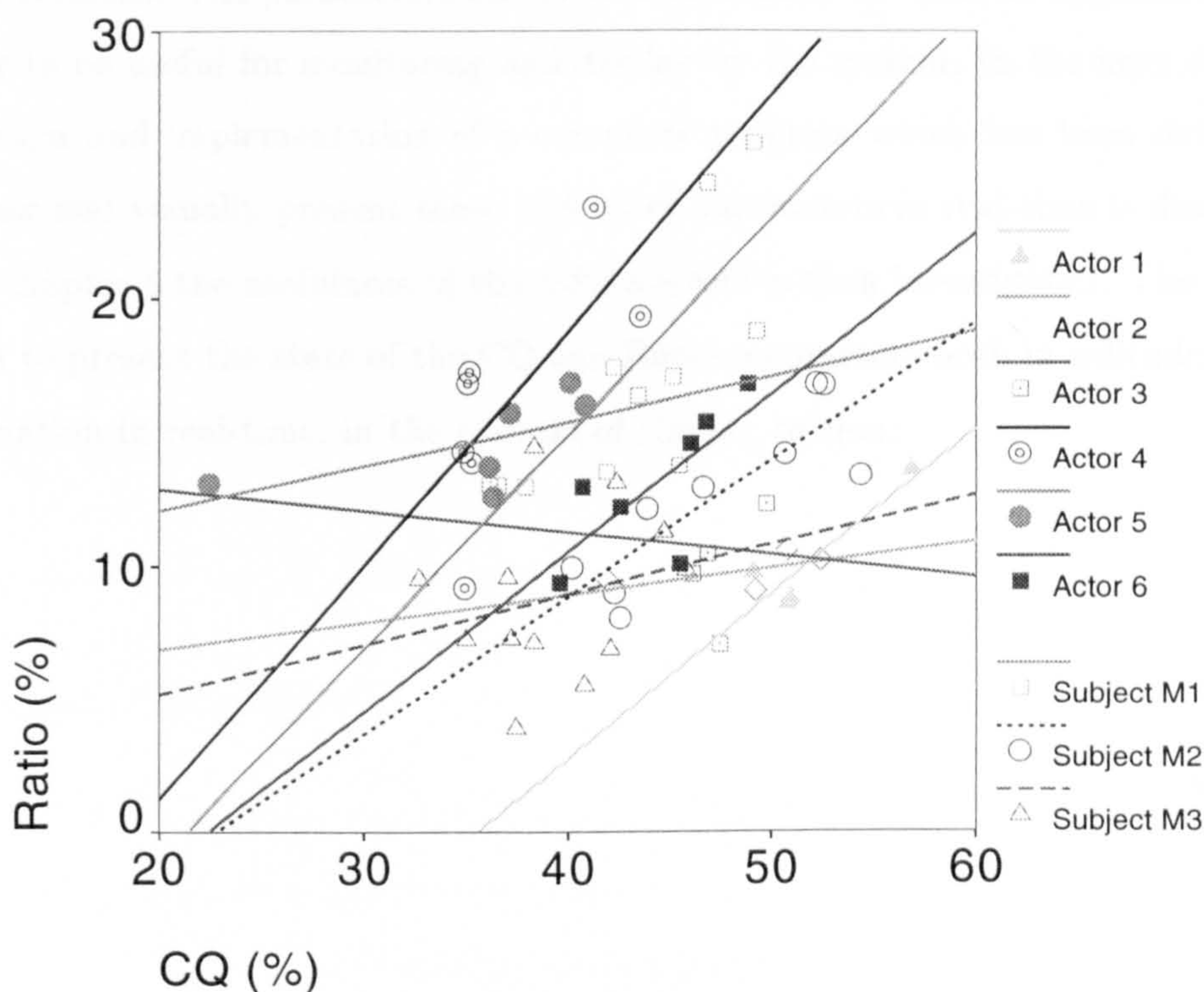


Figure 4.30: Scatterplot and line of best fit of Ratio and CQ for assessment of the singing voice, for the 6 male subjects assessed in study 2 and the 3 male subjects assessed in study 3.

4.7 Discussion of the three studies

There are clear indications from study 2 and study 3 that both CQ and Ratio parameters reflect vocal development for adult male subjects. Study 1 is regarded as an assessment of too brief a period for this phenomenon to have been expressed. The strength of the relationship between these two parameters is further illustrated by plotting the data for the two parameters derived from all the male subjects in both study 2 and study 3. This is shown for the singing voice assessment task in figure 4.30. The lines of best fit shown indicate a clear positive relationship between the two parameters for all except actor 3 from study 2, who has the least number of data points of the set.

These studies provide pointers to the design of a computer system which can

function for the provision of visual feedback for use as a complimentary tool in tuition sessions. The parameters observed as indicative of vocal development would appear to be useful for monitoring and display by the system. In the next chapter, the design and implementation of a computer program which has been developed to assess and visually present these and other parameters in real-time is discussed.

In chapter 6 the usefulness of the software tool is then investigated. The system is used to present the state of the CQ and Ratio parameters both in solitude and in combination in real-time, in the context of singing tuition.

Chapter 5

A tool for biofeedback

5.1 Introduction

In the conclusions to chapter 3 it was noted that existing real-time visual display systems only assessed a highly restricted set of voice parameters, presenting visual information concerning the state of those parameters in only a very few limited display styles. In this chapter a new software tool called ALBERT (Acoustic and Laryngeal Biofeedback Enhancement in Real-Time) is presented which has been designed for real-time visual feedback for use in a very wide range of voice tuition contexts, and for analysis which may take place in non-real-time for use in a similarly wide range of voice analysis contexts. For both modes, the analysis of the voice and the subsequent visual display of voice parameters are largely configurable. A flexible approach is provided so that voice parameters may be assessed individually or in combination. With regard to the visual displays and the graphical user interface through which the user is able to control the software, provision is made in order that the most suitable visual feedback can be presented for tuition and/or analysis. Several example applications are given based upon observations made in previous chapters, and from other research sources.

5.2 The benefit of visual feedback

Research considering the benefit of visual feedback *per se* has not yet been considered. To address this deficit, in this section an overview of previous research

concerning the benefits of visual feedback in enhancing a learning task is presented. This helps in building a context for the ALBERT system.

There are strong indications in the scientific literature that the addition of visual feedback to the normal auditory stream perceived by voice users strengthens the learning process. For example, during face-to-face interaction between people, expressions may count for as much as 55% of communication (Hadfield, 1994). Boyle et al. (1994) showed that communication between pairs of subjects who could see each other while performing a cooperative problem solving task was more efficient than that between pairs of subjects who could not see each other. Sumbly and Pollack (1954) demonstrated that allowing subjects to see the face of a talker who is uttering isolated words that must be identified against a background of noise is equivalent to improving the signal-to-noise ratio by up to 24dB. Walden et al. (1993) demonstrated that elderly and middle-aged people achieved recognition scores in the region of 90% when presented with information in combined audio and visual presentation, compared to 30% to 40% for either audio or visual data in solitude.

With regard to the application of computer based visual feedback for dysfunctional voice subjects, Michi et al. (1993) reported that computer based visual feedback was effective in treating /s/ sounds for patients with cleft palate, and Allen et al. (1991) indicated the merits of a study into the treatment of hyperfunctional dysphonia through the visual display of EMG signals. There is also a large body of literature concerning the application of visual feedback systems to the teaching of deaf people, with several reports indicating improved performance for subjects taught with computer based visual feedback when compared to similar subjects who used no computer tools (for example, Brooks et al., 1981; Arends et al., 1991) or who used tactile feedback instead (e.g., Youdelman et al., 1988). However, studies into the use of real-time visual displays with normal, non-dysfunctional voices appear to be rare. In one study, Welch et al. (1989) demonstrated that the SINGAD system discussed previously (section 3.5.5) can have a positive effect on the singing development of primary school children of normal ability, both with and without teacher interaction. There are therefore strong indications in general that the enhancement of feedback signals by the provision of complimentary visual displays can be beneficial to the learning process, with the caveat that the feedback must be

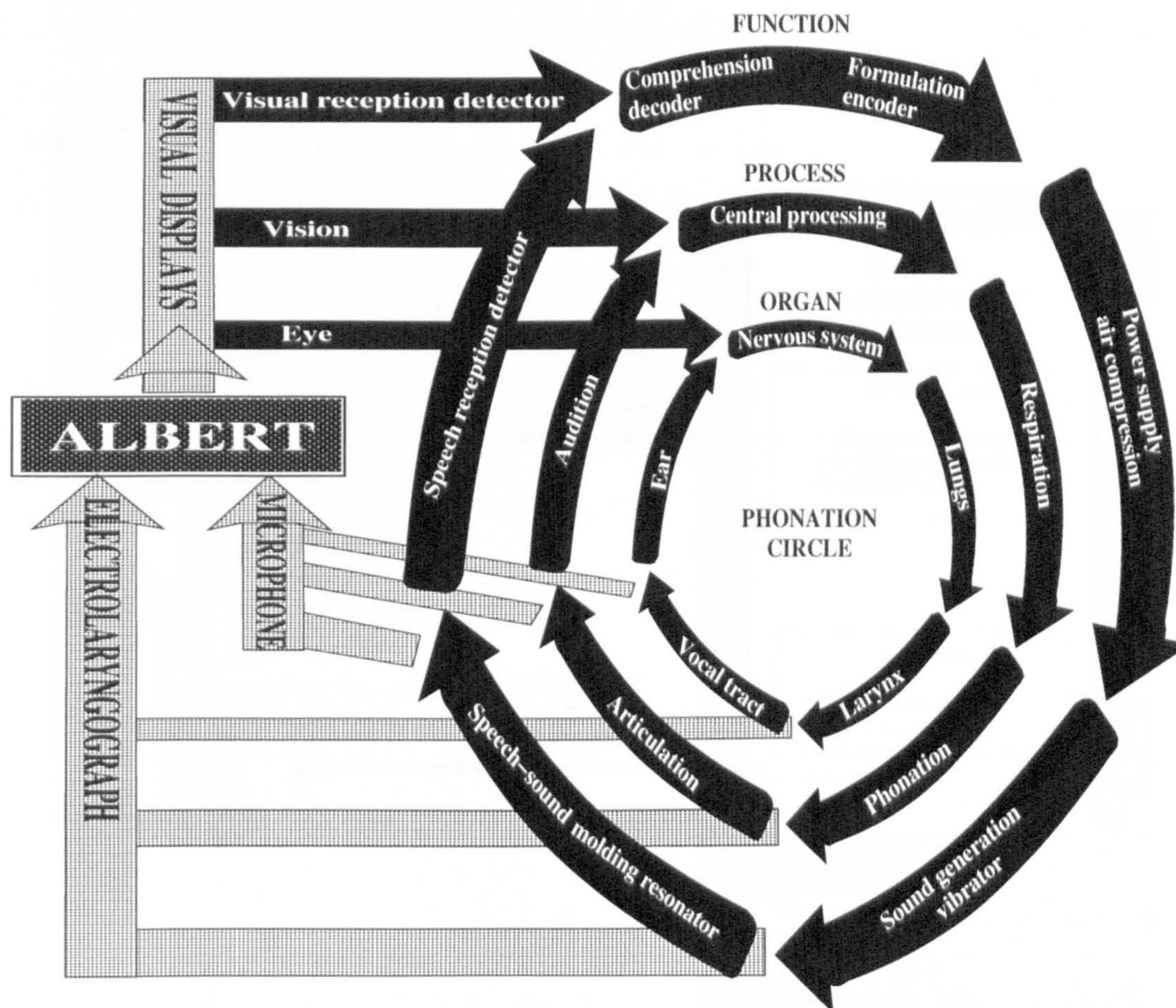


Figure 5.1: The phonation circle of Perkins & Kent (1986) with the domain of ALBERT superimposed

meaningful and readily understood.

5.3 ALBERT - Acoustic and Laryngeal Biofeedback Enhancement in Real-Time

As discussed in chapter 2, the process of speech or singing can be considered as a circular flow of information in which the acoustic product is perceived by the brain as a form of feedback. This enables any necessary adjustments to be made to the mode of phonation in order to better match a target state. A system called ALBERT (**A**coustic and **L**aryngeal **B**iofeedback **E**nhancement in **R**eal-**T**ime) has been developed to present visual feedback in real-time, or optionally in non-real-time

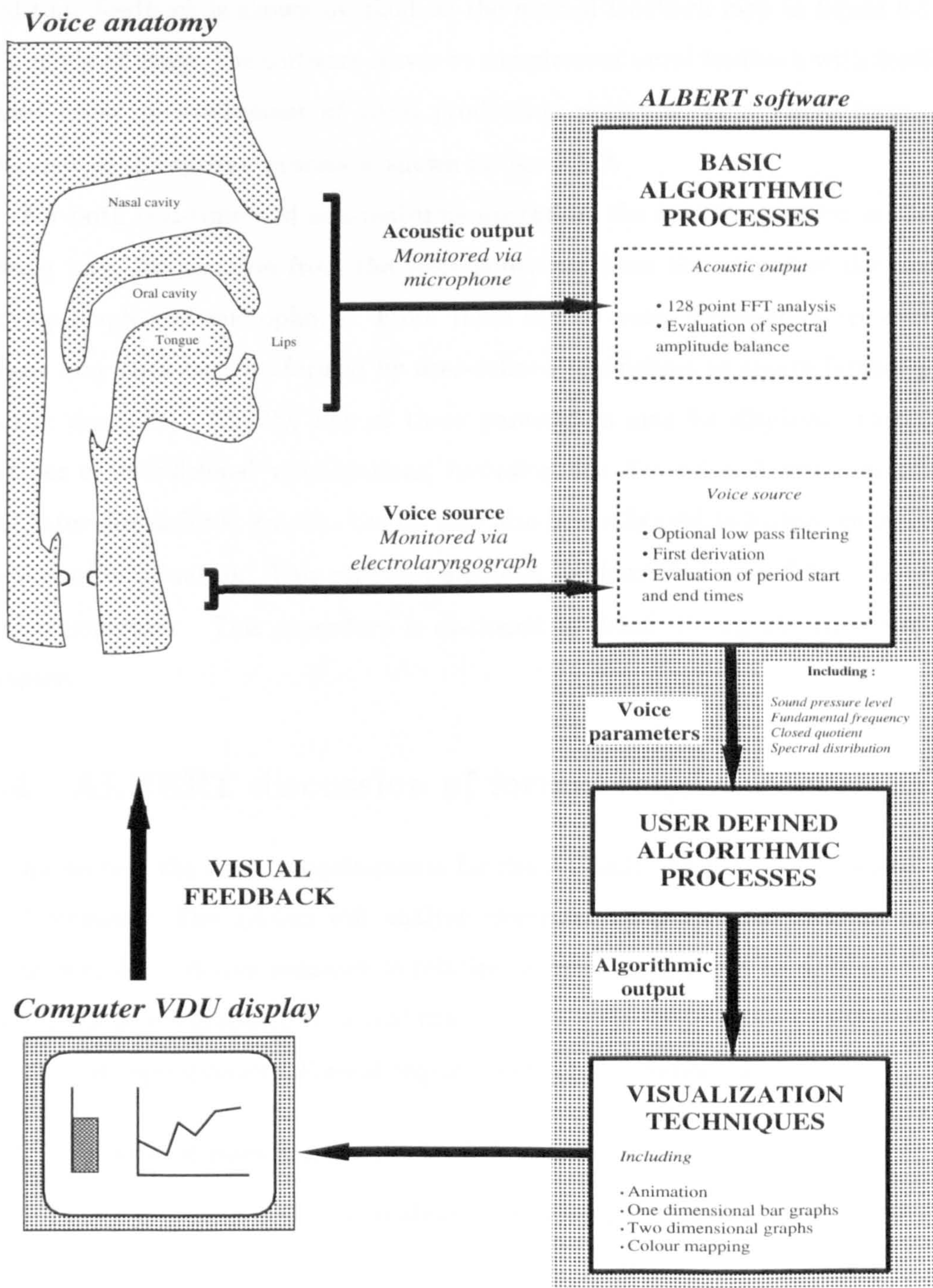


Figure 5.2: An overview of the ALBERT system process

when used for voice analysis only.

The domain within which the ALBERT software operates when employed for real-time feedback is shown overlaid on the normal feedback loop in figure 5.1. In this mode of usage, the software serves to complement aural feedback with feedback based upon an assessment of vocal production presented in a visual form. An overview of the system process is shown in figure 5.2.

For both real-time and non-real-time operation, the system operates in the following way. Information from the voice is derived from the output of the electro-laryngograph and microphone. From these signals voice parameters are derived. These may then be transformed by user-defined algorithms to create further parametric measures. Finally, any of these parameters may be displayed through a number of user-defined visualizations, including one dimensional, two dimensional and three dimensional graphs. Colour may also be configured to further enhance the display of information. This program has been implemented on a Silicon Graphics Indigo computer. This procedure is discussed in detail during the course of this chapter.

5.4 ALBERT discussion of formal requirements

In this section the formal requirements for the ALBERT system are discussed.

Overview : The system will analyse electrolaryngograph and acoustic microphone signals to derive parameters relating to voice usage, and be able to display these parameters graphically in real-time.

Formal requirements : Formal requirements of the systems are;

1. Analysis of parameters derived from the acoustic signal

That the system is able to analyse the following parameters from the acoustic signal:

- (a) the spectral distribution of the voice output (Ratio)
- (b) the loudness of the voice (SPL)
- (c) the change in loudness from one time instant to the next (shimmer)

2. Analysis of parameters derived from the electrolaryngograph signal

That the system is able to analyse the following parameters from the electrolaryngograph signal:

- (a) the fundamental frequency of the voice (F0)
- (b) the larynx closed quotient of the voice (CQ)
- (c) the rate of opening of the vocal folds (oa)
- (d) the rate of closing of the vocal folds (ca)
- (e) the peak-to-peak amplitude of the electrolaryngograph period (lx amp)
- (f) the change in amplitude of the electrolaryngograph period from one time instant to the next (lx shimmer)
- (g) the period-to-period change in fundamental frequency (jitter)

3. The graphical user interface

The user must be provided with easy-to-use controls over:

- (a) the processing and display of acoustic signal information.
- (b) the processing and display of electrolaryngograph signal information.
- (c) the properties of the voice parameters derived from the two input signals.
- (d) the way in which the state of the voice parameters may be visualized.

4. Visualization

The user must be provided with the ability to visualize the state of voice parameters in 1D, 2D and 3D display configurations. For reasons of clarity, a visualization must employ a minimum of 15% of the complete screen display.

5. Information on how to use the system

The user must have easy access to information on how to use the software.

6. Real-time performance

The system must be able to derive all voice parameters listed and visually present the state of any parameter at a frequency of not less than 10Hz.

In the following sections, the design and usage of the system created to satisfy the formal requirements is discussed in detail. At the end of the chapter (section 5.16) there is an assessment of the product with regard to each of the formal requirements.

5.5 Access to the ALBERT User Guide

A comprehensive user guide has been written for the ALBERT system. A copy is contained in Appendix E.

Access to the user guide and its contents has been designed to satisfy the following criteria:

1. **Centralised distribution** - in order that new versions of the manual may be easily accessed at a single distribution point.
2. **Hypertext implementation**- in order that the user may quickly and efficiently traverse the user guide contents.
3. **World access** - in order that anyone (with sufficient access) may access the user guide.

To satisfy these criteria, the user guide has been made available via the world wide web (www). It has been supported both in a format suitable for printing, so that a hard copy reference can be made by the user, and in a format suitable for efficient hypertext traversal.¹

5.6 Functional units

5.6.1 Introduction

In the design of ALBERT, the parameters that have been made available to the user for control over ALBERT functionality have been grouped into units according to

¹The user guide was written using the L^AT_EX(Lamport, 1986) formatting system. A version suitable for viewing on the www was obtained by converting the document using the 'latex2html' program (Drakos, 1993). This address at which the ALBERT user guide may be found is: http://www.york.ac.uk/~elec10/albert_manual.html. Both postscript and hypertext versions are presented. The postscript version can be printed out and used as a hardcopy reference. The hypertext version can be used to traverse, view and print specific sections of the manual.

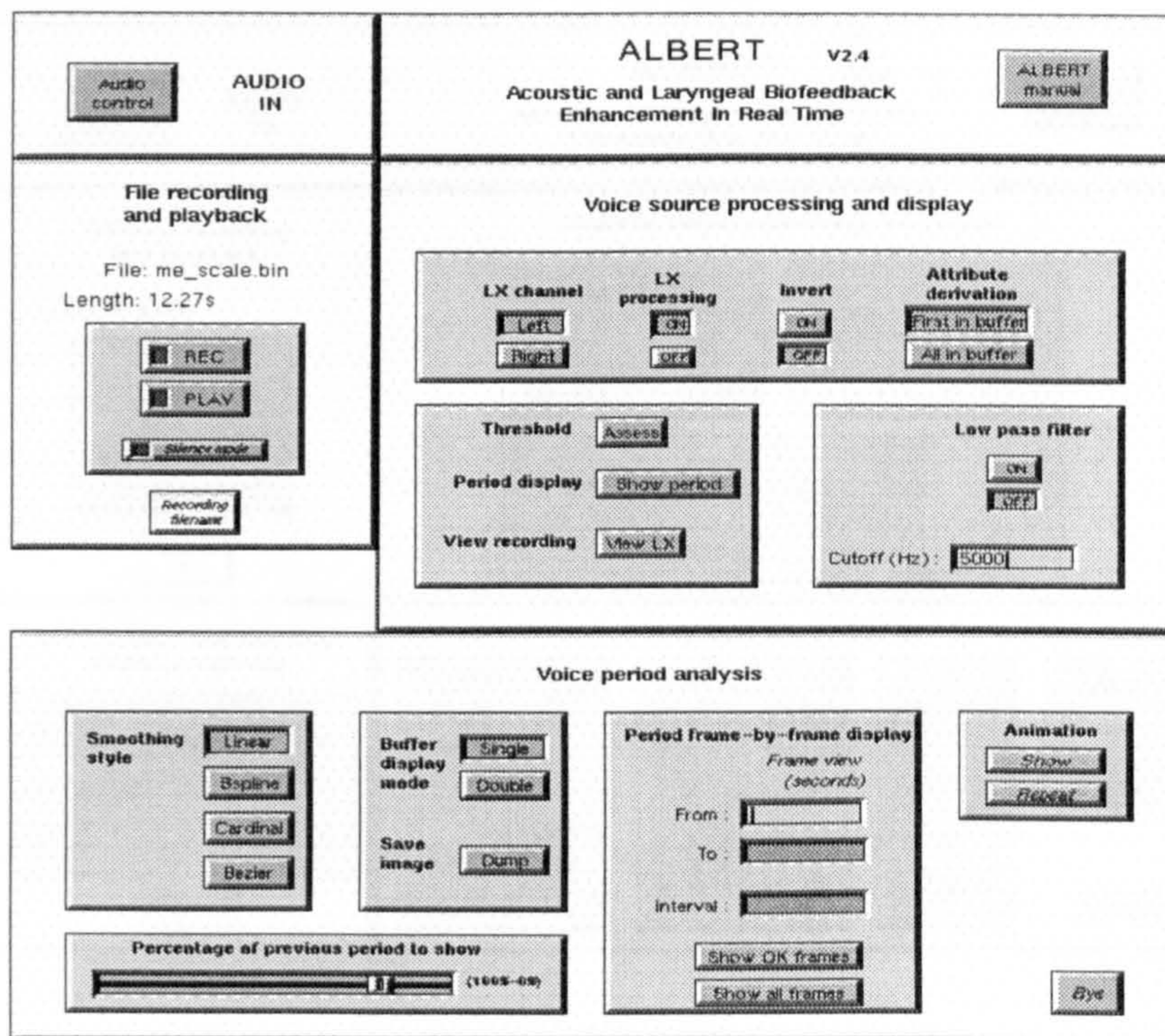
their function. These units are presented to the user on a number of *forms*. A form is a window in which a set of graphical user interface (GUI) controls such as buttons and sliders are presented to the user. For example, the user controls concerned with acoustic processing are presented on one form; the user controls concerned with voice parameter display and editing are contained on another form.

Several properties of the forms can be configured by the user. The user can control form position, height, width, and whether or not the component is to be displayed. In this way the display can be configured for maximum efficiency with regard to the context in which the system is being used. For example, if the user is only exploring visual feedback based on analysis of the voice source signal, then the forms which contain controls relating to acoustic monitoring are not required and can be toggled off. This is illustrated in the example shown in figure 5.3. Similarly, figure 5.4 illustrates an arrangement intended for efficient control of acoustic analysis alone.

5.6.2 An overview of the forms

A brief description of the 16 forms used by ALBERT follows.

1. **Header form.** This form provides a banner for the display of the name and version number of the software. It additionally provides the user with easy access to the ALBERT manual in electronic form.
2. **Audio form.** This form provides the user with an easy method to start the program used for audio input/output control.
3. **Acoustic form.** This form contains the main controls for acoustic processing.
4. **Acoustic and source form.** This form provides the user with the ability to change the processing and display update rate.
5. **File record form.** This form provides the user with easy control over the recording and playback of acoustic and laryngeal signals to file.
6. **Memory record form.** This form provides the user with easy control over the recording and playback of acoustic and laryngeal signals to memory.



(a) An example layout of some ALBERT controls. This example illustrates a possible configuration for voice source analysis.

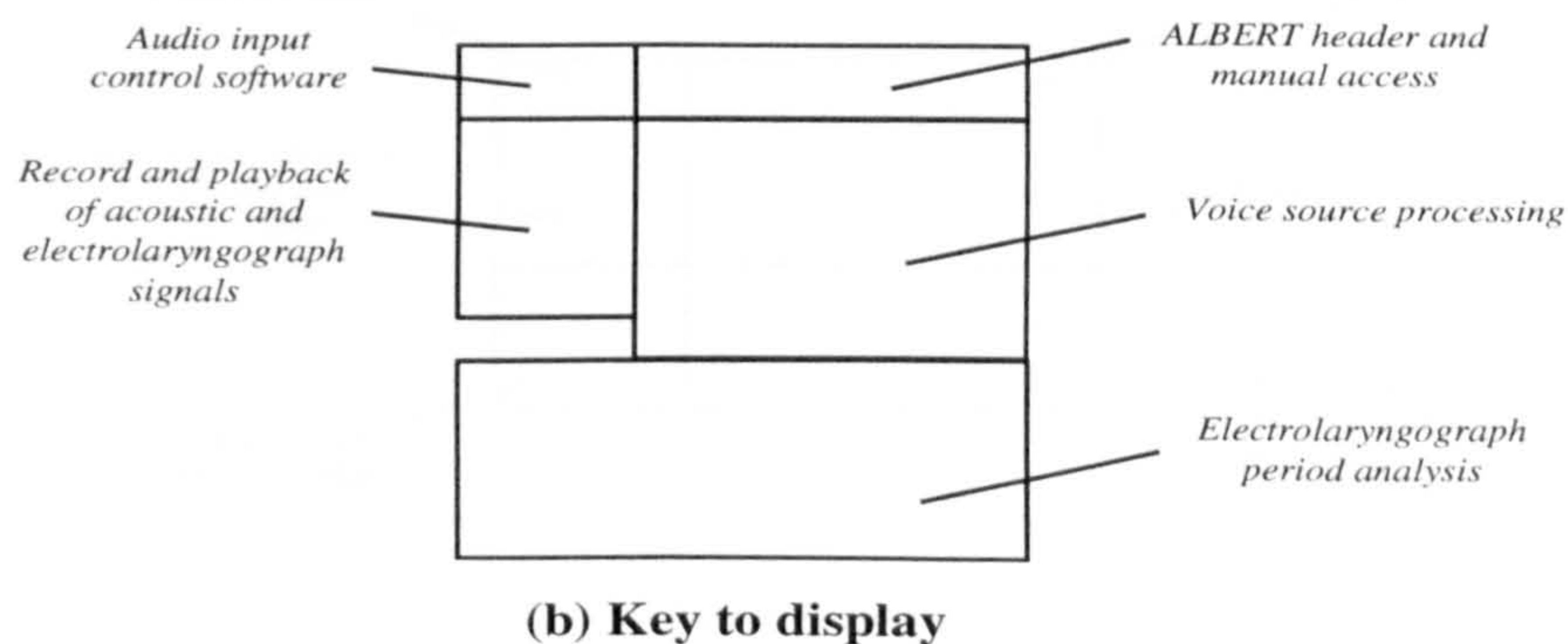
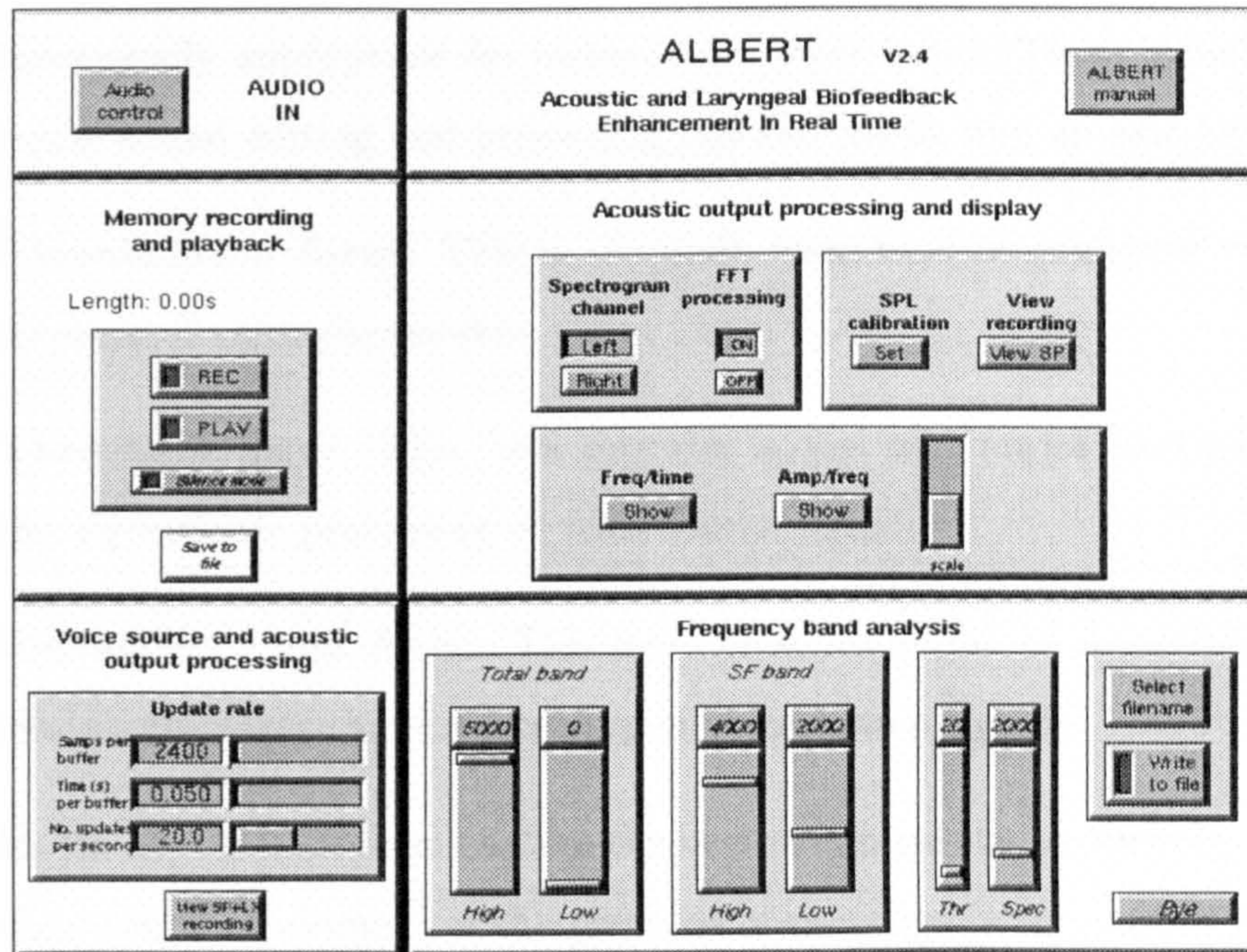
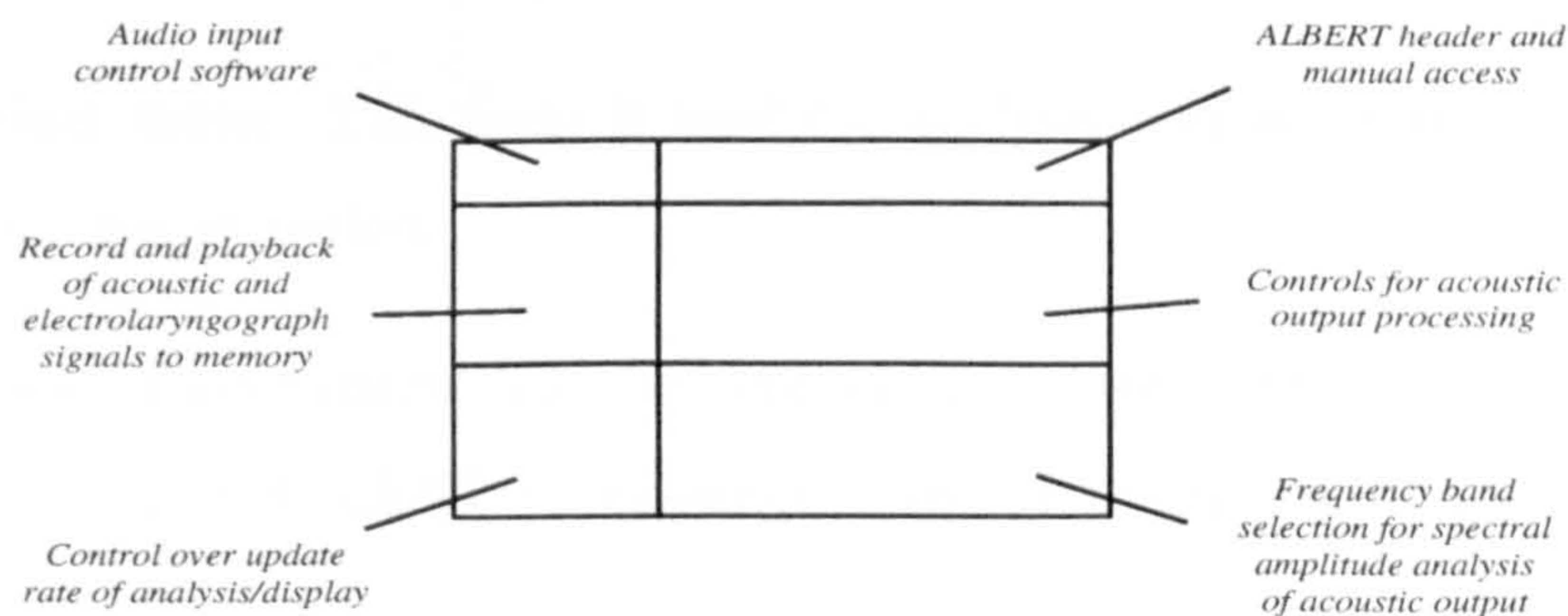


Figure 5.3: An example configuration of ALBERT for voice source analysis



(a) An example configuration of ALBERT for acoustic output analysis



(b) Key to display

Figure 5.4: An example configuration of ALBERT for acoustic output analysis

7. **Voice source form.** This form is the primary form for controlling the processing and display of voice source signals.
8. **Misc buttons form.** This form contains buttons which were judged as not functionally appropriate for inclusion on other forms. These include access to useful sound editing and processing software tools, and an exit button.
9. **Visualization form.** This is the main form used to configure visualization displays of voice parameters.
10. **Threshold form.** This form provides access to a major parameter used in the algorithmic processing of voice source signals.
11. **Refractory time form.** This form provides access to a second parameter used in the algorithmic processing of voice source signals.
12. **Ratio form.** This form enables the user to control the parameters used in the algorithmic assessment of the presence of the singer's formant in the acoustic signal.
13. **Params form.** This form is used to enable the user to view and alter voice parameter fields.
14. **Period form.** This form is used for analysis and animation of the electro-laryngograph period.
15. **Save and load form.** This form provides the user with the ability to configure part or all of ALBERT functionality through the saving and loading of files.
16. **Form visibility control.** This form provides a means for the user to control the visibility of each form.
17. **Form size control.** This form provides a means for the user to control the width and height of each form.

Each of these forms is discussed in context in the following sections. In the User Guide (Appendix E), a low-level description of the operation of each form element is discussed.

5.6.3 User control over form attributes

5.6.3.1 Introduction

Usually, the user is not provided with control over the user interface of a software program. ALBERT, however, has been designed so that the user is able to easily control some attributes of the user interface. Specifically, the user has control over the *visibility*, *width*, and *height* of each form.² In the next two sections the advantages of these features and the way in which they are accessed are discussed.

5.6.3.2 Form visibility

If all of the forms were presented on the screen at the same time the resulting display would be confusing and largely unmanagable because of the large screen area that would be used, and because of the many functions that would be presented to the user. The user is therefore provided with the ability to easily select whether each form is to be displayed on the screen, or is to be hidden from view.

The form through which this facility is accessed is shown on the left of figure 5.5. A sequence of toggle buttons are presented. Each button is labelled with the name of a form, and has an associated indicator light. If a form is currently displayed, the light is on. If the form is currently not displayed, the light is off. If a form is visible, the user may click on the button which is labelled with the name of the form. The form will then be removed from display. Alternatively, if the form is currently not on display (which would be indicated by the button light in 'off' mode) then a press of the button will cause the form to appear.

²The user can also control the position of each form. However, the way in which this is implemented is dependant on the windowing system used when ALBERT is executed, and control over this facility is outside the domain of ALBERT. For example, software running on Silicon Graphics computer systems are normally controlled through the '4Dwm' windowing system, and the control mechanism for moving a form within this system is engaged by default by holding down the ALT key and pressing the F7 key, but different window systems may have a different control mechanism for the same function.

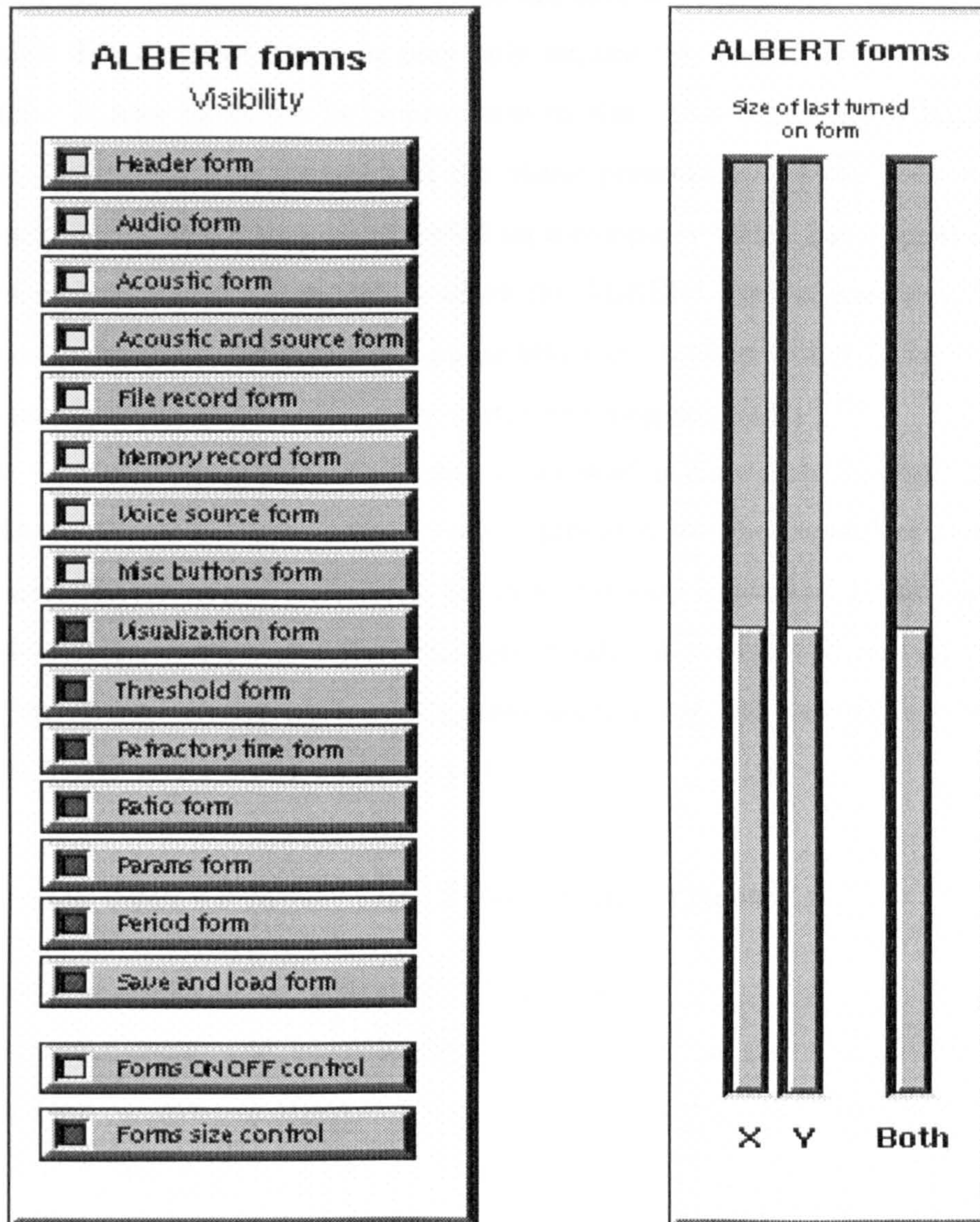


Figure 5.5: User control over form visibility, width and height. The control on the left can be used for managing the visibility of each form. The control on the right can be used for managing the width and height of each form.

5.6.3.3 Form size

In the design of ALBERT a method was designed so that the user could control the width and height of each form. There are several reasons why this facility can be useful. For example, the user may only require the display of a small number of forms. It may therefore be appropriate to size these forms much larger than normal, facilitating user control and the visual presentation of the user interface. Furthermore, ALBERT may be executed on a computer which has a greater visual resolution than that of the model on which the ALBERT system was developed. In this case the current form size will appear relatively smaller (other factors, such as the physical size of the visual display unit, remaining constant).

The form through which this facility is accessed is shown on the right of figure 5.5. Three sliders are provided. Two are provided for the control of the x axis and y axis dimensions. The third is provided for easy control of both dimensions in combination. The latter enables the aspect ratio of the form to remain constant regardless of any change in size. Changes made using the size control form are addressed only to the last form displayed.

5.7 An overview of the flow of information

The flow of information is illustrated in figure 5.6.

Both voice source and acoustic signals undergo unique forms of analysis. From this analysis, a number of voice parameters are derived. Two forms of display particular to each form of signal are supported.

The voice parameters derived during this process are then combined into a single list. The user may control the various parameter properties and the way in which the parameter information may be displayed in visualization configurations.

In the following sections, this process is discussed in full.

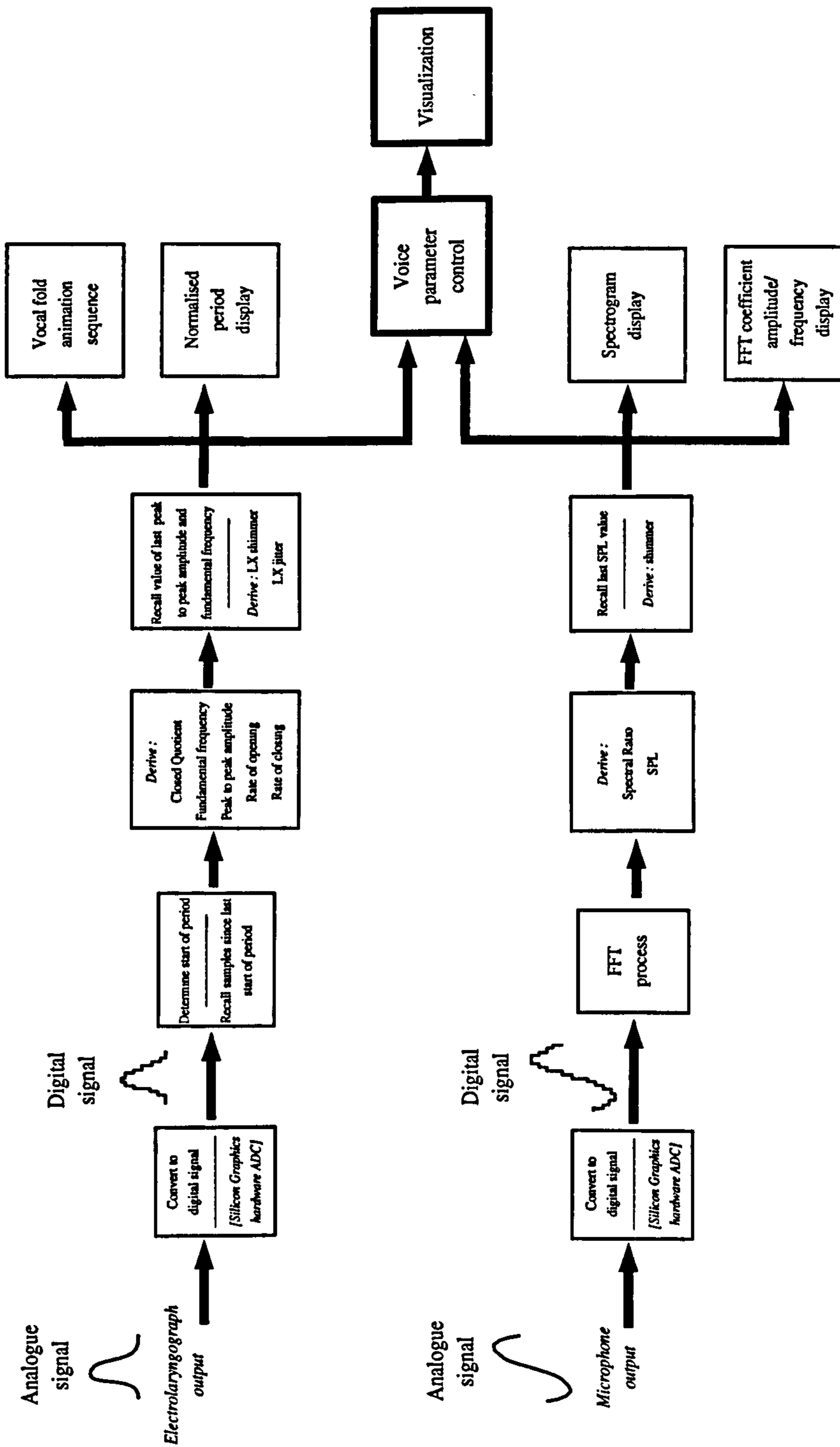


Figure 5.6: Overview of the system process. Key: ADC = Analogue to digital converter

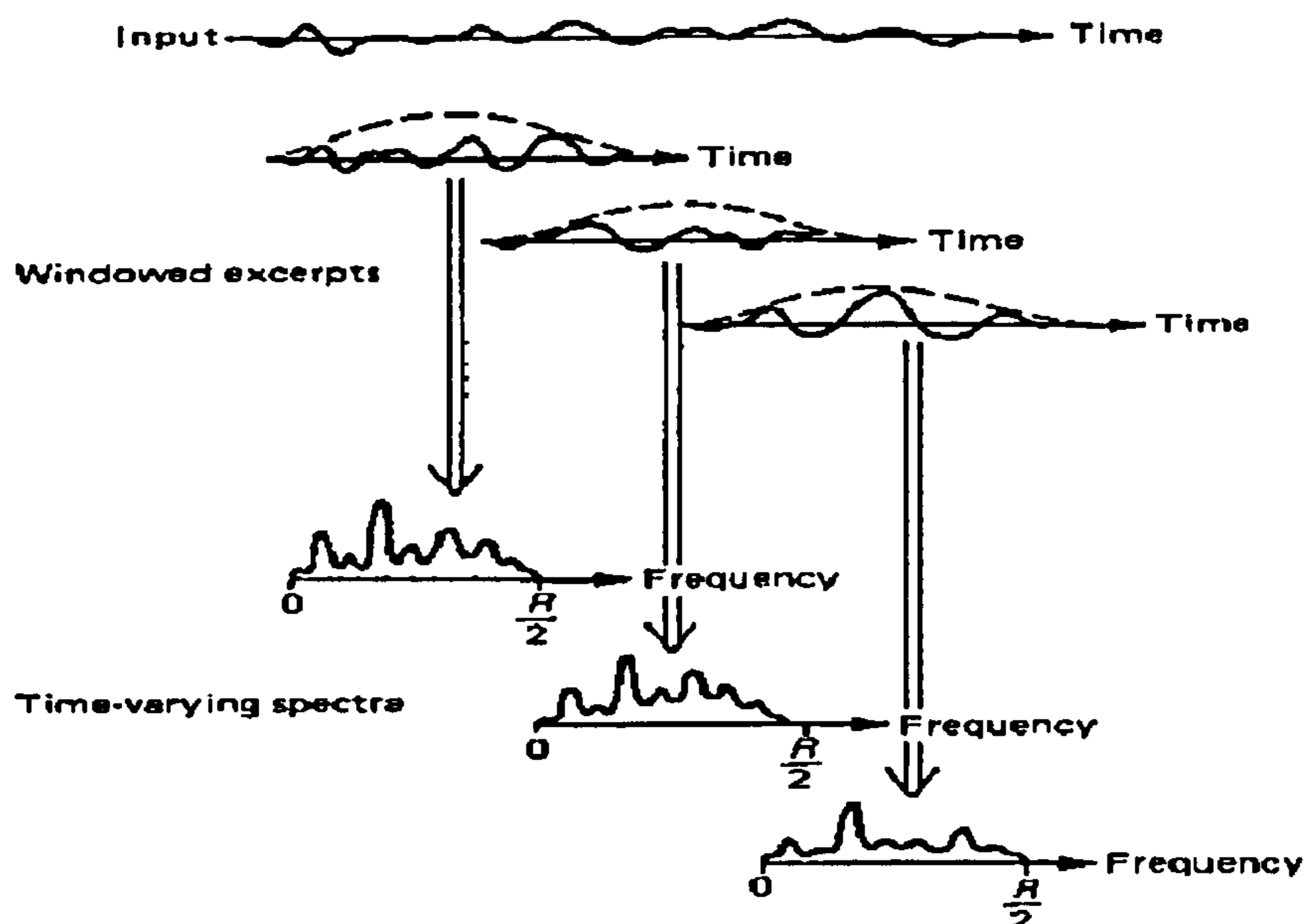


Figure 5.7: Fourier transform on a sequence of windows. R represents the sampling frequency of the signal. $R/2$ is called the Nyquist frequency. It is the theoretical limit on the highest frequency that can be represented in a sampled digital signal (From Dodge and Jerse, 1985).

5.8 Analysis of the acoustic signal

5.8.1 Deriving a measure of the singer's formant

The display of a single parameter representative of the singer's formant phenomenon presented in a feedback context other than the display of the entire acoustic spectrum was previously identified as likely to be a useful attribute for analysis and for provision as a real-time voice tuition system (section 3.4.4). To achieve this, a method used to evaluate the phenomenon of the singer's formant objectively from the digital representation of the acoustic signal was developed as follows. Prior to real-time operation, two spectral ranges were identified. The first indicates the entire spectral range for which changes in the acoustic output of the voice can be reliably detected (0-5kHz). The second spectral range covers the frequency range where the singer's formant is generally located (2kHz-4kHz).

A 128 point Fast Fourier Transform (FFT) (Cooley & Tukey, 1965) is carried out on the accumulated speech for each buffer of data. This process is illustrated in

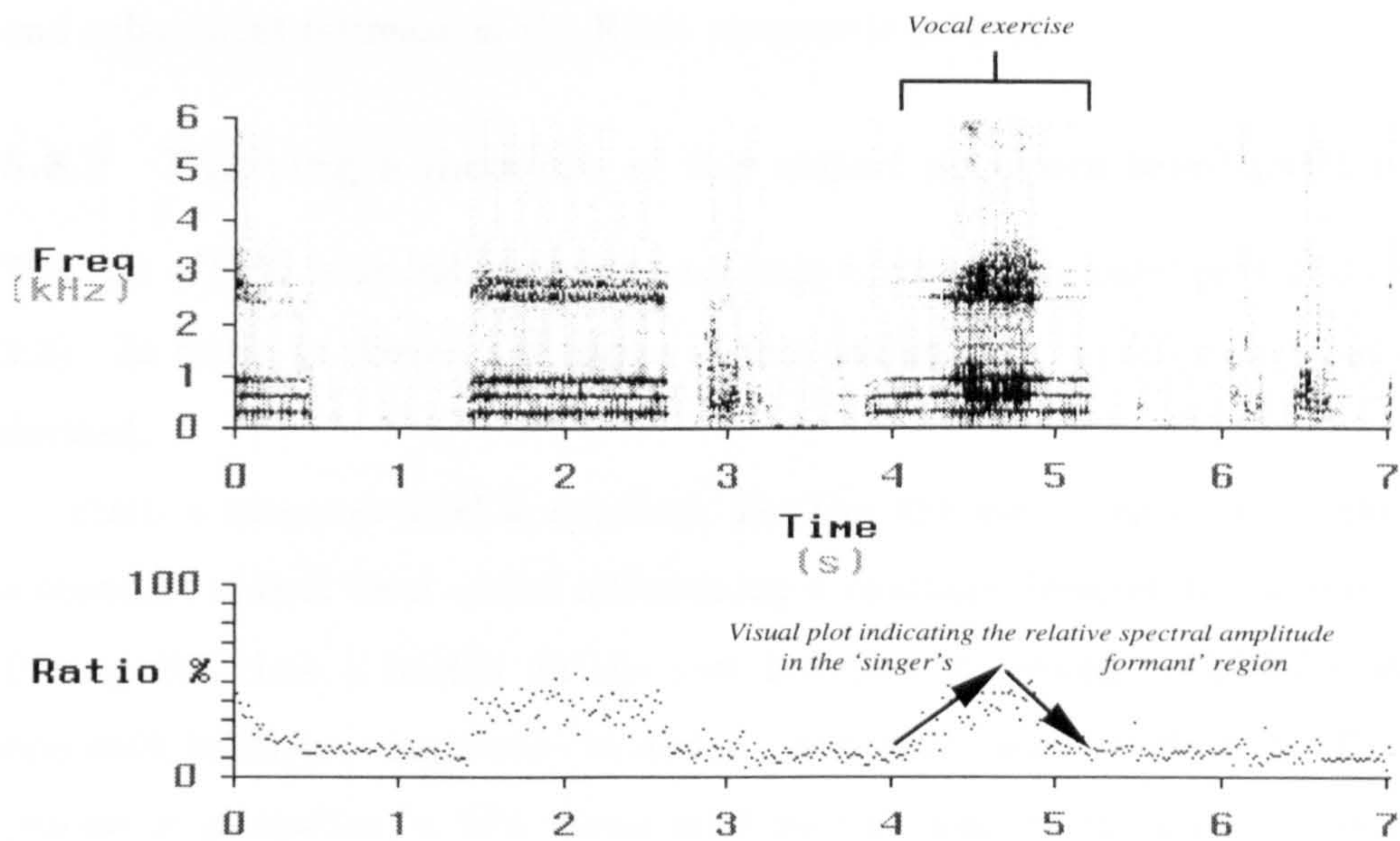


Figure 5.8: Spectrogram of a trained singer, with time aligned graph measure of the accumulated fft coefficient amplitudes in the region 2-4KHz relative to the accumulated amplitudes across the range 0-5KHz.

figure 5.7.

This produces an array of coefficients corresponding to the amplitude levels in the acoustic spectrum. The coefficient values for both pre-determined ranges are summed. A measure of the relative amplitude level of spectral partials in the singer's formant region (labelled *Ratio*) can then be derived as the ratio of the total fft coefficient amplitudes present in the singer's formant range (labelled A_{sf}) to the total fft coefficient amplitudes present in the spectral range used by the human voice as a whole (labelled A_{total}). This index is then multiplied by 100 in order to produce a percentage measure. In summary;

$$Ratio(\%) = \frac{A_{sf}}{A_{total}} * 100$$

A simple indication of the way in which this measure fluctuates with vocal quality is shown in figure 5.8. The area labelled 'vocal exercise' indicated in the figure illustrates the relative increase of spectral amplitudes in the singer's formant region during the expression of a particular vocal exercise in which good vocal quality is

gradually 'brought in' to the sound and then released.³ A corresponding increase and subsequent decrease in the Ratio measure is evident.

5.8.2 Deriving a measure of the sound pressure level (SPL)

The role of SPL as an indicator of vocal usage has been discussed previously (section 3.3). In order to derive a measure of this parameter, the following method was devised.

First, a reference level is required. For this the user is required to phonate at a constant output level whilst maintaining a constant distance to the microphone. During this time a button on the user interface is pressed. ALBERT will then sum each coefficient amplitude to derive a reference value, labelled A_0 . During the process of phonation, a SPL meter must be held next to the microphone in order to record the level at that position. This value is now entered by the user into ALBERT. It is labelled SPL_0 .

During acoustic processing, each fft coefficient amplitude is summed to derive a total amplitude value (A). The increase in sound level (labelled SPL_i) is then derived by the following formula (Blitz, 1964).

$$SPL_i(dB) = 20 \log_{10} \frac{A}{A_0}$$

To get the current dB level, the SPL_i value is simply added to the reference value originally entered by the user.

$$SPL(dB) = SPL_0 + SPL_i$$

Assuming the distance between microphone and subject remains constant, ALBERT will subsequently supply a correct SPL level reading. This is expressed as the 'spl' variable.

A measure of the stability of the SPL data, known as *shimmer* was also discussed previously (section 3.3).

This is derived on a buffer-by-buffer basis by subtracting the current SPL value with the previous SPL value.

$$shimmer = SPL_n - SPL_{n-1}$$

³This example exercise, from the *Bel Canto* school of singing, is called the *Mesa Di Voce*.

5.8.3 Display of acoustic signal information

Information derived from the acoustic signal can optionally be displayed as both a spectral cross-section and a spectrogram. Two examples of each are shown in figure 5.9.

5.8.4 User control over acoustic signal analysis

5.8.4.1 Introduction

There are two forms which support controls concerning acoustic signal analysis. The first contains high-level controls, such as the selection of channel and a control for turning on and off the fft process. The second contains sliders for control over the algorithm used to derive the 'Ratio' parameter, such as the frequency range boundaries.

5.8.4.2 Primary control form

This form is shown in figure 5.10.

It contains the main controls for acoustic processing. These are :

1. channel selection.
2. turning on the fft algorithm.
3. SPL calibration (section 5.8.2).
4. The viewing of a recording. This facility enables the user to zoom in and play or print part or all of the signal sequence.
5. Access to a spectrogram display.
6. Access to a display of spectral amplitude across frequency.
7. The ability to vertically scale the display of fft coefficient amplitudes in the Amp/freq window.

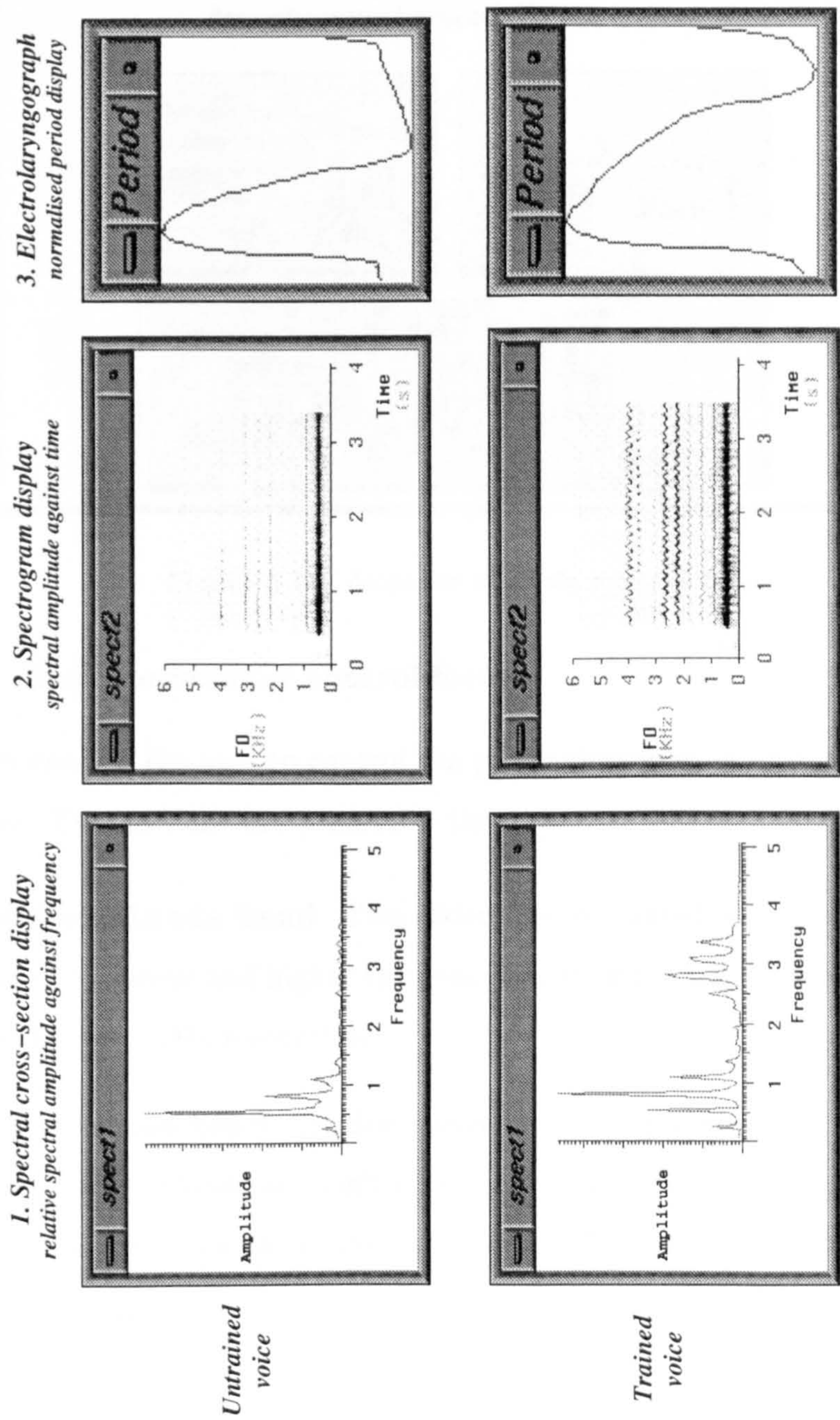


Figure 5.9: The three standard display modes for voice source and acoustic output signals illustrated for 'trained' and 'untrained' modes of phonation

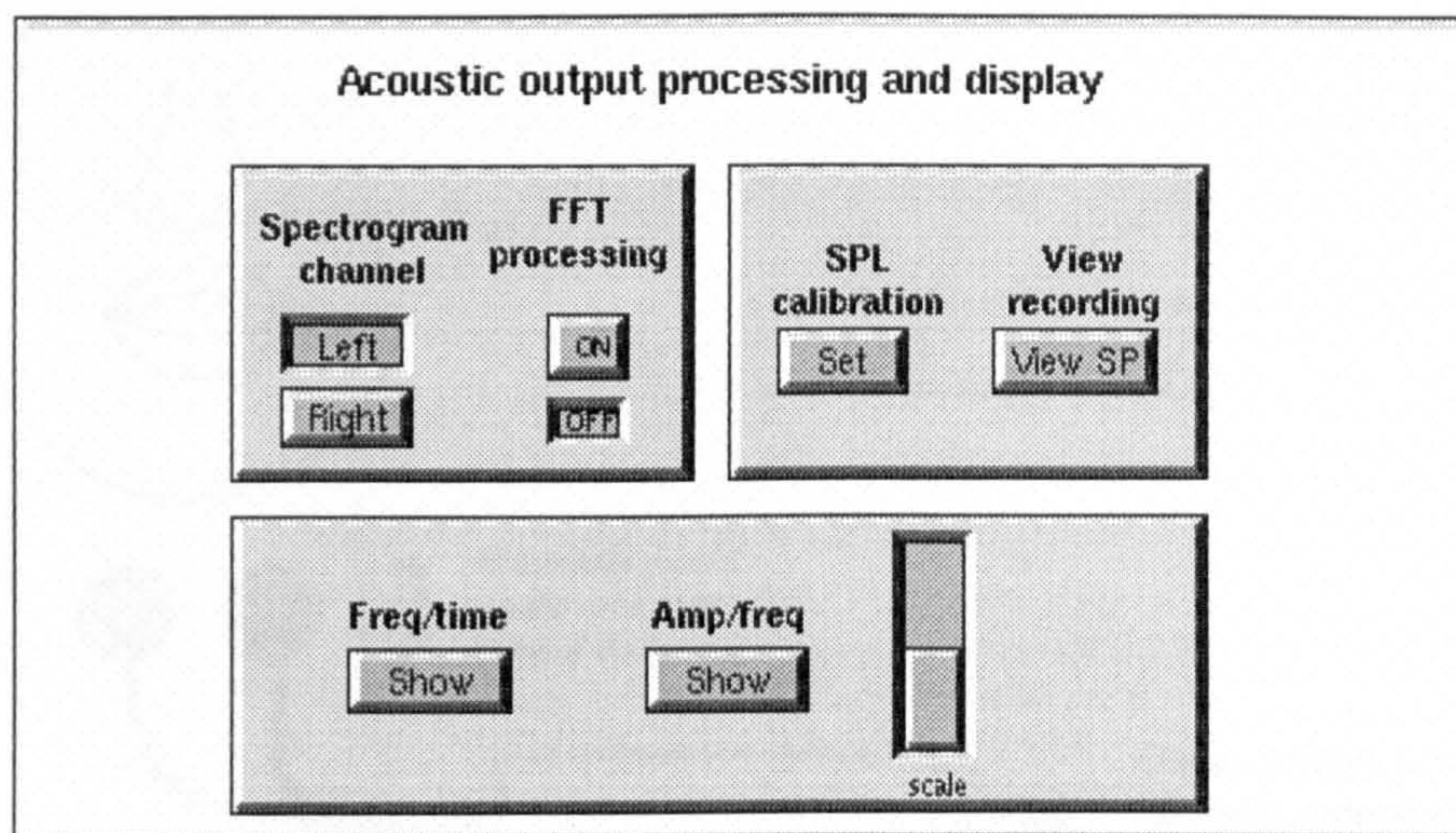


Figure 5.10: Acoustic analysis control form

5.8.4.3 Ratio parameter control form

This form enables the user to control the parameters used in evaluating the 'Ratio' parameter. Two sections are present in the form;

1. **Total amplitude band.** Two sliders are presented with which the user can control the lower and higher frequency limits of the larger band. These default to 0Hz and 5KHz respectively.
2. **SF amplitude band.** Similarly, two sliders are presented with which the user can control the lower and higher frequency band within which the phenomenon of the singer's formant is known to occur. These default to 2KHz and 4KHz respectively.

5.9 Analysis of the electrolaryngograph signal

5.9.1 Introduction

In order to design a system that is able to provide feedback on the vocal activity of a subject, a decision needs to be made on the most appropriate method of analysing the vocal actions of the subject.

One method for deriving the closed quotient of the laryngeal period is through

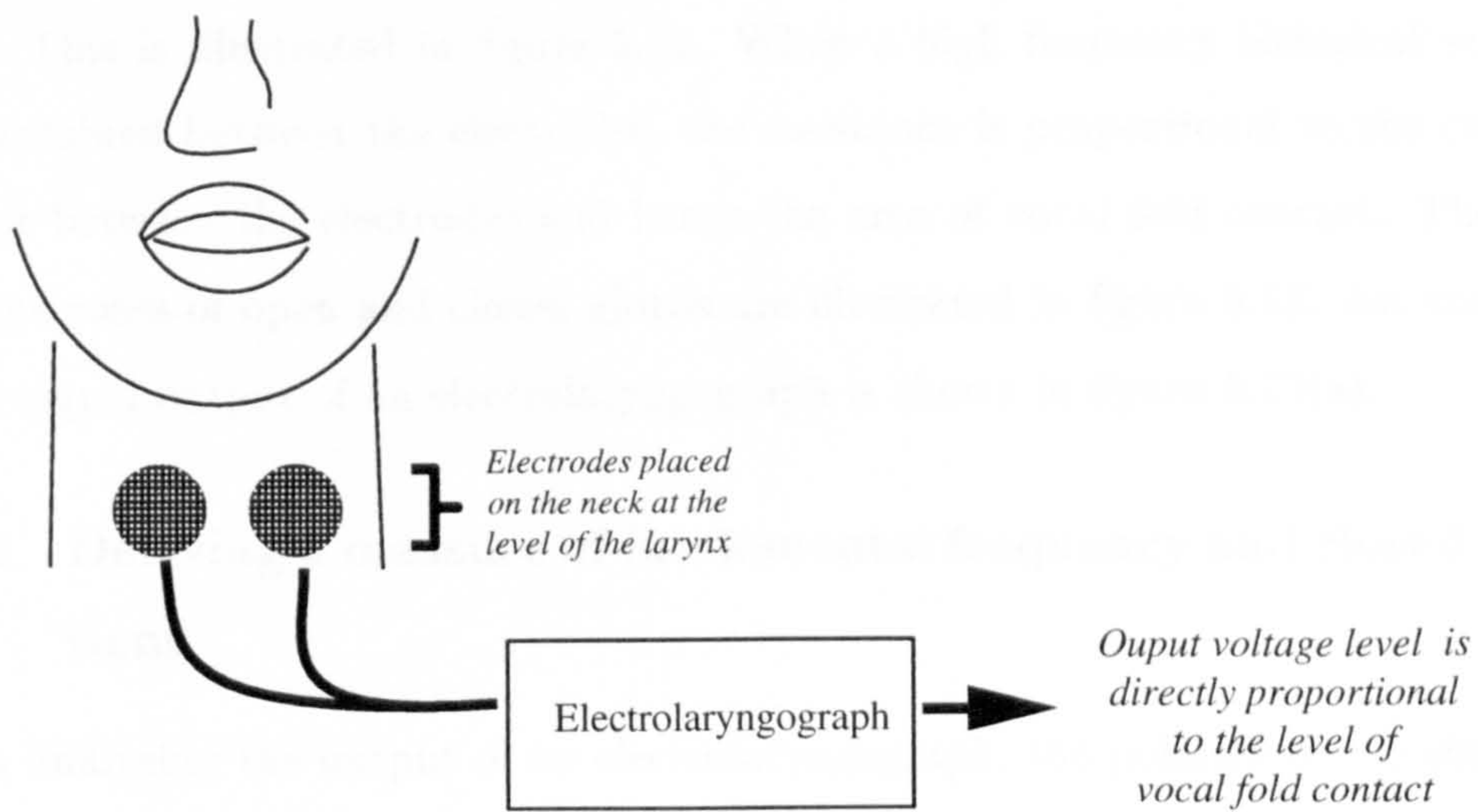


Figure 5.11: Use of the electrolyaryngograph

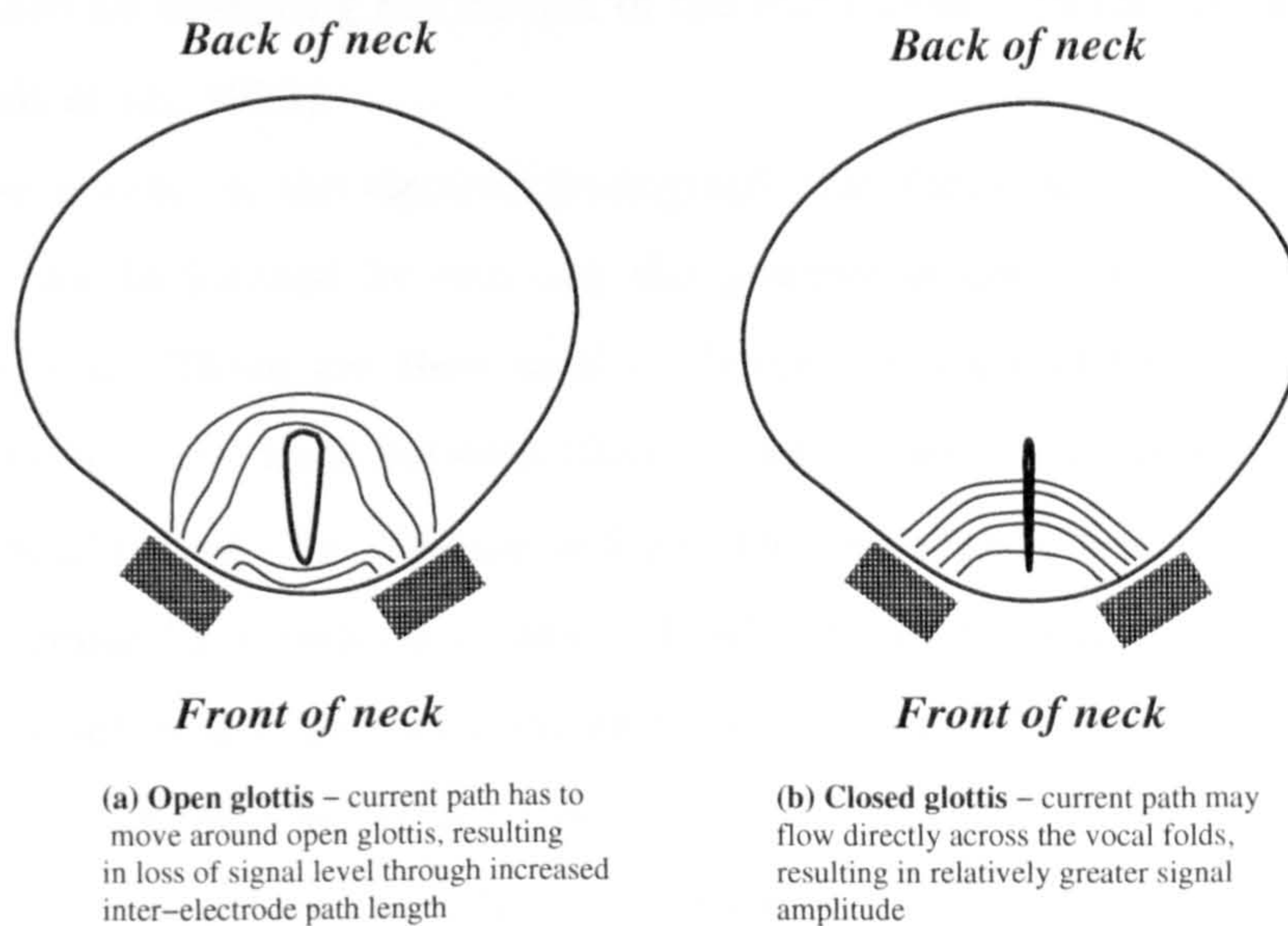


Figure 5.12: Cross-section of the neck at the level of the larynx indicating current path across the vocal folds for open and closed glottis (After Garner, 1993)

the use of an electrolaryngograph (Abberton et al., 1989). A pair of electrodes is placed on either side of the thyroid cartilage at approximately the level of the vocal folds. This is illustrated in figure 5.11. When a high frequency electrical voltage is maintained between the electrodes, the resistance is proportional to the current flowing between the electrodes and hence the area of vocal fold contact. The two extreme cases of open and closed glottis are illustrated in figure 5.12. An example of the signal output of an electrolaryngograph is shown in figure 5.13(a).

5.9.2 Deriving a measure of fundamental frequency and closed quotient

When analysing the output of an electrolaryngograph, the polarity of the signal is first checked to ensure that positive changes reflect increased inter-electrode current flow. The detection of the start of the closed phase is based on the assumption that the vocal fold contact area changes more rapidly when it is increasing than when it is decreasing (that is, the folds snap together more rapidly than they part) (Davies et al, 1986). This relationship has been verified by several studies. For example, this was verified by observing the motion of the vocal folds with the aid of a stroboscope (Askenfelt et al., 1980).

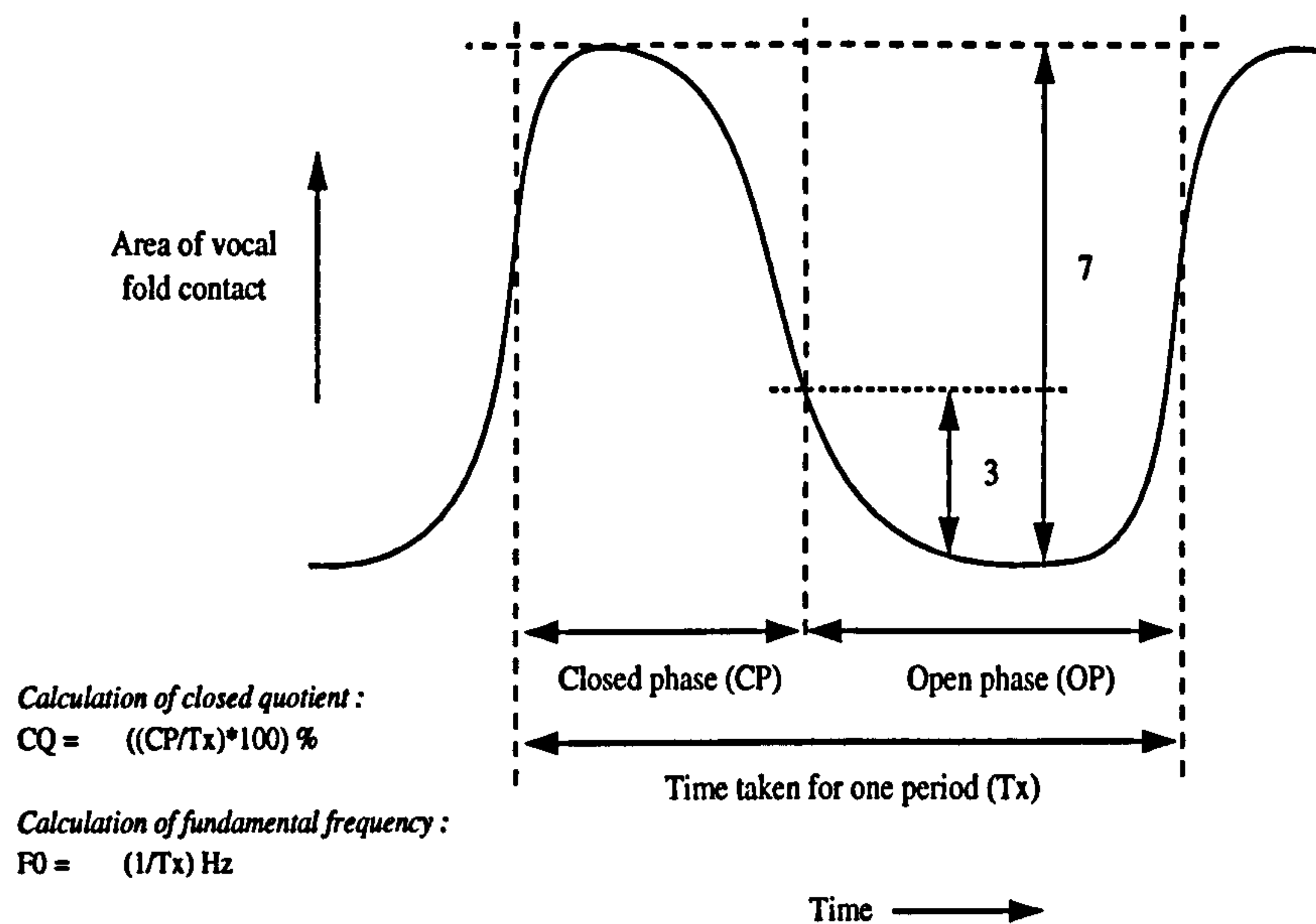
These points on the electrolaryngograph waveform (which is often referred to as 'Lx') can be located by deriving the positive peaks in the time differentiated Lx waveform. These are then used to define the start of the closed phase (CP) in each cycle. The time between these peaks is used to provide a measure of the fundamental period (Tx). See figure 5.13. The end of the CP is the instant when the negative-going Lx waveform crosses a fixed ratio of the current cycle's amplitude. The ratio, set at 3:7, has been shown to exhibit a result close to that obtained by inverse filtering (Howard et al., 1990).

Electrolaryngographically derived CQ is then obtained by the following equation:

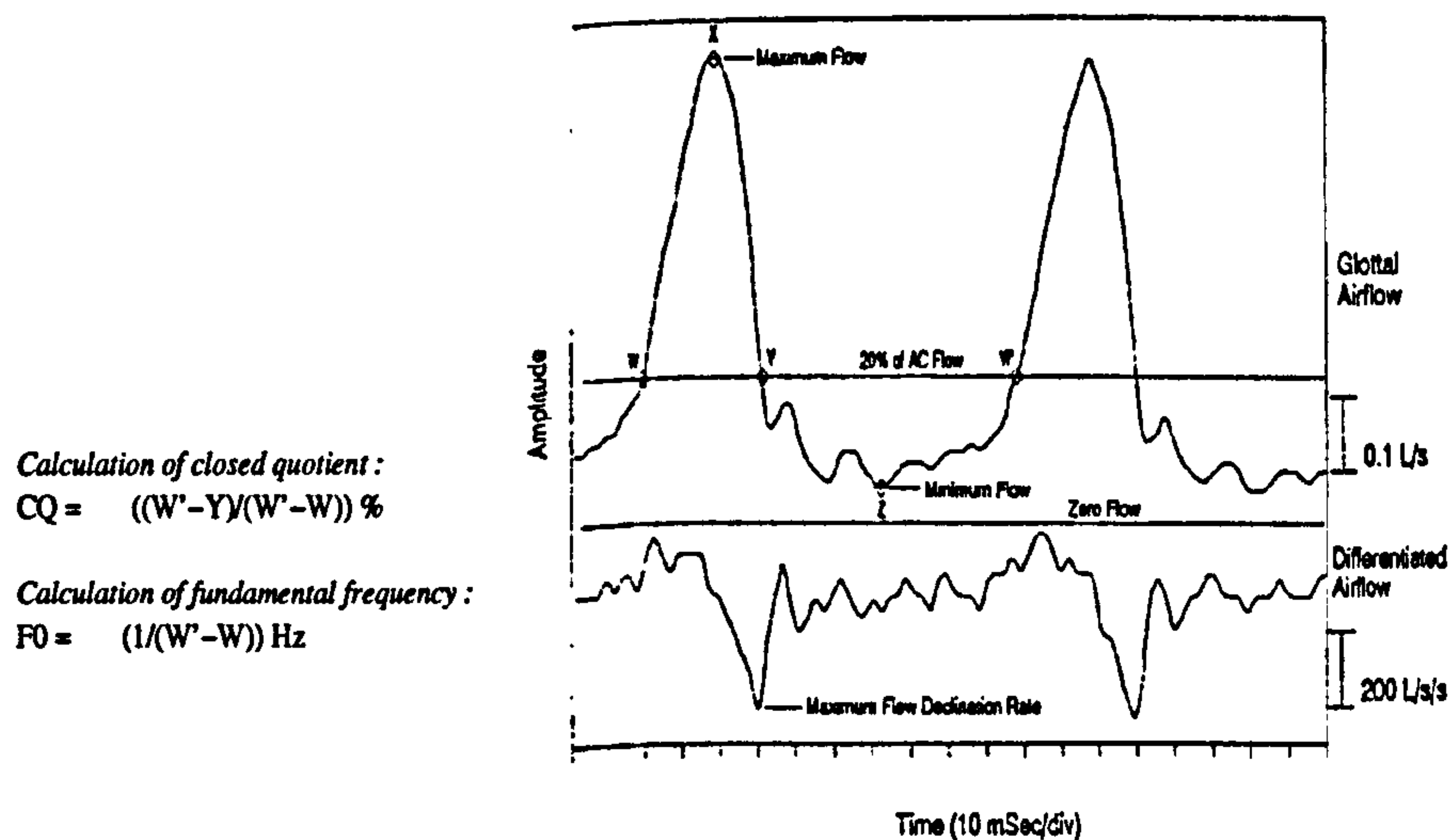
$$CQ(\%) = \left(\frac{CP}{Tx} * 100 \right)$$

This is illustrated in figure 5.13(a).

A measure of fundamental frequency and closed quotient is also available from analysis of the airflow produced during phonation. This can be obtained from a flow



(a) The derivation of Closed Quotient (CQ) from an electrolyngograph signal



(b) The derivation of Closed Quotient (CQ) from an airflow waveform. As for (a), the maximum flow declination rate is defined from the greatest negative peak of the differential glottal airflow waveform. (From Stathopoulos & Sapienza, 1993)

Figure 5.13: The derivation of fundamental frequency and closed quotient from electrolyngograph and airflow waveforms

mask (Rothenburg, 1973) which is placed over the face of the subject. All air intake and exhalation then takes place through the mask, enabling precise monitoring. The air flow signal is first inverse filtered to counter the effects of the supralaryngeal vocal tract on the voice source signal. The method used to subsequently derive CQ is almost identical to that for deriving CQ from the electrolaryngograph output described previously, although a slightly different ratio constant is used in the definition of the end point of vocal fold closure. This is illustrated in figure 5.13(b). However, due to the nature of the flow mask design, the acoustic signal produced by the subject is altered prior to reaching his/her ear. This introduces a high level of error into the acoustic feedback signal, which is probably the most important feedback parameter for the process of vocal tuition. An example of disquiet expressed by two professional singers may be found recounted in Sundberg & Gauffin (1979; pp. 304). The consequences of changing or depriving the subject of the sound they produce are hard to formalise (Ward & Burns, 1978) but are clearly unacceptable liabilities to a voice tutoring process.

5.9.3 Deriving a measure of the peak-to-peak signal amplitude

Most forms of electrolaryngograph design employ some form of automatic gain control (e.g., see Baken, 1987). This includes the electrolaryngograph used for the studies presented in this thesis. However, if the design is altered to remove this, or if an early model is used which does not have this design feature, then the peak-to-peak amplitude of the electrolaryngograph period will change in proportion to the distance of the vocal folds from the electrodes in the vertical plane (other factors remaining constant) (see, for example, Gilbert et al., 1984). A measure of the peak-to-peak amplitude might then be useful in several voice tuition contexts. For example, trained singers are known to exhibit markedly less variation in larynx height than untrained singers (Sundberg, 1987). For this reason, a measure of the peak-to-peak period amplitude is provided as one of the parameters supported by ALBERT.

A measure of the associated level of shimmer is also supplied.

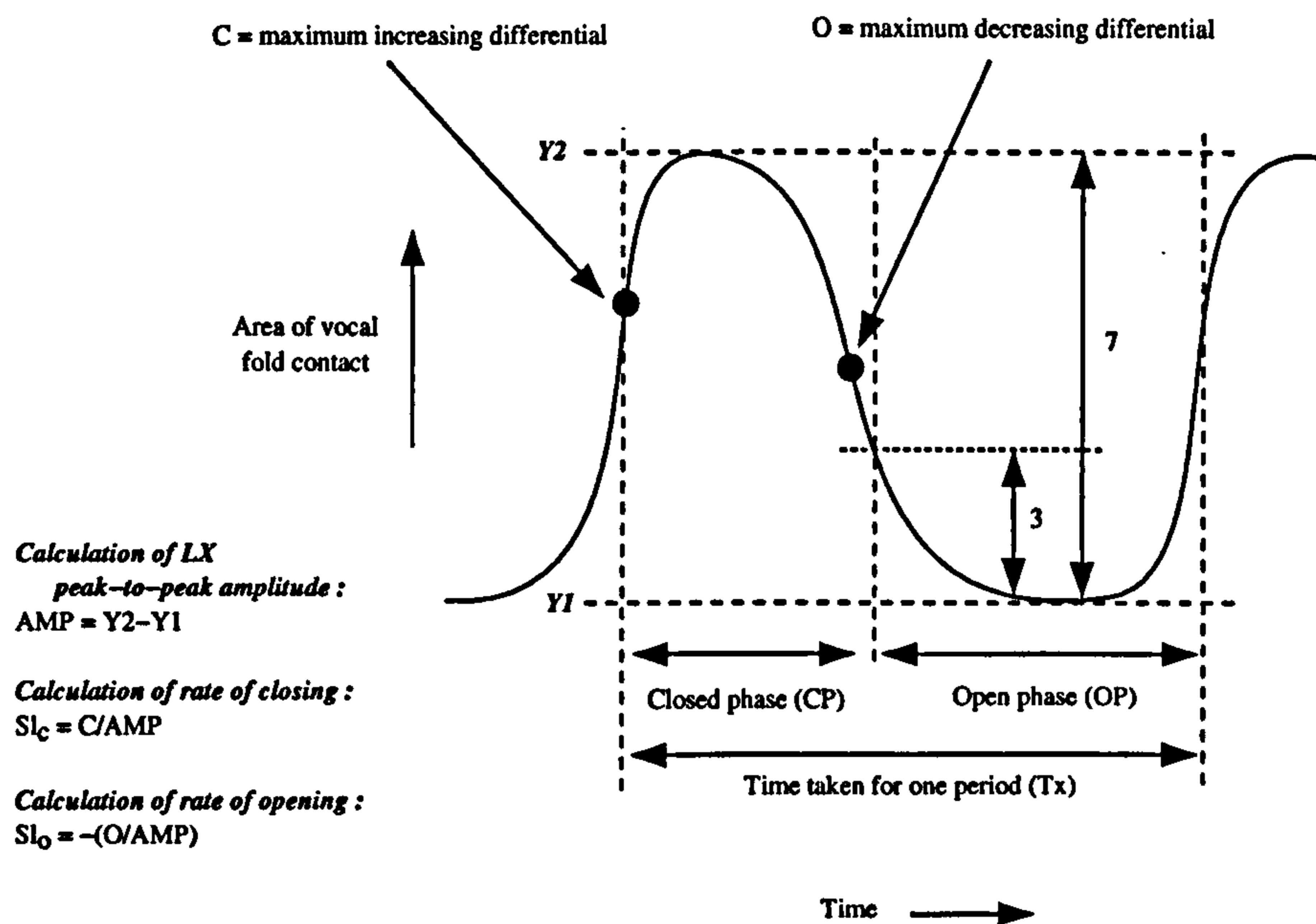


Figure 5.14: Deriving a measure of the rate of larynx opening and closing from the electrolaryngograph waveform

5.9.4 Deriving a measure of the rates of larynx opening and closing

In chapter 2 it was established that the *rates* of larynx opening and closing were of importance in vocal function. A measure of these may be derived by obtaining a value of the maximum and minimum derivatives of the electrolaryngograph period. These values are further divided by the peak-to-peak amplitude of the period to take into account the recording level. This is illustrated in figure 5.14.

These two parameters may be useful in the process of voice monitoring. They may be beneficial when presented as real-time feedback for voice tuition. However, as discussed previously in section 3.7, there appears to be no previous research in which comparable parameters are used in the assessment of the developing voice. The interpretation of these parameters must therefore be tentative and cautious. For example, the model of the vocal folds presented by Childers & Krishnamurthy (1985) demonstrated that an increase in the *angles* of opening or closing of the vocal folds would result in a change to the electrolaryngograph signal (Childers et al., 1987), although the authors concluded that a measure of the differential of the signal was still appropriate for determining instants of opening and closure.

5.9.5 The effect of quantisation

The rate at which ALBERT samples electrolaryngograph (and acoustic) signals is maintained at 48kHz (which is the highest sampling rate that the hardware is capable of). The Silicon Graphics audio hardware incorporates anti-aliasing filters to ensure that only frequencies lower than the nyquist frequency are sampled.

However, because of the quantisation of the signal, a high frequency signal is poorly represented by the digital sampling process when compared to the sampling of a low frequency signal. This contrast is illustrated in figure 5.15. At a frequency of 125Hz, one period will be represented by 384 samples. This resolution has an associated maximum error of 0.325Hz, which corresponds to 0.26%. In contrast, at a frequency of 1000Hz, one period will be represented by 48 samples. This has an associated maximum error of 20.8Hz, which corresponds to 2.08%. Clearly, the effects of quantisation at high frequency must be noted when considering the results of ALBERT assessment, although phonation at such high fundamental frequencies is normal for only a small subset of voice users (such as soprano singers).

5.9.6 Display of the electrolaryngograph signal

5.9.6.1 Normalised period display

As discussed previously, the electrolaryngograph signal is analysed in order to determine the laryngeal period start and end points. This information is required for the subsequent derivation of the voice source parameters, which is carried out on a cycle-by-cycle basis. The portion of electrolaryngograph signal between these two points can optionally be displayed with signal amplitude and length of time normalised to fit a window of arbitrary height and width. Two examples are shown in figure 5.9. This display enables the viewer to easily visualize the relative closed and open phases of the larynx period regardless of signal amplitude or frequency.

During use of the normalised period display, it was realised that the ability to see the electrolaryngograph signal immediately prior to the currently displayed period signal would be of further aid in visualizing the performance of the electrolaryngograph signal. A simple slider mechanism was provided for the user to select the extent to which the electrolaryngograph signal preceding the period is to be dis-

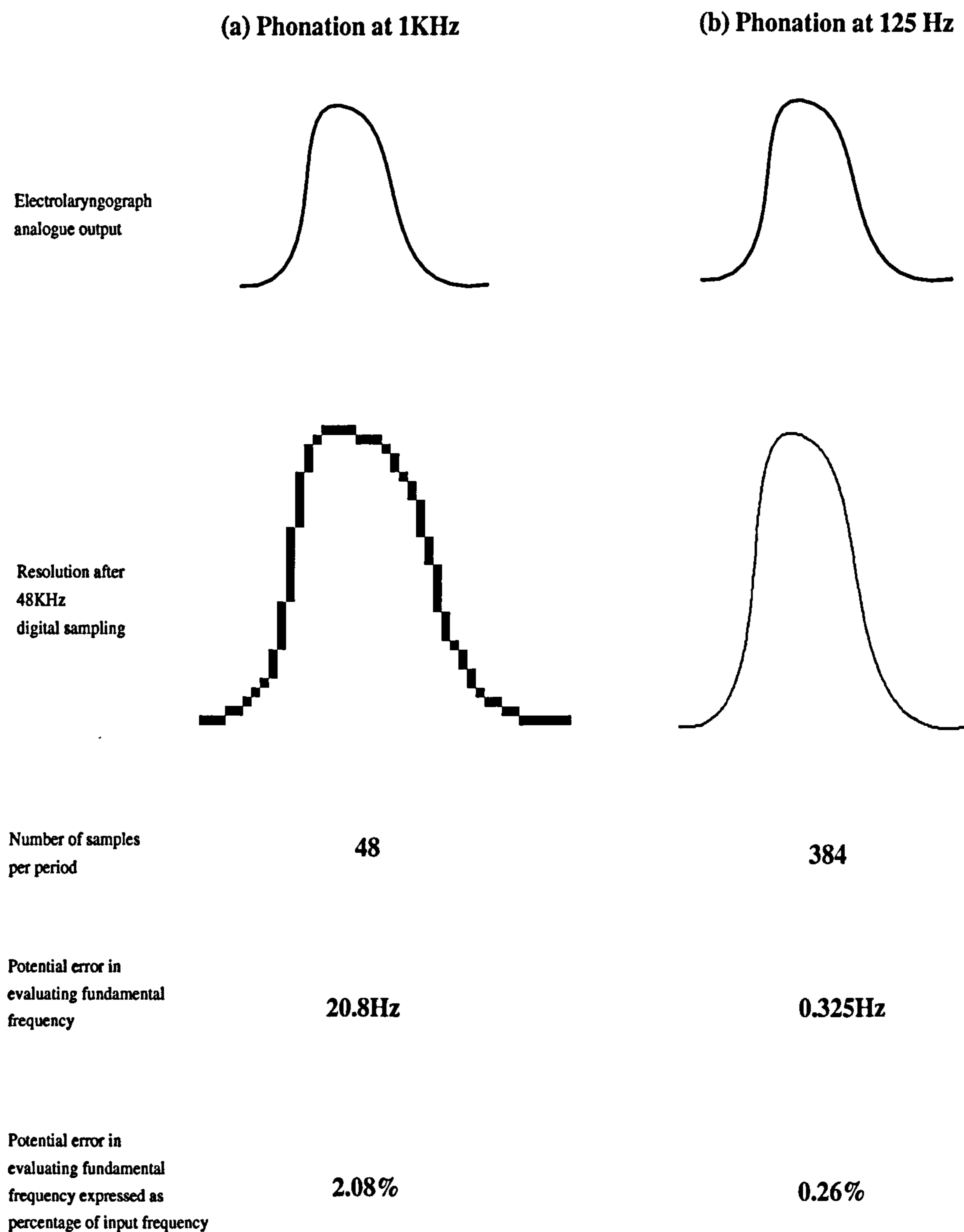


Figure 5.15: Accuracy of deriving fundamental frequency from the electrolaryngograph output for phonation at 1KHz and 125Hz digitally sampled at 48KHz

played. The user is able to select a value between 0% (show none of the previous period) and 100% (show both the current period and the previous in the normalised period display). An intermediate value of 50% has been found appropriate during use.

5.9.6.2 Animation of vocal folds

A facility was developed to animate an artificial image of the vocal folds so that users may more easily understand the relationship between the normalised signal and the action of the vocal folds. This is illustrated in figure 5.16. Different levels of CQ derived from the last period are reflected in the time for which the animated vocal folds are shown in contact. The user may select a continuous repetition of this sequence, or a single sequence. A slider is provided so that the user may alter the speed of the animation.

5.9.7 User control over electrolaryngograph signal analysis

5.9.7.1 Introduction

A total of four separate forms have been provided for the user to control the way in which ALBERT assesses the electrolaryngograph signal. These forms are discussed in the following sections.

5.9.7.2 Primary control form

This form is the primary form for controlling the processing and display of voice source signals. It is shown in figure 5.17. The user is given access to the following attributes.

1. The channel of the incoming electrolaryngograph signal
2. The ability to turn the entire voice source processing mechanism on or off. For example, if the user is only interested in processing the acoustic signal, then this button would sensibly be switched off.
3. The ability to invert the incoming electrolaryngograph signal. This option is useful because often there are differences between DAT machines made by

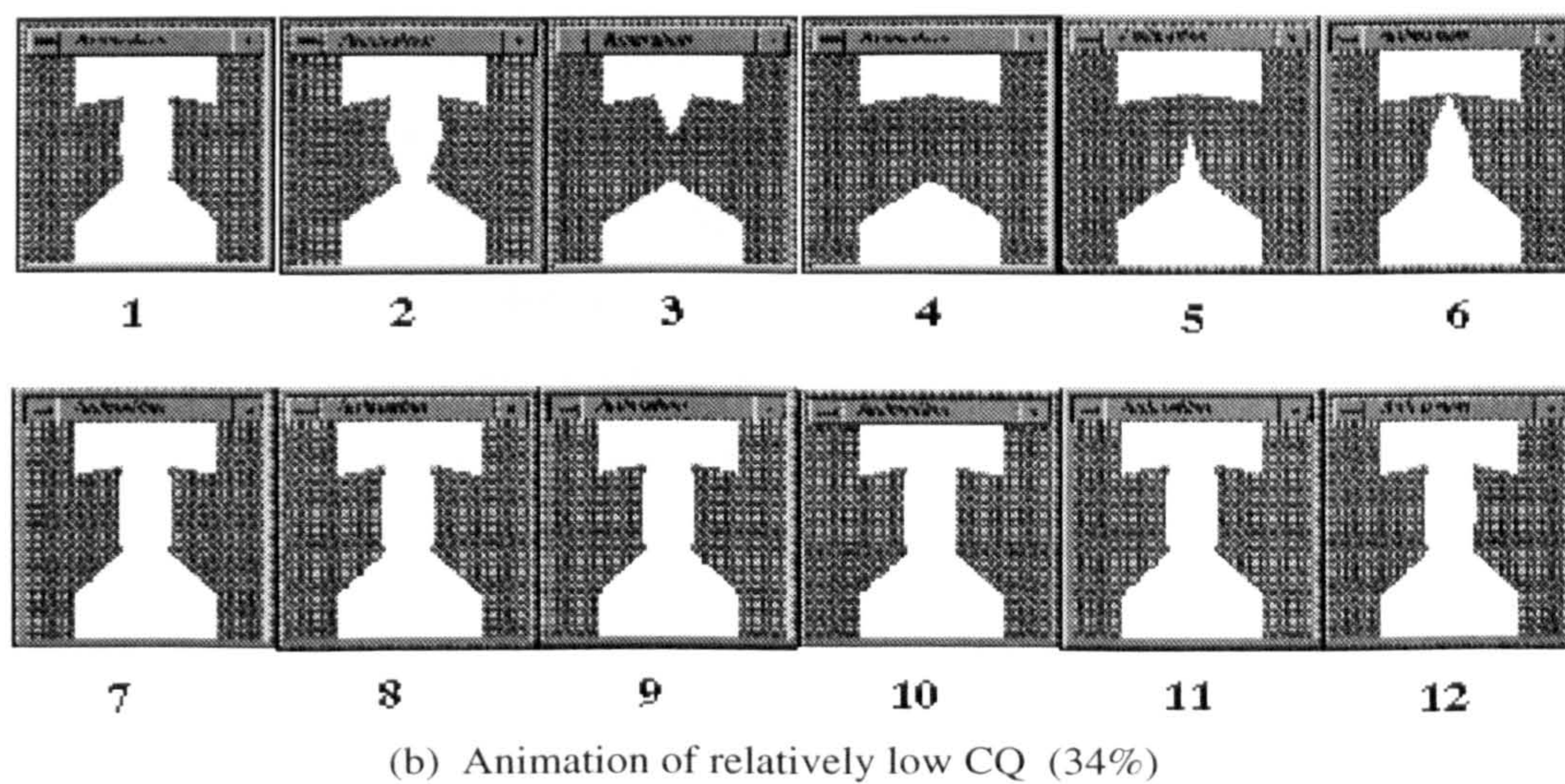
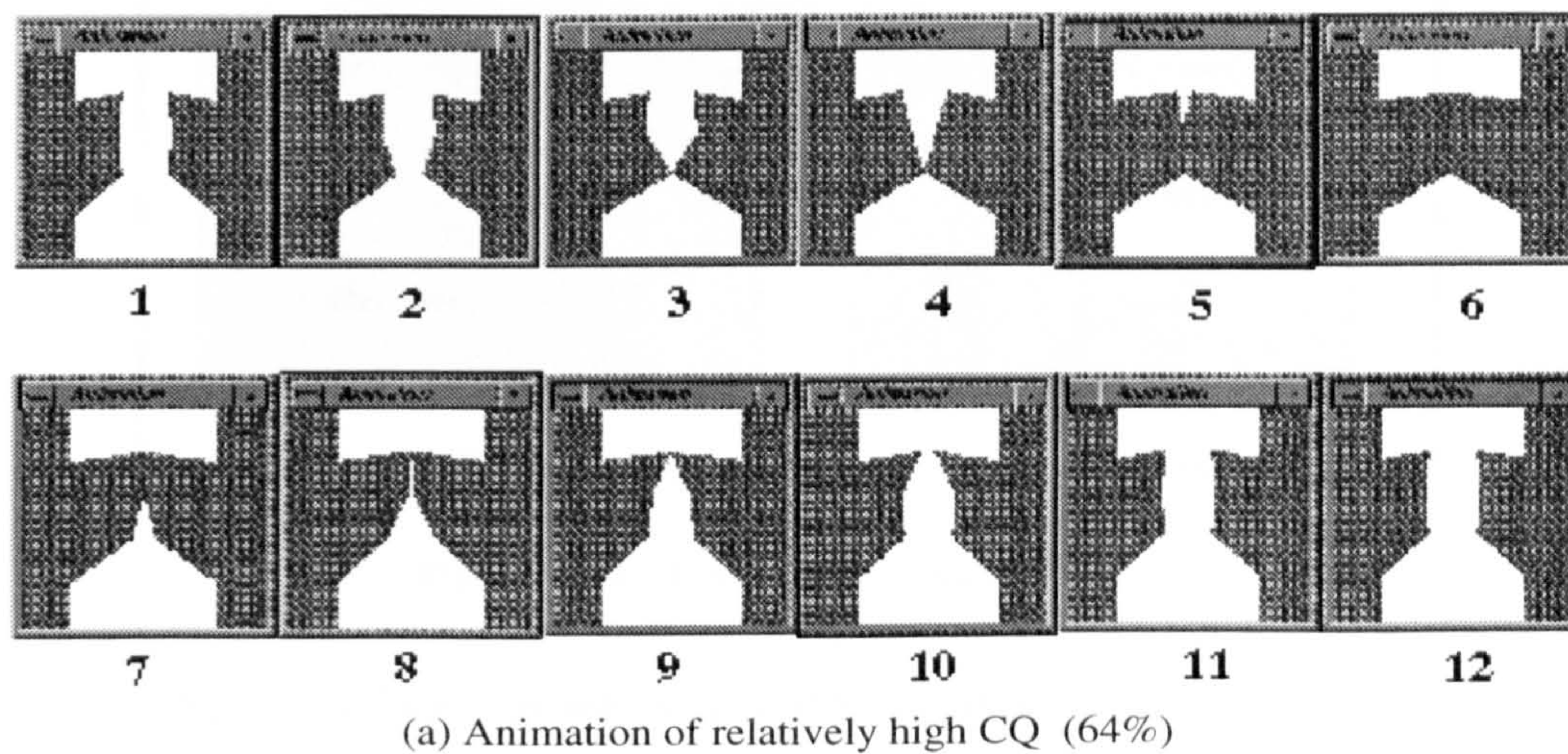


Figure 5.16: Animation sequences illustrating two contrasting levels of CQ

Voice source processing and display

LX channel <input type="button" value="Left"/> <input type="button" value="Right"/>	LX processing <input type="button" value="ON"/> <input type="button" value="OFF"/>	Invert <input type="button" value="ON"/> <input type="button" value="OFF"/>	Attribute derivation <input type="button" value="First in buffer"/> <input type="button" value="All in buffer"/>
--	---	--	---

<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Threshold</td> <td><input type="button" value="Assess"/></td> </tr> <tr> <td style="text-align: center;">Period display</td> <td><input type="button" value="Show period"/></td> </tr> <tr> <td style="text-align: center;">View recording</td> <td><input type="button" value="View LX"/></td> </tr> </table>	Threshold	<input type="button" value="Assess"/>	Period display	<input type="button" value="Show period"/>	View recording	<input type="button" value="View LX"/>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Low pass filter</td> </tr> <tr> <td style="text-align: center;"><input type="button" value="ON"/></td> </tr> <tr> <td style="text-align: center;"><input type="button" value="OFF"/></td> </tr> <tr> <td>Cutoff (Hz): <input style="width: 50px;" type="text" value="5000"/></td> </tr> </table>	Low pass filter	<input type="button" value="ON"/>	<input type="button" value="OFF"/>	Cutoff (Hz): <input style="width: 50px;" type="text" value="5000"/>
Threshold	<input type="button" value="Assess"/>										
Period display	<input type="button" value="Show period"/>										
View recording	<input type="button" value="View LX"/>										
Low pass filter											
<input type="button" value="ON"/>											
<input type="button" value="OFF"/>											
Cutoff (Hz): <input style="width: 50px;" type="text" value="5000"/>											

Figure 5.17: Voice source control form

different manufacturers which may mean that a signal recorded on one DAT machine will played back on another machine in an inverted form.

4. Attribute derivation. The user can select the extent to which each buffer of incoming electrolaryngograph signal is processed and graphical displays based on analysis of the signal are updated. There are two options; a mode in which only the first electrolaryngograph period present in each buffer will be processed, and a mode in which all periods are processed. The former option is typically used for real-time function; the latter for analysis when real-time performance is not important, but a thorough analysis of the signal is.
5. An automated process for setting the value of a parameter used in the assessment of the electrolaryngograph signal (discussed in the following section).
6. Access to the normalised period display (section 5.9.6.1).
7. The ability to view and print selected areas of the recorded voice source signal.
8. The ability to engage a low pass filter. This can be useful in processing noisy electrolaryngograph signals. The cut-off frequency, which has a default of 5000 Hz, is shown in a small window. If required, the user can alter this value.

5.9.7.3 Threshold form

This form provides easy access to a parameter used in the assessment of the electrolaryngograph signal. This parameter represents the smallest signal level that the voice source processing algorithm will judge to be a rising edge of the electrolaryngograph signal. If this value is too low, then the algorithm may be too eager to interpret the incoming signal as a rising edge. If this value is too high, then the algorithm will never detect a rising edge. A sensible intermediate value is required. This parameter may be automatically set by the press of a button on the voice source form. Alternatively, the user may wish to override the automatic setting. This can be achieved by simply clicking on the slider provided in this form and selecting the desired level.

5.9.7.4 Refractory time form

This form provides easy access to a parameter used in the processing of the electrolaryngograph signal. It represents the smallest interval of time immediately following the detection of a rising signal edge indicative of the start of a laryngeal period within which the voice source analysis algorithm is not permitted to detect a further rising edge. This parameter can be used to optionally improve the performance of the voice source processing algorithm. For example, if it is known that the vocal task under assessment does not involve phonation at a fundamental frequency higher than 500Hz, this slider can be set to 2ms; if it is known that the vocal task does not involve phonation at a fundamental frequency higher than 250Hz, then the slider may be set to 4ms. This may be achieved by simply clicking on the slider and selecting the desired level.

5.9.7.5 Period form

This form is used for analysis and animation of the electrolaryngograph period. It is shown in figure 5.18.

1. **Smoothing style.** This enables the user to select a form of smoothing for use in the normalised period display.

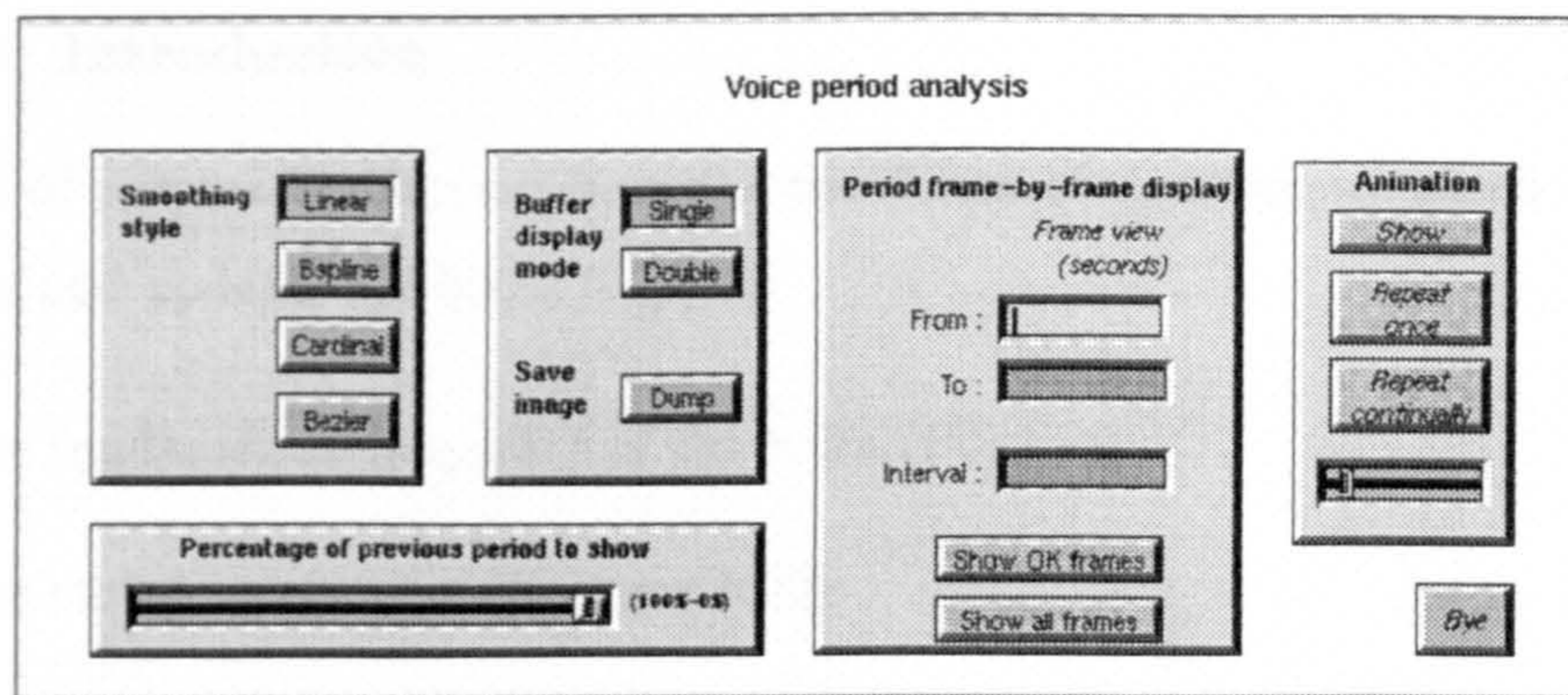


Figure 5.18: Voice period control form

2. **Period display buffering.** In this section the user is able to control the form of buffering used in the normalised period display. Single or double buffering may be selected. In the latter mode, the sequence of electrolyngograph periods appears visually smoother. However, a small time delay is incurred (of the order of one fiftieth of a second) each time the display is updated. This is due to the hardware operation of the Silicon Graphics computer. This delay may be inappropriate when a heavy work-load is required for real-time operation.
3. **Period frame-by-frame display.** This facility is provided to enable the user to examine a sequence of voice source activity that has been recorded to file. The user can enter the start time, end time and interval of analysis. One period of electrolyngograph signal at each interval is shown normalised. Timings and frame numbers are shown underneath each displayed period.
4. **Animation.** This controls the animation feature discussed previously (section 5.9.6.2). The user is provided with control over the animation display, mode of repetition, and speed of animation.
5. **Display of the signal prior to the current period.** This section contains the slider control discussed previously provided for control over optional display of the electrolyngograph signal immediately prior to the currently assessed period.

5.10 Voice parameters

5.10.1 Introduction

The list of parameters derived from the two forms of signal input which ALBERT analyses and updates are listed below.

1. the fundamental frequency of the voice (F0)
2. the closed quotient of the voice (CQ)
3. the rate of opening of the vocal folds (oa⁴)
4. the rate of closing of the vocal folds (ca⁵)
5. the spectral distribution of the voice output (Ratio)
6. the loudness of the voice (SPL)
7. the change in loudness from one time instant to the next (shimmer)
8. the period-to-period change in fundamental frequency (jitter)
9. the peak-to-peak amplitude of the electrolaryngograph period (lx amp)
10. the change in amplitude of the electrolaryngograph period from one time instant to the next (lx shimmer)
11. time

In section 4.5, it was concluded that the rate of opening and closing of the vocal folds are not appropriate indicators for vocal development at this time. These two parameters, listed above as numbers 3 and 4, may be more appropriate as aids to analysis rather than as suitable parameters for real-time feedback.

A formal evaluation of the assessment of these parameters by the ALBERT system is presented in Appendix F.

⁴This label was chosen to reflect the method in which the parameter is derived; the rate of opening (indicated by the maximum differential) divided by the peak-to-peak amplitude of the period.

⁵This label was chosen to reflect the method in which the parameter is derived; the rate of closing (indicated by the minimum differential) divided by the peak-to-peak amplitude of the period.

5.10. User control over the parameters

5.10.1. The parameter control window

5.10.1.1. List of parameters derived from voice analysis

List of parameters derived from voice analysis
 ('LX' means that the parameter was derived from the electrolaryngograph)

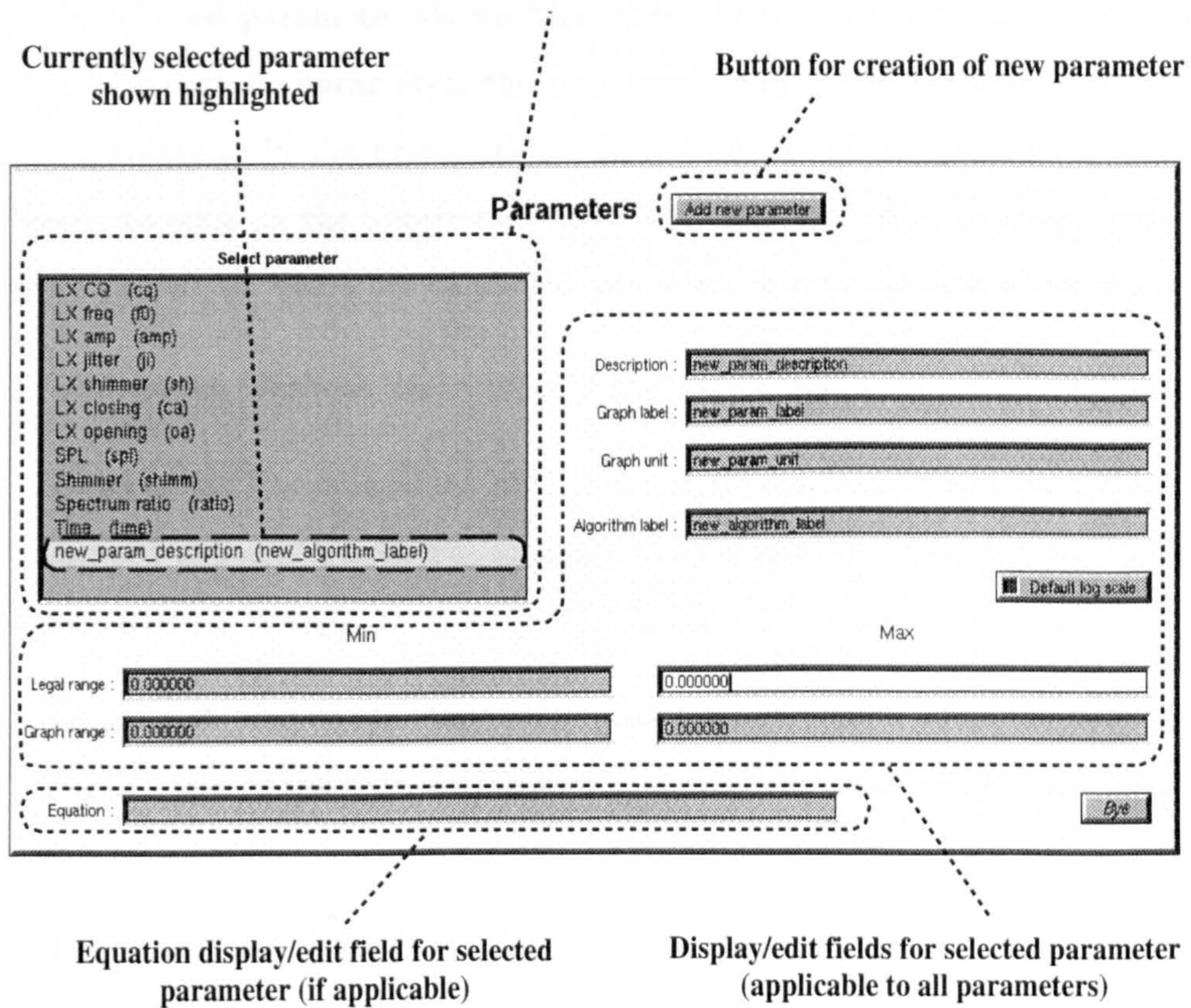


Figure 5.19: ALBERT parameter control window

5.10.2 User control over the parameters

In figure 5.19 these parameters are shown listed in the user interface display through which their properties may be viewed and edited by the user. The form is divided into two sections. The first section, shown in the top left hand corner of the form, is a simple list of each voice parameter. In the second section, a list of the properties, referred to as *fields*, of the parameter is shown filled with the relevant data of the currently selected parameter shown highlighted in the first section. To look at the fields of a particular parameter, the user needs only to click on the name of the relevant parameter in the first section. To adjust any of the parameter fields, the user needs to click on the appropriate field and alter the value as appropriate. The following parameter fields are displayed, all of which may be altered by the user.

1. *Description.* Verbose description of the parameter.
2. *Graph label.* The parameter label (optionally) displayed when the parameter is mapped to a graph axis. For example, the default label used for the ‘frequency’ parameter is ‘Freq’.
3. *Graph unit.* The parameter unit (optionally) displayed when the parameter is mapped to a graph axis. For example, the default unit used for the ‘frequency’ parameter is ‘Hz’.
4. *Algorithm label.* This is the label used to refer to the parameter in an algorithm reference (discussed below). For example, the default label used for the ‘frequency’ parameter is ‘f0’.
5. *Legal range - minimum.* This is the minimum legal value of the parameter. For example, the default legal minimum value of the ‘frequency’ parameter is 30.
6. *Legal range - maximum.* This is the maximum legal value of the parameter. For example, the default maximum legal range for the ‘frequency’ parameter is 1200.
7. *Graph range - minimum.* This is the minimum parameter value to be plotted on a graph. For example, the default minimum graph range for the ‘frequency’

parameter is 30.

8. *Graph range - maximum.* This is the maximum parameter value to be plotted on a graph. For example, the default maximum graph range for the 'frequency' parameter is 1200.

In addition, a boolean field indicates whether the visualization of the parameter will be mapped logarithmically by default. This may also be altered. If the state of this variable is not true, then any visualization of the parameter will be mapped linearly by default. For example, this is usually true for the 'frequency' parameter, but is not true for the 'time' parameter.

5.10.3 Extending the list of voice parameters

A new parameter may be created by the user by pressing the *add new parameter*. A 'skeleton' parameter is then created by ALBERT with dummy fields. The user may then alter these fields according to his/her requirements. The user is able to define a new parameter based on an algorithmic treatment of other parameters. For example, to provide a new parameter which gives an indication of the performance of CQ with regard to frequency, the user can request a new parameter button, provide a new name for the parameter (for example, CQ_freq), and enter the equation for the new parameter (in this case, CQ/Freq). Once the other fields have also been entered, the new parameter may be used for visualization or for reference by a further algorithm.

5.11 Recording acoustic and electrolaryngograph signals

5.11.1 Introduction

There are many advantages in being able to record a sequence of voice source and acoustic signals. For example, a library of acoustic and electrolaryngograph recordings can be constructed for easy reference, and a recorded sequence may be played back any number of times, facilitating the exploration of different forms of analysis and visualization in the interpretation of the signals. Furthermore, by recording a

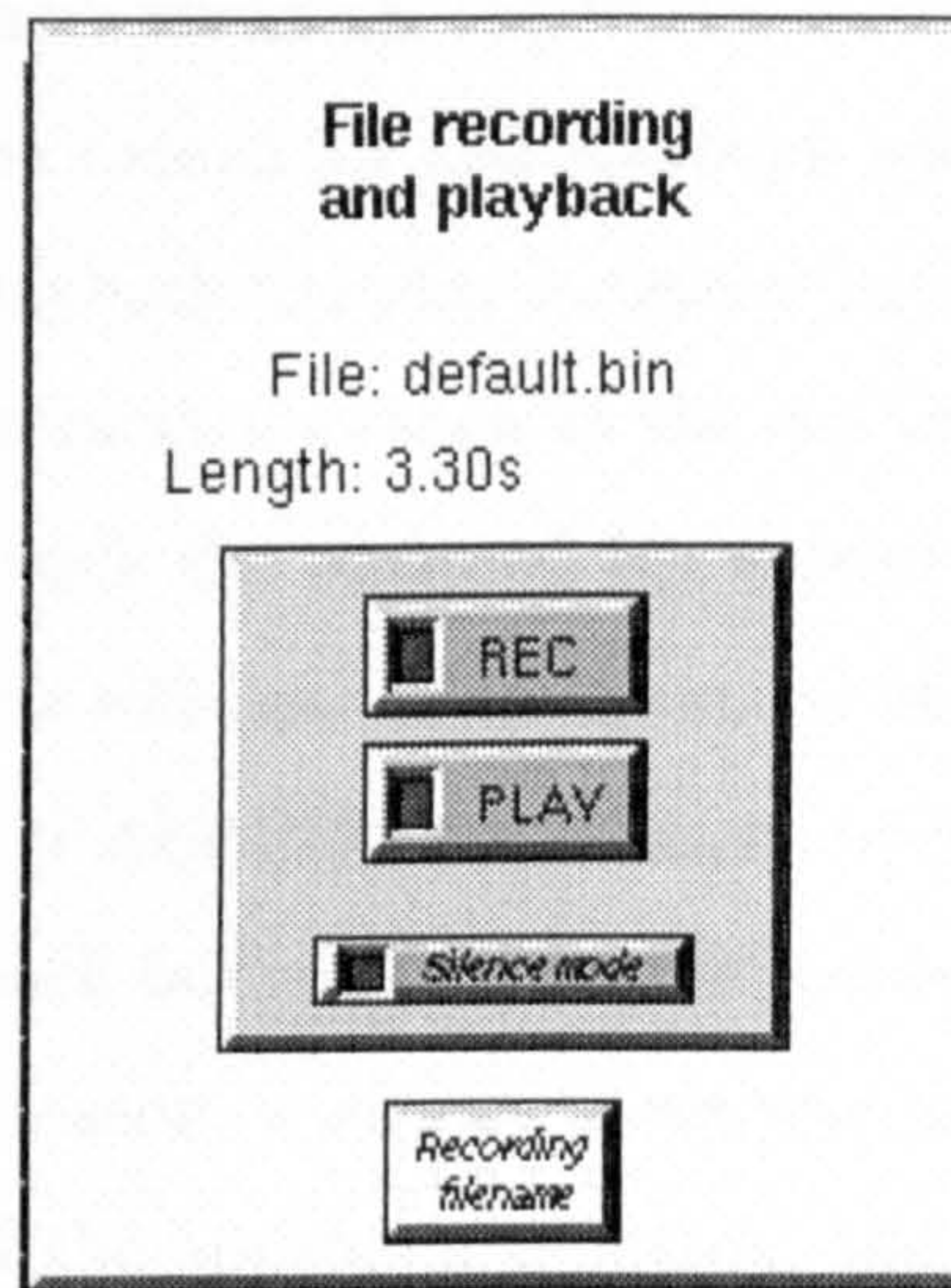


Figure 5.20: User control form for file record and playback

sequence prior to analysis, the user is assured of a thorough analysis of the sequence regardless of the analysis and display load placed on the system.

Two recording facilities are provided. The first is the ability to record and playback to and from a file; the second is the ability to record and playback to and from memory. These are discussed in the following sections.

5.11.2 Recording to file

Controls over the recording and playback of acoustic and laryngeal signals to file are presented on the form shown in figure 5.20. To start the recording of the incoming signal, the user clicks on the **REC** button. A recording indicator light will turn on. To finish recording, the user presses the button again. The recording light will turn off and the length of the recording shown at the top of the form will be updated. This is expressed in seconds. To play the file, the user clicks on the **PLAY** button. The play indicator light will turn on and the recorded sequence will be played back. The light will turn off when playback is complete.

It was realised that there would be occasions when it would be useful to be able to specifically instruct ALBERT to halt playback of a recorded sequence without the entire recording being played. For example, the user might wish to quickly check the contents of a selected file by listening to the first few seconds. A mechanism was developed so that the user could halt a playback sequence prior to the end of the

recording. For this, the user clicks on the PLAY button a second time. The play indicator light will then be turned off and playback will cease. Alternatively, if the user does not interrupt playback the play indicator light will turn off and play back will cease without user interaction as soon as the end of the recorded file is reached.

Normally during playback the recorded file is read and played via the internal speakers. To an extent, this will slow down the speed at which ALBERT will be able to proceed. Furthermore, if ALBERT has a heavy work load, there will be a delay between the playback of each buffer. This can be perceptually very disconcerting to the human ear. For this reason, a *silence mode* was implemented. When engaged by the user via a press of the appropriate button, the audio buffers read from file are subsequently processed in silence.

5.11.3 Recording to memory

If the host computer running ALBERT has a lot of memory, it is useful to be able to have the option to use this memory for storing signals as an alternative to using file storage space on the hard disk. In addition, the speed of processing for signals stored in memory is faster than for signals recorded to file, because the access time is much quicker.

For these reasons, the function and user interface design of the form discussed in the previous section for recording and playback from file was duplicated for the purposes of recording and playback to and from memory.

One functional alteration was made. When a program such as ALBERT ceases operation, the contents of the memory used by that program are normally lost. This would mean that any recorded sequence held in memory would also be lost. For this reason, the ability to save a memory recording to file was developed. To this end, the function and name of the Recording filename button originally present on the 'Record to file' form was replaced by one for the selection of a filename to which the memory recording is to be transferred.

5.12 Methods for visualization

5.12.1 Introduction

Prior to the design of the system some consideration was given concerning voice parameters and their display in the analytical scientific literature. A summary of the display methods adopted in forty research papers and their application is shown in table 5.1. There is a wide selection of parameters, display configurations, and of tasks to which the configuration sets were applied. However, all the studies utilise fundamentally identical graphical techniques. Predominantly, this is a two dimensional representation of information with one parameter mapped to the x axis and another mapped to the y axis. It is only the selection of parameters which is different for each visualization case.

A system was therefore designed in which the visual structure for information display was not fixed in any predetermined form, but could be easily configured by the user in a matter of seconds in any combination s/he required. In this way the manner of information display is not fixed to a handful of select display modes in the manner of all voice parameter display software currently available, but is left largely free for exploration and configuration by users. The visualization approach adopted supports all of the display modes listed in table 5.1 and the tasks for which they were used, in addition to other modes created by the user employing any of the parameters derived for display. Within the ALBERT software one dimensional and three dimensional displays are also supported in addition to the two dimensional displays previously discussed. The former is intended predominantly for real-time display as part of a vocal task given to a subject, the latter for investigation and analysis of the interaction between a selection of voice parameters.⁶

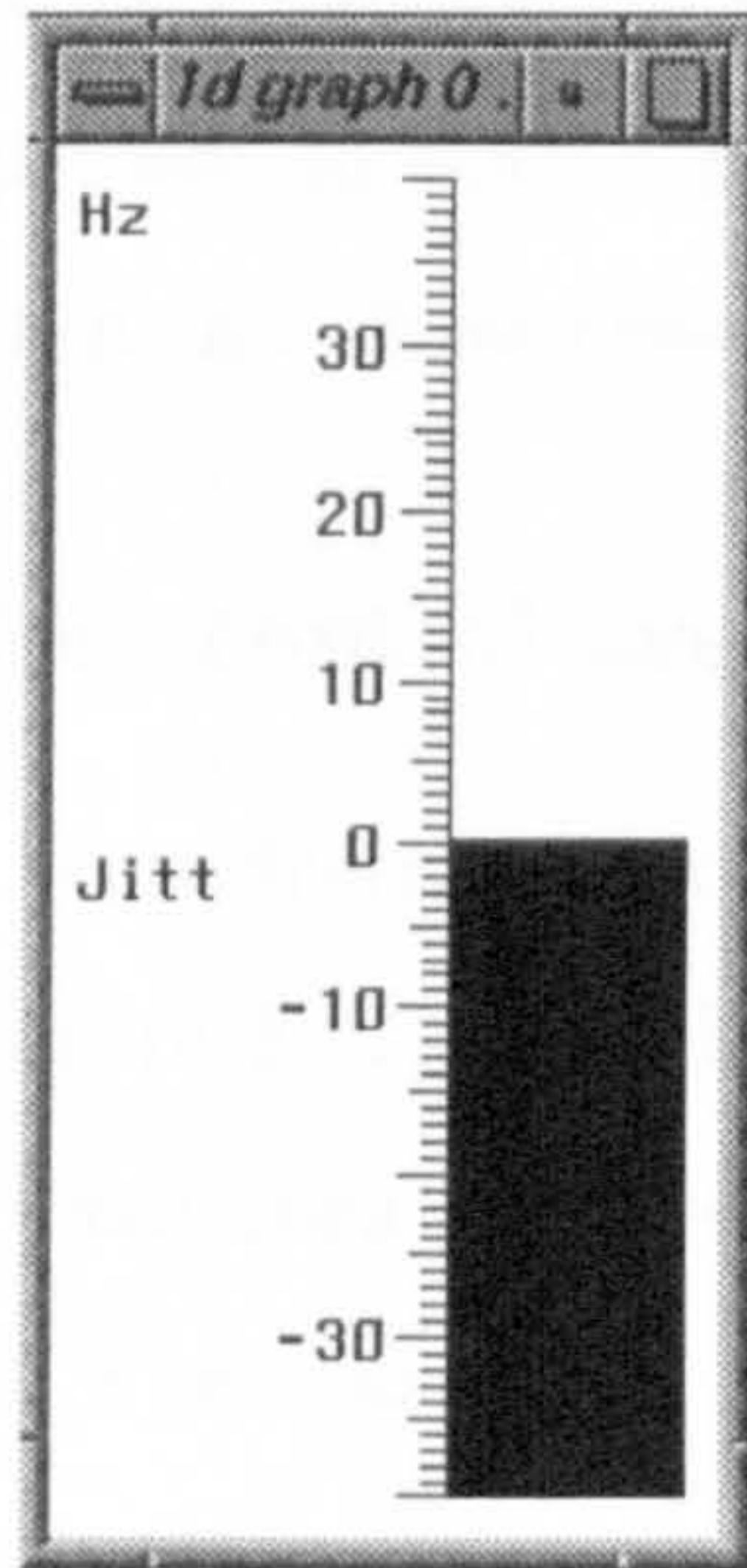
5.12.2 An example application

Examples of display configurations which could be used for real-time feedback are shown in figure 6 with application to the real-time display of jitter. A bar graph

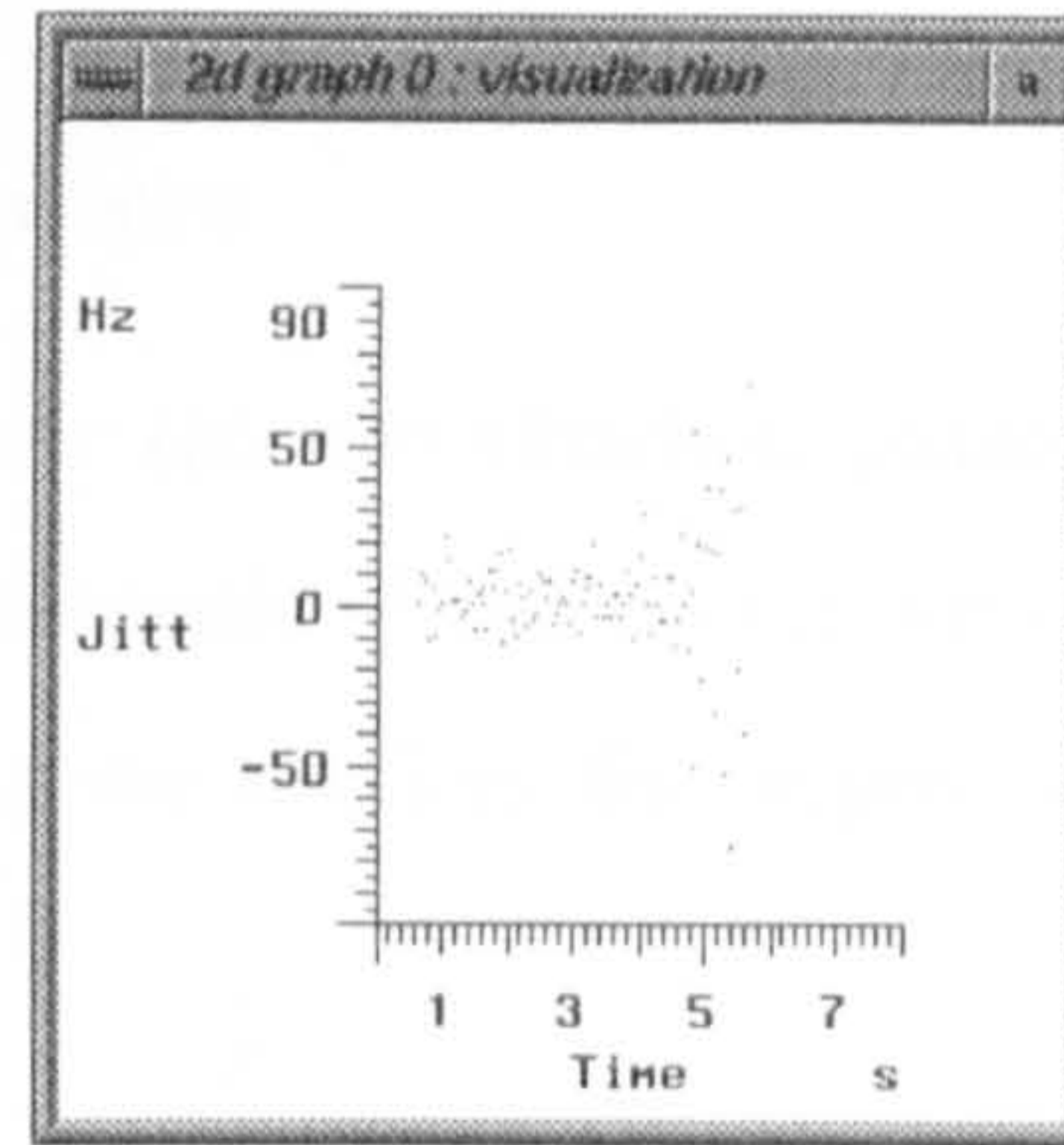
⁶At the time of writing the ALBERT 3D visualization options are underdeveloped, as priority was given to 1D and 2D visualization support which are of greater application for use as real-time tuition tools.

- Time against F0** – Transient time in singers (Leonard et al., 1987) – with superimposed Intensity against time – vowel example (Hori, 1983)
- LX signal amplitude against Time** – Voice register change (Roubeau et al., 1987)
- F0 against F0, scattergram** – Structured distribution of successive pitch periods for young woman with moderate creak (Hirson & Roe, 1993)
- F0 against time** – Vibrato patterns for different singing styles (Howard, 1992). Pitch stimulus for assessing junior high school singers (Cooksey, 1993). Trained male baritone two octave scale (Lindsey & Howard, 1989). Incentive displays for encouraging voice development in primary school users (Welch et al., 1989). Improving the speech of deaf children (Wirz & Anthony, 1979). Teaching intonation to deaf persons (Spaai et al., 1993). The vowel in 'dart' spoken by a healthy adult female (Abberton et al., 1989). The effect of articulation in singers and speakers (Baken & Orlikoff, 1987). Voice register change (Roubeau et al., 1987). Analysis of Parkinson's disease (Watson & Macaulay, 1994). Performance demands (Coleman, 1987)
- CQ/OQ against time** – Trained male baritone two octave scale (Lindsey & Howard, 1989). The vowel in 'dart' spoken by a healthy adult female (Abberton et al., 1989).
- CQ against F0, scattergram** – Different singing styles (Howard, 1992). Belt and opera qualities in the female voice (Evans & Howard, 1993). 10 trained/untrained male singers (Lindsey & Howard, 1989). 21 trained/untrained female singers (Howard et al., 1991). a trained/untrained singer, read passage/sung two-octave scale (Howard et al., 1990). Healthy adult male reading a 2 minute passage (Abberton et al., 1989).
- CQ against F0** – Male tenor/baritone/bass voices (Cleveland & Sundberg, 1983)
- Ratio against intensity** – Artists/students/musicians/non-singers, male/female subjects (Hollien, 1993). Male tenor/baritone/bass voices (Cleveland & Sundberg, 1983).
- Ratio (energy in 3KHz filter) against time** – Stage singers (Winckel, 1971)
- SPL against F0 (PHONETOGRAM)** – Voice change during puberty for boys (Pederson et al., 1982). Voice change during puberty for girls (Pederson et al., 1984). Male/female singers/nonsingers (Gramming, 1991). Male/female singers/nonsingers/vocal dysfunction (Gramming, 1988). Trained/untrained vocalists (Awan, 1993). Performance demands (Coleman, 1987).
- SPL against F0, scattergram (density mapped to SNR)** – Visual representations of the voice (Klingholz, 1990)
- Shimmer against Jitter** – Normal and dysfunctional voice users (Haji et al., 1986).
- Relative power against frequency (SPECTROGRAM)** – Singing formant analysis for different types of western operatic singers (Walker, 1992). Acoustical efficiency in singing (Simonson, 1992). Belting and pop modes of female singing (Schutte & Miller, 1993). Singing voice analysis (Miller & Schutte, 1990).

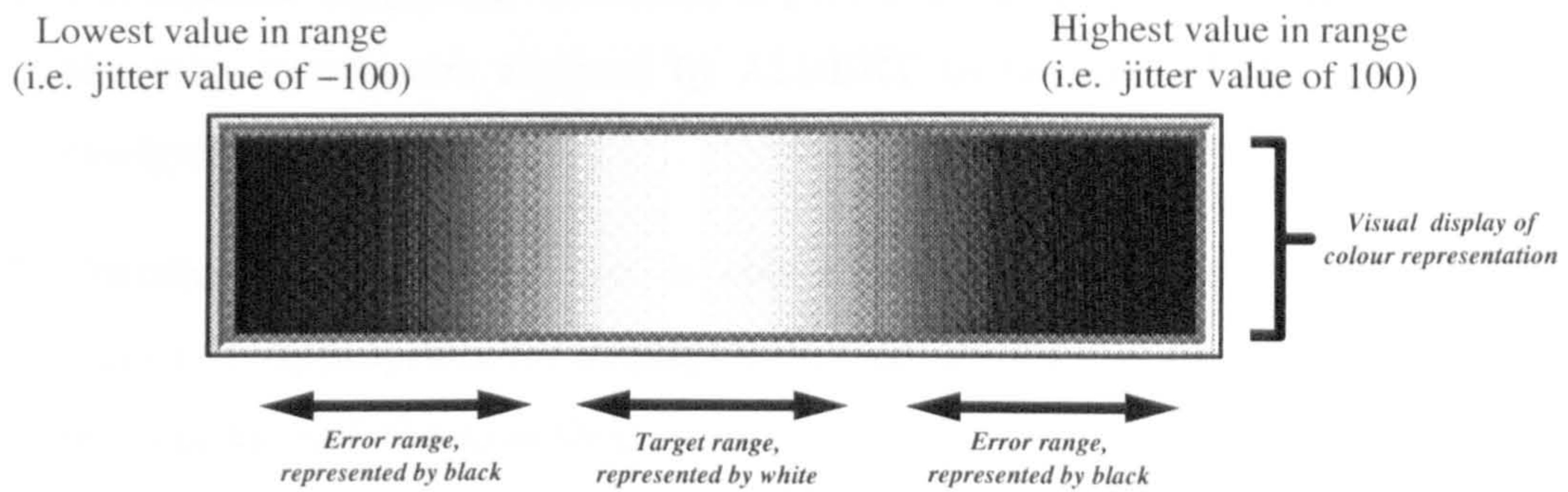
Table 5.1: Example visualization configurations and function presented in the scientific literature, and example configurations and applications in which ALBERT may be used. Order of entries: y-axis followed by x-axis, coloured black unless specified as scattergram. 'LX' refers to the signal derived from an electrolaryngograph. 'Ratio' refers to the level of spectral amplitude in the singer's formant region relative to all voice spectrum output. 'SNR' stands for signal-to-noise ratio.



(a) 1D bar graph display of jitter



(b) 2D graph display of jitter plotted against time



(c) A suitable colour scheme for real-time feedback of jitter. This may be used as a complimentary medium of information by, for example, either the 1D or 2D display shown above. The colour of the plotted data will then reflect a measure of the acceptability of the level of jitter.

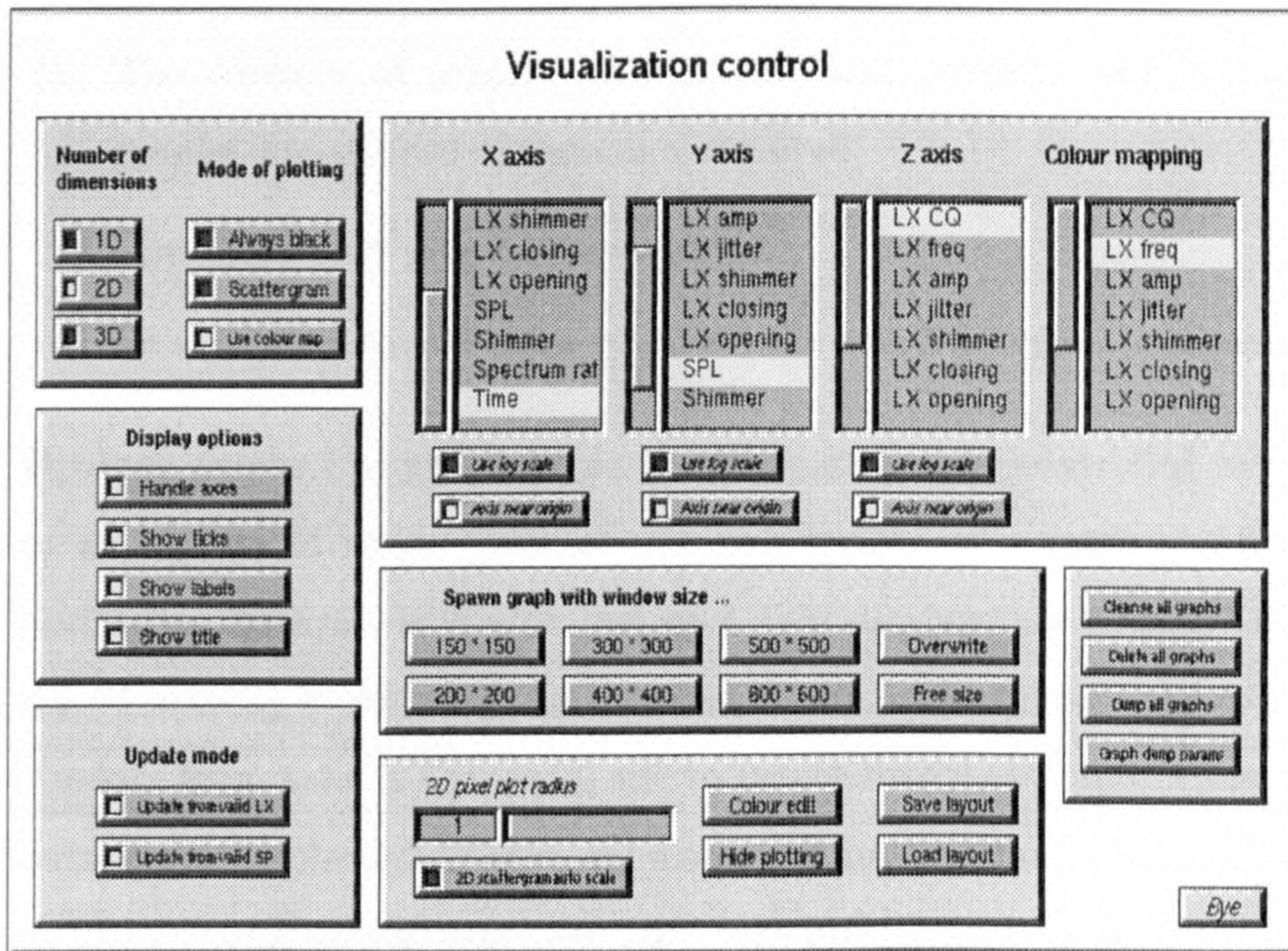
Figure 5.21: Two example visualization configurations and a complimentary colour scheme with application to the real-time display of jitter

display is shown in figure 5.21(a), and a 2D display of jitter against time is shown in figure 5.21(b). In both instances, the subject can use the display during sustained phonation to concentrate on maintaining a small deviation from the target pitch (visually expressed in both 5.21(a) and 5.21(b) as minimal fluctuation in the vertical plane, with a mean value close to zero).

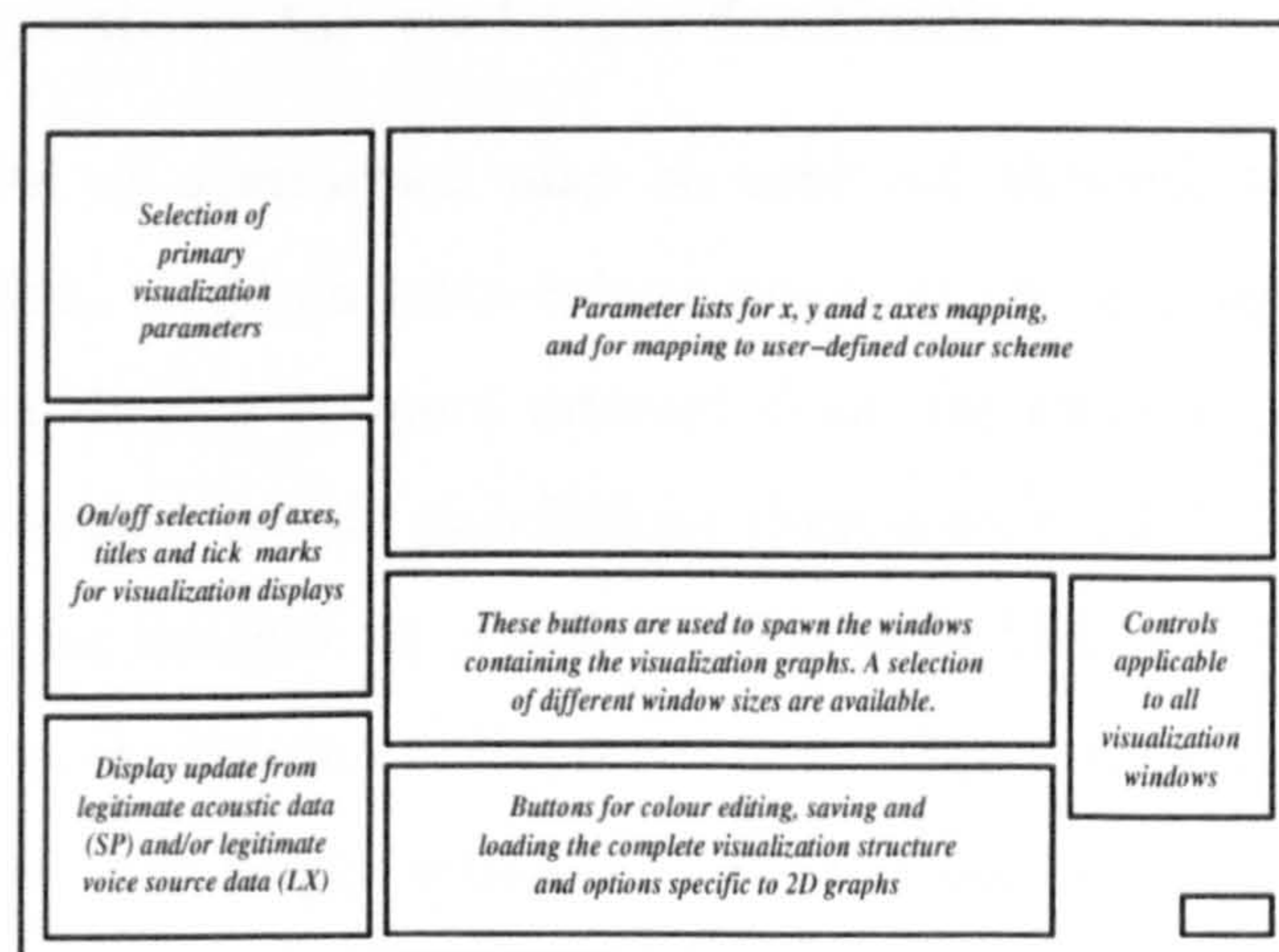
5.12.3 User control over the parameters

A form has been designed for easy control over the visualization parameters. This is shown in figure 5.22. Functionally similar elements of the user interface presented on the form have been grouped together into seven sections for improved access and presentation. These are:

1. **High level attributes of the visualization.** This includes the selection of 1D, 2D, or 3D graph configurations, and the mode of colour display.
2. **Parameter mapping.** This section provides the user with the means to map the voice parameters assessed by ALBERT to the axes of the visualization configuration.
3. **Display options.** The user is able to select which axes elements he/she regards as appropriate for display, including control over the display of labels, tick marks, and the axes themselves.
4. **Spawn graph.** This section presents the user with a list of possible window sizes from which an appropriate selection may be made. Alternatively, the user may spawn a window of a specific width and height. It is also possible to 'overlay' a new graph on top of a graph of the same format which is already present on the screen. For reasons of clarity, the smallest graph size which the user can select is 15% of the total screen size. This size has been found an appropriate minimum during use of the system.
5. **General graph control.** Four buttons are available for high-level graph control. These are for the following functions.



(a) visualization control window display



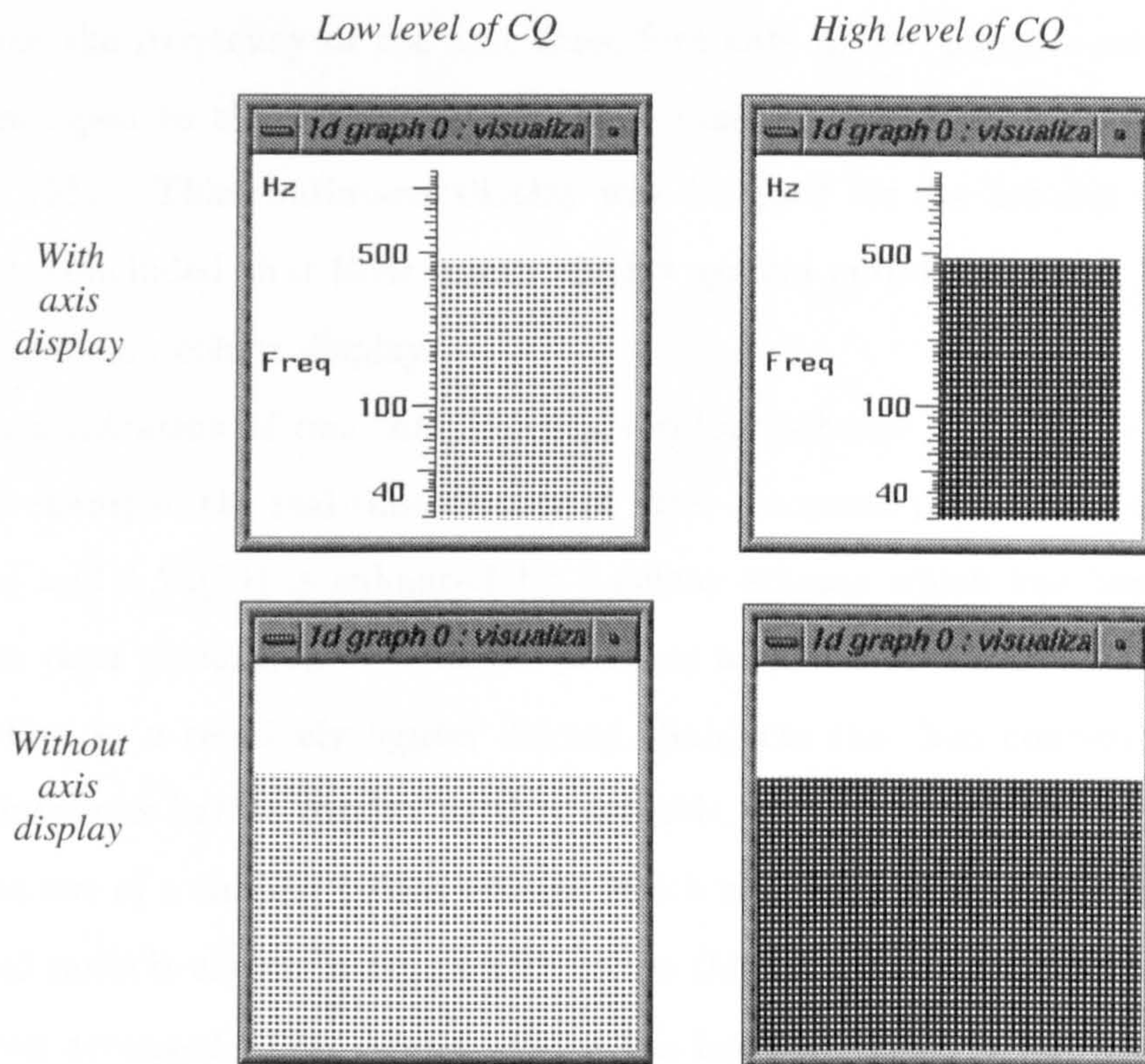
(b) key to display

Figure 5.22: ALBERT visualization control window

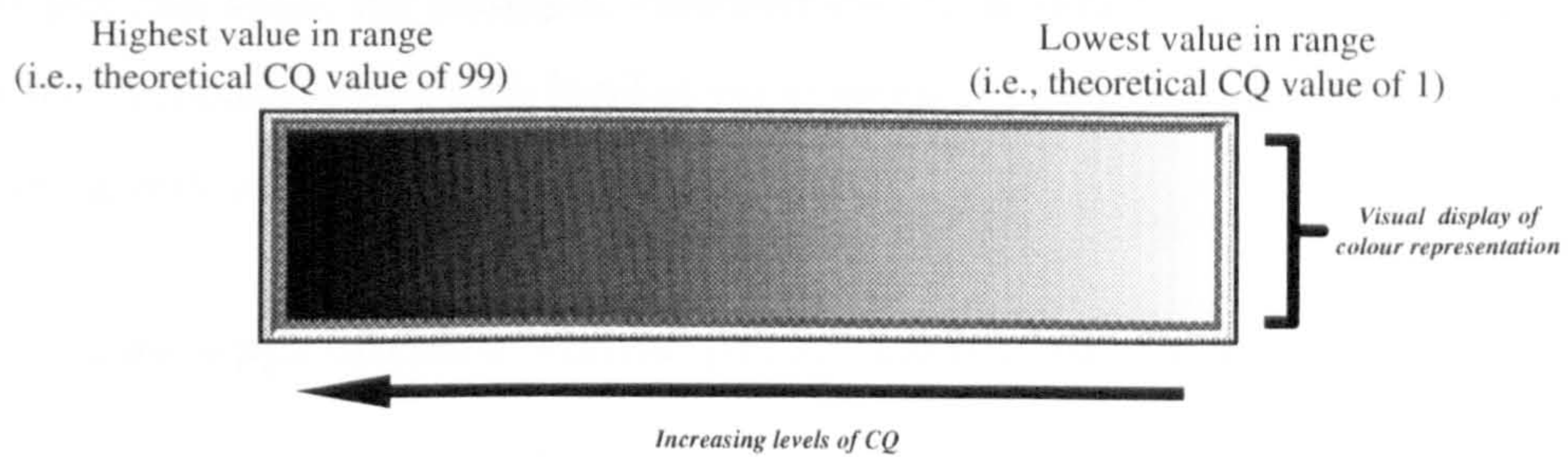
- (a) The cleansing of graphs. All graphs are cleared of data and reset to their original state. All statistics associated with the graphs (such as mean and standard deviation values) are also reset.
 - (b) The deletion of graphs. All graphs, and the windows containing the graphs, are cleared of data and removed.
 - (c) The transfer of graph images to file.
 - (d) Control over the filename to which the graph images are transferred.
6. **Update mode** The graph displays may be optionally controlled to react only to an update of either voice source parameters or acoustic parameters, or both. For example, it is often required that displays are updated only when voiced activity is detected. In this case it would be appropriate to select the Update from valid LX button, and to ensure that the Update from valid SP button was deselected. The visual display of any parameter (even if it is based on the acoustic signal) will then only take place if voicing action is detected.

5.12.4 Using colour for real-time feedback

A further level of abstraction may be achieved through the use of colour to present information. Children learn colour names at an early age and may well find a changing colour display of more interest than, for example, a constantly black display. There are many colour associations (Macdonald, 1990), all of which may be actively exploited for the process of voice tuition with ALBERT. For example, colour has the ability to evoke emotional responses or to trigger memories. Warm hues (i.e., red, orange, yellow) can imply action or danger. Cool hues (i.e., green, blue, grey) can imply passivity or sadness. Guidelines for the selection of appropriate colour selections already exist. For example, Hunt (1987) recommends the use of Green and Red to indicate a two band scheme, and white, cyan, and orange to indicate a three band scheme. Some appreciation of the subject's abilities, limitations and cultural and educational background are appropriate in the creation of a colour mapping scheme for real-time information display. Colour schemes can then be 'personalised' to suit the subject and vocal task.



(a) 1D bar graph displays of fundamental frequency and CQ. Two contrasting levels of CQ are shown, with and without axis display.



(b) The colour scheme used in the examples shown in (a) above for the complimentary display of CQ

Figure 5.23: Real-time graph display of fundamental frequency with levels of closed quotient (CQ) expressed by colour changes

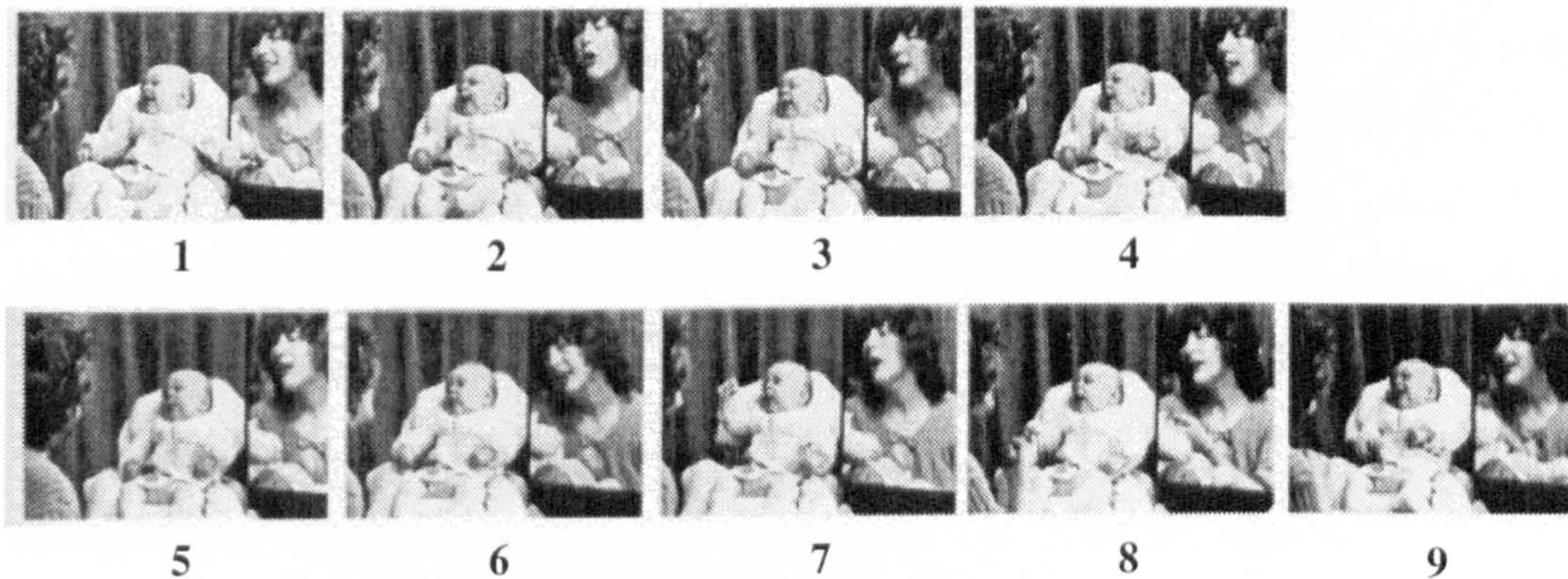
The use of colour in real-time voice displays appears to be very rare. In one instance the frequency of the first three formants in the acoustic output spectrum were mapped to the red, green and blue guns of a television monitor (Watanabe et al., 1985). This continuous display was designed for the hearing impaired. The authors concluded that their colour display system performed much better than an equivalent non-colour display.

An illustration of one complimentary colour scheme is shown in figure 5.21(c). In this example, the real-time display of jitter discussed previously (shown in figure 5.21(a) and 5.21(b)) is enhanced by a colour scheme which has been designed to express poor phonation of a target pitch as a relatively darker display, and good phonation as a relatively lighter display. Subjects can then concentrate on trying to maintain as light a display as they are able.

The use of a simpler colour scheme which is appropriate for a different selection of vocal tasks is shown in figure 5.23(b). In this example, a simple linear gray scale has been arranged. This colour scheme has been employed in the graphs shown in figure 5.23(a), where the height of the bar graphs have been mapped to fundamental frequency, and the level of grayness of the bar graphs have been mapped to CQ. Relatively higher levels of CQ are manifested as relatively darker bar graph displays. The subject can then, for example, concentrate on maintaining a consistently dark display (corresponding to a high level of vocal efficiency) whilst controlling the height of the bar graph display during pitch exercises.

5.12.5 The appropriate visual presentation of data

There are dangers in presenting information concerning the physiology of the voice in the literal form which many devices provide (such as the videoendoscope, MRI and X-ray devices). By the nature of their operation these devices tend to invite the focus of attention on particular elements in the process of vocal production. This reflects their role predominantly as tools for diagnosis and analysis concerning particular areas of vocal anatomy, rather than their ability to effectively present information suitable for a developing voice user who is unlikely to have a thorough knowledge of physiological mechanisms used in the voice process and who may accordingly be unable to concentrate on a specific physiological task. Sundberg (1987), for example,



A conversational cycle at age 6 weeks

1. Greeting smiles. Mother says 'Come on then'
- 2, 3. 'Oh aaa. Come on.' Baby coos.
- 4, 5, 6. Baby coos 'Aghoo.' Mother says, 'Come on. Hoo hoo.'
- 7, 8, 9. As before, then with baby's smile, 'That's right. Woo hoo.'

Figure 5.24: Mother-child interaction (From Trevarthen, 1983)

notes that given the many different forms of voice anatomy and usage,

“A danger with a more specific type of instruction, utilizing terms such as tongue, jaw, subglottic pressure, and so on, is that it might cause students to focus attention on these features, which are... actually means rather than goals”.

Information for visual feedback would benefit from the presentation of data in an appropriate form for the developing voice user. An example taken from the earliest stages of vocal development is that of a parent encouraging the vocal activities of his/her baby. Exaggerated articulatory actions are used by the parent to produce target stimuli. An example interaction sequence is shown in figure 5.24 (from Trevarthen, 1983). Facial responses are then used as a form of visual feedback to encourage and reward appropriate reaction from the child. Despite the fact that a baby is unable to concentrate on any single aspect of voice development, some of the fundamental steps of language learning, such as basic vowel development, are learned through this process. To re-phrase this in terms of Sundberg's words, visual feedback is presented to the subject based on the goals rather than on the means of voice production.

ALBERT has been designed so that detail deemed by the user to be at too low a level to be appropriate for the tuition task may be discarded. This approach has been supported through two main features. The first is the ability to combine any number of voice parameters algorithmically into a single parametric measure, which can then be updated and displayed in real-time. The second is the ability to configure the display to present parameter values in one of many different levels of abstraction. For example, any parameter may be displayed as a dynamically changing bar measure. However, the labelling of the axes may appear alien to many subjects, such as children. The user can therefore deselect several attributes of axis labelling, including the entire axis display. Examples of this are shown in figure 5.23(a).

Tufte (1983) and others espouse the maximising of the data-ink ratio criterion. This measure is defined as the ratio of ink used for data to the ink used for the whole display. This parameter holds true for computer displays also with the description of 'ink' changed to 'pixels' (the smallest element on a computer display). The free control given to the user for configuring the visualization enables the maximisation of this measure for highly effective information display.

5.13 The rate at which information is updated

5.13.1 Introduction

In the design of any system for real-time feedback, some consideration must be given to the rate at which the process of deriving information and subsequent display is carried out to ensure appropriate modes of operation for the learning process. Some vocal tasks result in phonation of very high pitch; for example, the singing of a soprano top C requires phonation at a pitch of approximately 1046 Hz. Therefore, for each second of phonation in this mode, the pattern of vocal fold movement which imprints the fundamental waveform onto the flow of air from the lungs occurs more than 1000 times a second. For the sake of real-time feedback there is no advantage to analysing each of the occasions for which the vocal folds oscillate. The human brain is not able to process such a high rate of information usefully.

For example, the propagation of neural signals ranges from a tenth of a second

to a hundredth of a second for eye to hand coordination (Hendee & Wells, 1993). Therefore, it is reasonable to assume that a similar range of rates of information transfer occurs from the presentation of an image during real-time feedback and subsequent reaction to that image. In addition, a visual image is known to linger on the retina for about a tenth of a second, and if during this instant it is replaced by a new image the two will seem to merge (Jones, 1994). Practical experience in virtual reality environments also suggests that an update period of 90ms, very close to the rate of image merging noted by Jones, is adequate to give an illusion of continuous motion (Wenzel, 1992).

On the basis of these observations, ALBERT has been designed with a default update rate of twenty times a second, which has been found comfortable and appropriate during use. However, there are occasions for which a different update rate may be suitable. For example, about 1 in 2000 people are susceptible to photo-seizures (McCrone, 1994). The frequency most likely to trigger an epileptic fit in a susceptible person is 18 times a second. A user of ALBERT may therefore wish to alter the update rate from 20 times a second to one further removed from that of the frequency domain in which photo-seizures are known to occur, such as 25Hz. Furthermore, in dim light the persistence of vision increases (Jones, 1994). This phenomenon could be usefully exploited by reducing the update rate. This would enable ALBERT to support a higher level of real-time analysis and visualization. Provision has been made within ALBERT to enable the user to change the update rate at any point in time through access to a simple slider mechanism on the GUI.

As discussed previously (section 5.9.7.2), a comprehensive analysis mode of operation is available in which each and every vocal fold oscillation is analysed and the derived information displayed as required. For some high load instances, this facility is supported at the sacrifice of real-time performance.

5.13.2 User control over the update rate

Three sliders are provided on the form. One displays the number of updates per second, another displays the number of samples per buffer, and a third displays the length of time of each buffer. All three slider values are inter-related. A change in one of the three will result in automatic changes to the other two.

Save and Load		
Visualization	Save	Load
Items	Save	Load
Forms	Save	Load
Acoustic display + flags	Save	Load
Voice source display + flags	Save	Load
All the above	Save	Load
Bye		

Figure 5.25: User control form for saving and loading ALBERT configurations

5.14 Saving and loading ALBERT configurations

The user may wish to use a particular configuration of analysis and display on many occasions. In order that the arrangement does not have to be recreated on each occasion, the user is able to save and load the state of ALBERT. The form used for this is shown in figure 5.25. In addition to the ability to save and load the entire state of ALBERT, the user is also able to save and load subsets of the state. For example, the user may have two different arrangements saved to file (such as one for voice source analysis and one for acoustic analysis) and may wish to incorporate *elements* of both into a single arrangement. The ability to access the various *domains* within ALBERT would greatly facilitate this task.

The following domains may be saved and loaded. They are listed here in the order in which they appear on the control form.

1. *Visualization*. This includes all 1D, 2D and 3D graphs and colour configurations, in addition to the mapping information of voice parameters to graphs and the positions of graphs on the screen.
2. *Items*. This includes the fields of all voice parameters, including any parameters that may have been created by the user.

3. *Forms*. This includes the position, size and visibility of every form.
4. *Acoustic display and flags*. This includes the position and size of both acoustic displays and all flags associated with acoustic processing. This does not include any information that may be held within the 'Items' and 'Visualization' domains.
5. *Voice source display and flags*. This includes the position and size of the normalized period display and all flags associated with electrolaryngograph processing. This does not include any information that may be held within the 'Items' and 'Visualization' domains.
6. *All the above*. This covers all of the domains listed above, which in aggregate completely define a state of ALBERT.

The ability to load each of the domains listed were also made accessible to the user via command line options for added usefulness. Please refer to the User Guide for usage information (Appendix E).

5.15 Some example real-time applications

5.15.1 Examples based on the state of one parameter

5.15.1.1 Adult singers

As discussed previously, a real-time display based upon quantitative evaluation of the singer's formant phenomenon may be of use in the process of singing tuition (section 5.8.1). The ALBERT software could be configured to function as a tool for this purpose based on the Ratio parameter. The parameter could then be displayed in a number of ways, such as a plot on a 2D graph with F0 or time parameters mapped to the x-axis.

5.15.1.2 Child singers

A similar phenomenon has also been observed in child singers. Current research has indicated that children phonating at high fundamental frequencies exhibit increased spectral amplitude in the 700-1800Hz region as a function of vocal training

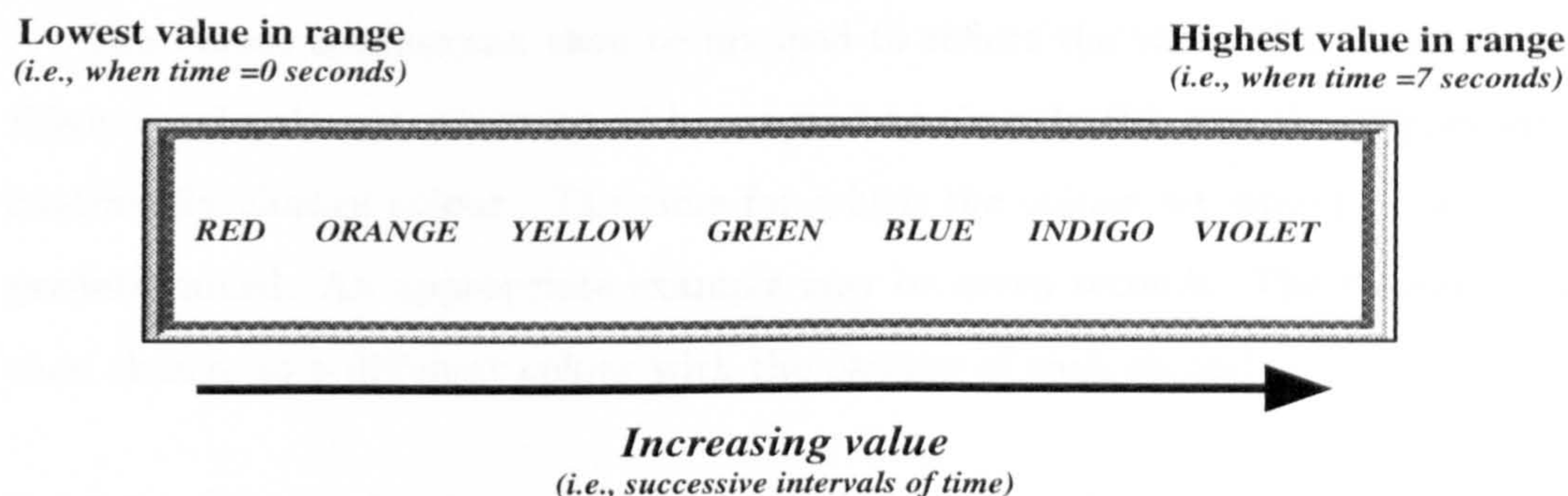


Figure 5.26: An example colour scheme for use in mapping time as a mechanism to encourage attention

(White et al., 1994). A real-time display based upon quantitative evaluation of this phenomenon may therefore be useful in the tuition of child singers.⁷

This could be achieved by using the same process as that described for adult singers, but with selection of the frequency range 700-1800Hz instead of 2-4KHz. A further modification may be to de-select the drawing of the axes in the display, which may serve to confuse or distract the subject.

5.15.1.3 Subjects with short attention span

A number of voice users are unable to concentrate for long periods of time. This includes children and adults with learning difficulties. The ALBERT software can be configured to complement the display of a parameter with colour in a number of ways. This facility can be used to create a real-time display which continually changes colour.

This first requires the design of an appropriate colour scheme. An example is illustrated in figure 5.26. All the colours of the rainbow have been employed in the colour set. The lowest data values are then mapped to the first colour in the range (red), with the highest data values mapped to the last colour in the range (violet). Values between these extremes are mapped to the other colours according to their

⁷The potential usefulness of this configuration appears strengthened in view of Moore's (1994) comments that 'the absence of any guide to follow' is one of four major causes of inaccurate child singing.

relative position between the two extremes.

This colour scheme can then be mapped to reflect the value of a parameter. In this example, the colour set would be mapped to *time*. In this way the display would continually change colour. The time for which the colour set would cycle can be predetermined. An appropriate example may be seven seconds. The display would then change to a different colour with the passing of each second.

5.15.2 Examples based on the performance of two voice parameters in combination

5.15.2.1 Introduction

In section 4.4.9, it was demonstrated that the performance of many voice parameters are interlinked. The suitability of a real-time display of one of the parameters may be increased by deriving an algorithmic measure which takes into account the performance of one or more other parameters at the same time instant. The following examples illustrate some tuition contexts in which such an adjustment is appropriate, and the way in which ALBERT may be configured to accommodate these contexts.

5.15.2.2 The display of CQ with regard to SPL changes

In section 3.6.6, it was concluded that a real-time display of the level of CQ of the vocal folds may be of considerable benefit in the tutoring of vocal efficiency. In chapter 3 it was shown that male subjects exhibit progressively higher levels of CQ with increasing levels of acoustic output (SPL). A real-time display of CQ may therefore appear to demonstrate improved efficiency which has in fact been expressed only as a consequence of increased output volume level. An adjustment is required in which the effects of increased SPL are accounted for in the monitoring process. This can be achieved by taking the two coefficients m and c which have been previously determined for the linear relationship $y = mx + c$. In this case, $m = 0.4$ and $c = 41$. Using the ALBERT system, these may be entered into formula G.5 presented in Appendix G, with the parameter name 'cq' substituted for the variable y , and the parameter name 'SPL' substituted for the variable x . The result

of this formula may be given a new label, such as 'quality' or 'target', and selected for display in any of the visualization styles discussed previously. Any increase in CQ expressed by the subject is then assessed relative to the volume level of acoustic output, and subsequently displayed in real-time.

5.15.2.3 The display of Ratio with regard to SPL changes

In a context similar to that of a real-time display of CQ, a display of the level of spectral amplitudes in the singer's formant region may be useful in encouraging a professional quality in the singing voice. However, in the second voice study this parameter was also shown to increase in proportion to the level of SPL output (section 4.4.9; figure 4.20). A real-time display of the level of spectral amplitudes in the singer's formant band may therefore appear to demonstrate improved efficiency which has in fact been expressed only as a consequence of increased output volume level. An assessment of this parameter which has been adjusted to take into account the effects of loudness would be more appropriate. This may be achieved by the same process as that described for the real-time display of CQ, with the parameter name 'Ratio' substituted for the parameter name 'CQ'.

5.15.2.4 The display of development along a CQ/singer's formant continuum

After the assessment of developing voice parameters presented in chapter 3 it was concluded that for male subjects there was a strong correlation between the level of CQ and the level of spectral amplitude in the singer's formant frequency band. This is illustrated in figure 5.27. The study also concluded that there were indications that developing voice users could be regarded as aiming to improve vocal performance along the CQ/Ratio line of correlation illustrated. In order to create a single parameter representative of progress along this continuum, an equation was derived (number G.4 in Appendix G). Based upon this equation, a new parameter can be created and subsequently displayed for use as a real-time indicator of progress along the line of continuum.

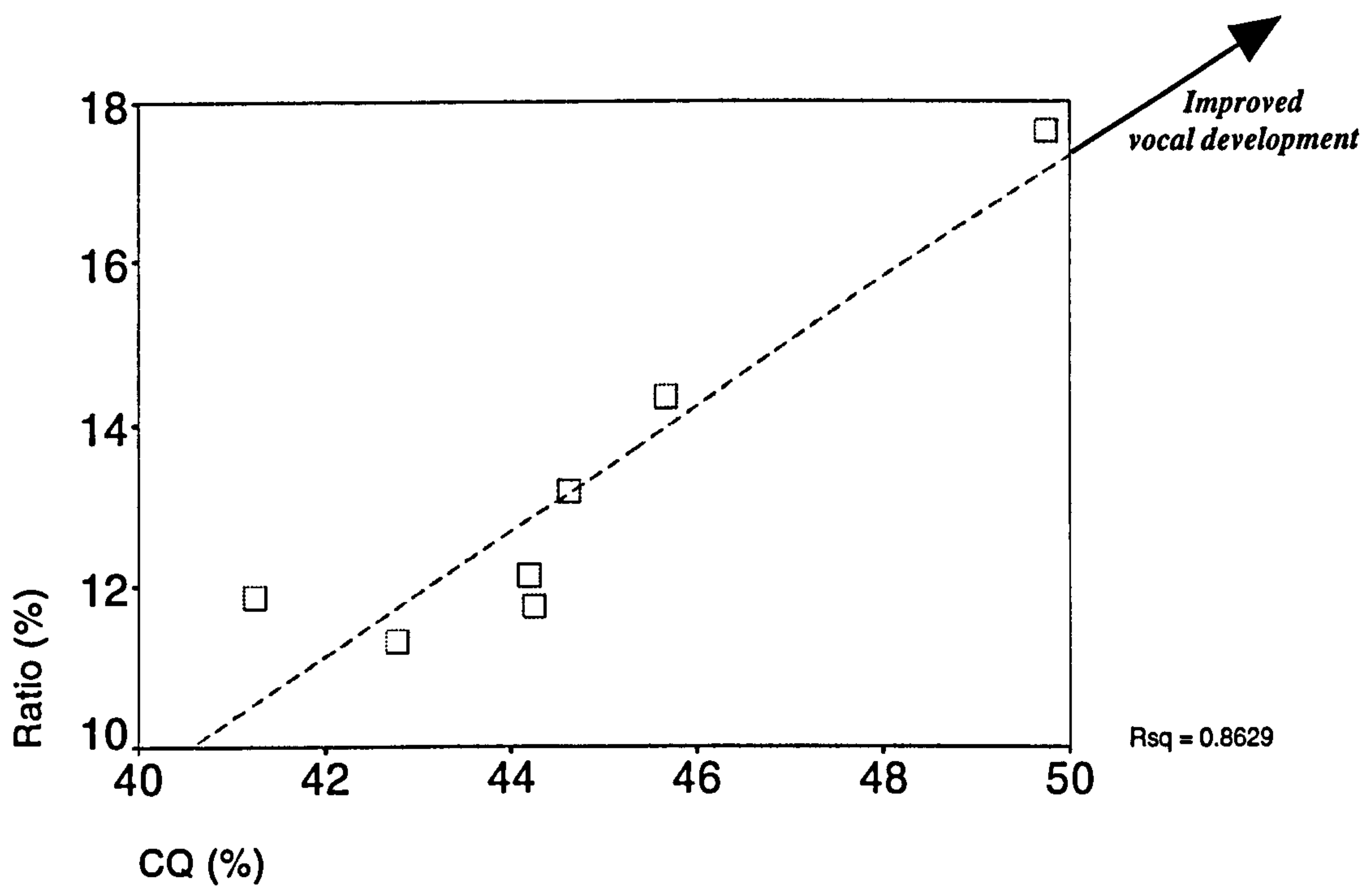


Figure 5.27: Mean longitudinal CQ and Ratio data for 6 male subjects singing the nursery rhyme 'Baa Baa Black sheep' across an eight month period of vocal training (from voice study 2)

5.16 An examination of the product compared to the original requirements

In this section the product is judged with regard to the original formal requirements (section 5.4). For this task ALBERT was run on a Silicon Graphics Indigo computer.⁸ The original requirements of the system are listed together with a description of the way in which these requirements have been met.

1. *Analysis of parameters derived from the acoustic signal*

That the system is able to analyse the following parameters from the acoustic signal:

- (a) the spectral distribution of the voice output (Ratio)*
- (b) the loudness of the voice (SPL)*
- (c) the change in loudness from one time instant to the next (shimmer)*

Provision has been made for the system to analyse each of these parameters (sections 5.8.1 and 5.8.2).

2. *Analysis of parameters derived from the electrolaryngograph signal*

That the system is able to analyse the following parameters from the electrolaryngograph signal:

- (a) the fundamental frequency of the voice (F0)*
- (b) the closed quotient of the voice (CQ)*
- (c) the rate of opening of the vocal folds (oa)*
- (d) the rate of closing of the vocal folds (ca)*
- (e) the peak-to-peak amplitude of the electrolaryngograph period (lx amp)*
- (f) the change in amplitude of the electrolaryngograph period from one time instant to the next (lx shimmer)*
- (g) the period-to-period change in fundamental frequency (jitter)*

⁸There are several different models which operate at a faster rate than this model (such as the Silicon Graphics Indy computer). However, access to these machines was not available.

Provision has been made for the system to analyse each of these parameters (sections 5.9.2, 5.9.3, and 5.9.4).

3. *The graphical user interface*

The user must be provided with easy-to-use controls over:

- (a) the processing and display of acoustic signal information.*
- (b) the processing and display of electrolaryngograph signal information.*
- (c) the properties of the voice parameters derived from the two input signals.*
- (d) the way in which the state of the voice parameters may be visualized.*

The user is provided with a series of forms. Each form contains controls concerning a particular aspect of system function. All of the aspects listed may be easily controlled by the user through the controls presented on these forms (sections 5.6.1 and 5.6.2). Furthermore, several properties of the user interface may be controlled by the user for enhanced ease of use (section 5.6.3).

4. *Visualization*

The user must be provided with the ability to visualize the state of voice parameters in 1D, 2D and 3D display configurations. For reasons of clarity, a visualization must employ a minimum of 15% of the complete screen display.

The user is supported with control over a number of visualization configurations, including 1D, 2D and 3D displays. Colour mapping may optionally be employed for enhanced presentation of information (sections 5.12.3 and 5.12.4).

5. *Information on how to use the system*

The user must have easy access to information on how to use the software.

The user guide has been arranged on the world wide web to provide easy access, at a centralised information point (section 5.5).

6. *Real-time performance*

The system must be able to derive all voice parameters listed and visually present the state of any parameter at a frequency of not less than 10Hz.

An assessment of the ALBERT system running on a Silicon Graphics Indigo computer has demonstrated that this criteria has been met.

Appendix F contains an assessment of the performance of the system in evaluating each of the voice parameters.

5.17 Conclusions

This chapter has considered the nature of visual feedback for real-time voice analysis and subsequent enhancement of vocal qualities. It has introduced a highly configurable system which may be used in real-time as an aid for voice tuition in a number of contexts, or in non-real-time as a tool for voice analysis. The system is able to support advanced analysis and display of parameters derived from voice source and acoustic output signals. The user is able to control many attributes of the way in which the system functions so that the most appropriate form of analysis and display may be arranged. This includes several properties of the graphical user interface, the rate of analysis and display, parameter control and visualization configuration. The arrangements may be stored to file.

Several examples of configurations for real-time voice feedback have been presented.

The next chapter presents an investigation into the advantages of the system when used as a complimentary tool in the context of voice tuition.

Chapter 6

An assessment of biofeedback

6.1 Introduction

This chapter presents an assessment of the usefulness of the ALBERT system when used to provide visual feedback in real-time. The development of several parameters relating to vocal usage in two subjects have been considered as the subjects underwent consecutive sessions of singing tuition with a professional teacher. One subject was taught conventionally and the other was taught by the tutor with the aid of the ALBERT system configured to provide visually presented biofeedback based on the CQ and Ratio parameters identified in chapter 3 as indicative of vocal development. These parameters were presented as visual feedback both in solitude and then in combination.

Previously, a discussion of the benefits of real-time visual feedback noted that several reports have indicated a relatively greater progress in the process of vocal tuition when subjects were taught with the aid of graphically presented biofeedback (i.e., Michi et al., 1993; Welch et al., 1989) (section 5.2). However, some studies neglect to employ a reference subject (i.e. Wirz & Anthony, 1979; Oster, 1988; Allen et al., 1991). This means that any observed trends are subject to the criticism that they may have occurred regardless of the use of biofeedback. A formal study into the role of biofeedback in vocal development may benefit by employing a reference subject against which the progress of a subject taught using biofeedback can be compared. This is the approach taken in this study.

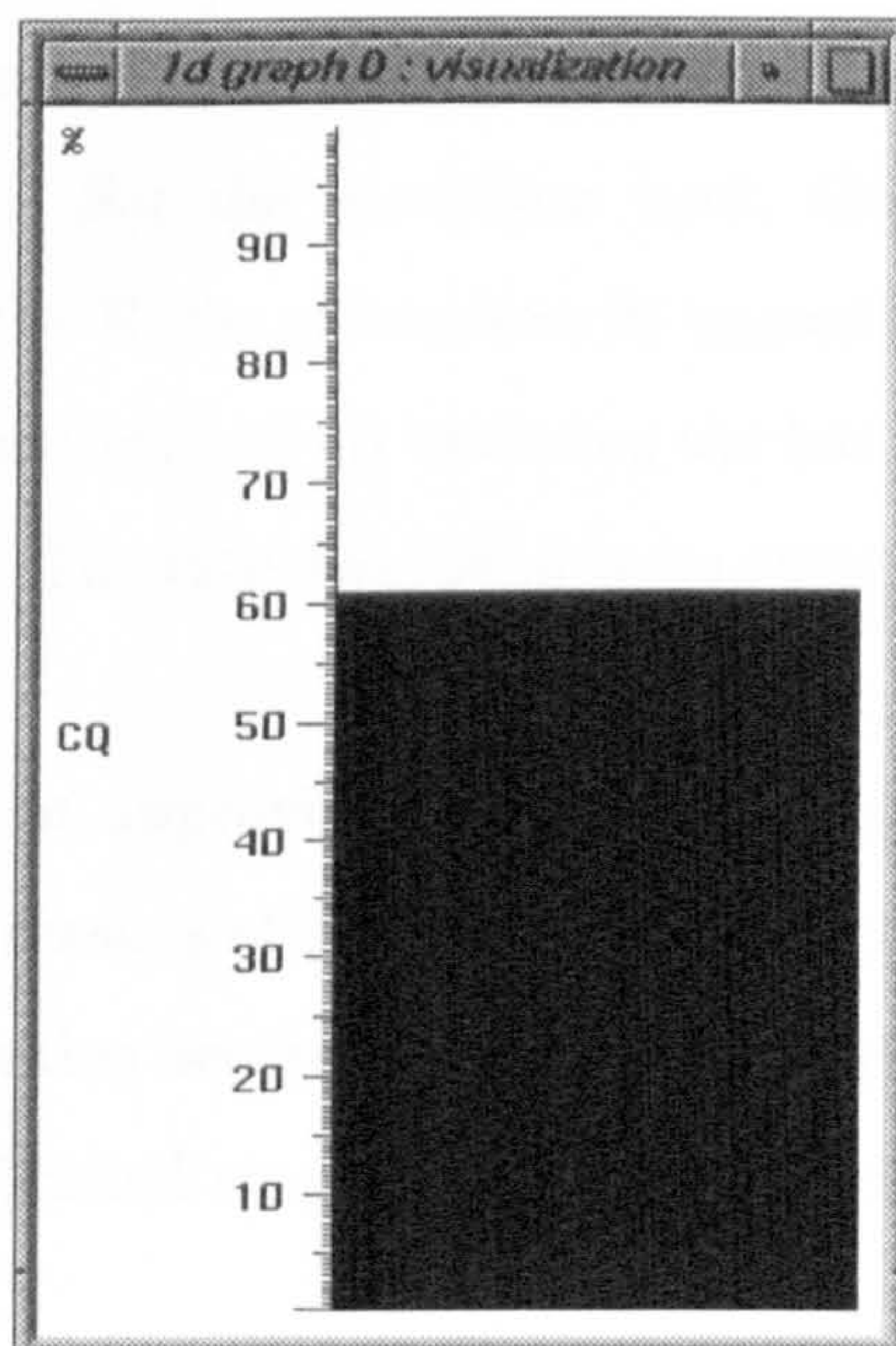


Figure 6.1: Bar graph display used for real-time biofeedback

6.2 Tuition

Two male subjects were selected for the study. Neither had any previous experience of vocal tuition. A total of six periods of singing tuition were received by each subject. Across the six sessions, subject A was taught using conventional techniques with instruction being provided aurally by the tutor, with appropriate practical demonstrations. Subject B, in contrast, was additionally provided with visually presented biofeedback of the CQ and Ratio parameters. These were employed in the following order:

1. CQ - used during lessons 1 and 2
2. Ratio - used during lessons 3 and 4
3. CQ and Ratio combined - used during lessons 5 and 6

For (3) above, the CQ and Ratio parameters were assessed relative to progress along the linear correlation between the two parameters observed during study 2 (see section 4.4.9; figure 5.27). Details are given in Appendix G.

Information was presented as a one dimensional graph identical or similar in appearance to figure 6.1. The biofeedback update rate was twelve times a second.

This rate was based upon the assessment of the demands of the human visual system discussed in the previous chapter and the real-time load which the ALBERT system is required to handle. For the maximum load, this involved i) engaging the fft process for assessing the Ratio parameter, ii) engaging the voice source process for assessing the CQ parameter, and iii) updating the bar graph display of the parameter presented as feedback. The rate was found to be comfortable and appropriate during use.

Lessons were held at approximately the same time of day, on the same day of successive weeks. For reasons of diary incompatibility a delay of one week had to be introduced between tuition sessions 3 and 4, and between sessions 5 and 6. Lessons were given by an established singing tutor and professional singer.

6.3 Recording

In order to assess their speaking voice, subjects were required to read aloud a sequence of phonetically balanced text frequently used for speaking voice assessment. In order to assess their singing voice, subjects were required to sing a two-octave scale up and down, with inhalation between each octave. The assessment procedure is the same as that used in assessing the subjects for study 3 (see Appendix D). A frequency range for the singing assessment task was selected by the tutor for each subject individually. In terms of frequency range, the same allocated two octave scale was sung for each recording session. Subjects were initially recorded in the hour preceding the first tuition session. Subsequent recordings were made in the hour following each of the tuition sessions. Singers were not specifically asked to warm up prior to recording. Subjects were not given tuition or biofeedback during the recording period, but were simply instructed to perform the two assessment tasks as well as they could. The times taken by the subjects to read the spoken passage and sing the two octave scale were typically 80 seconds and 25 seconds respectively.

Stereo recordings were made using a Sony TCD-D10-PRO DAT recorder at a sampling frequency of 48KHz. The acoustic output of the subjects was monitored via a Sennheiser MKH/40/P48 cardioid microphone on one channel, and the output

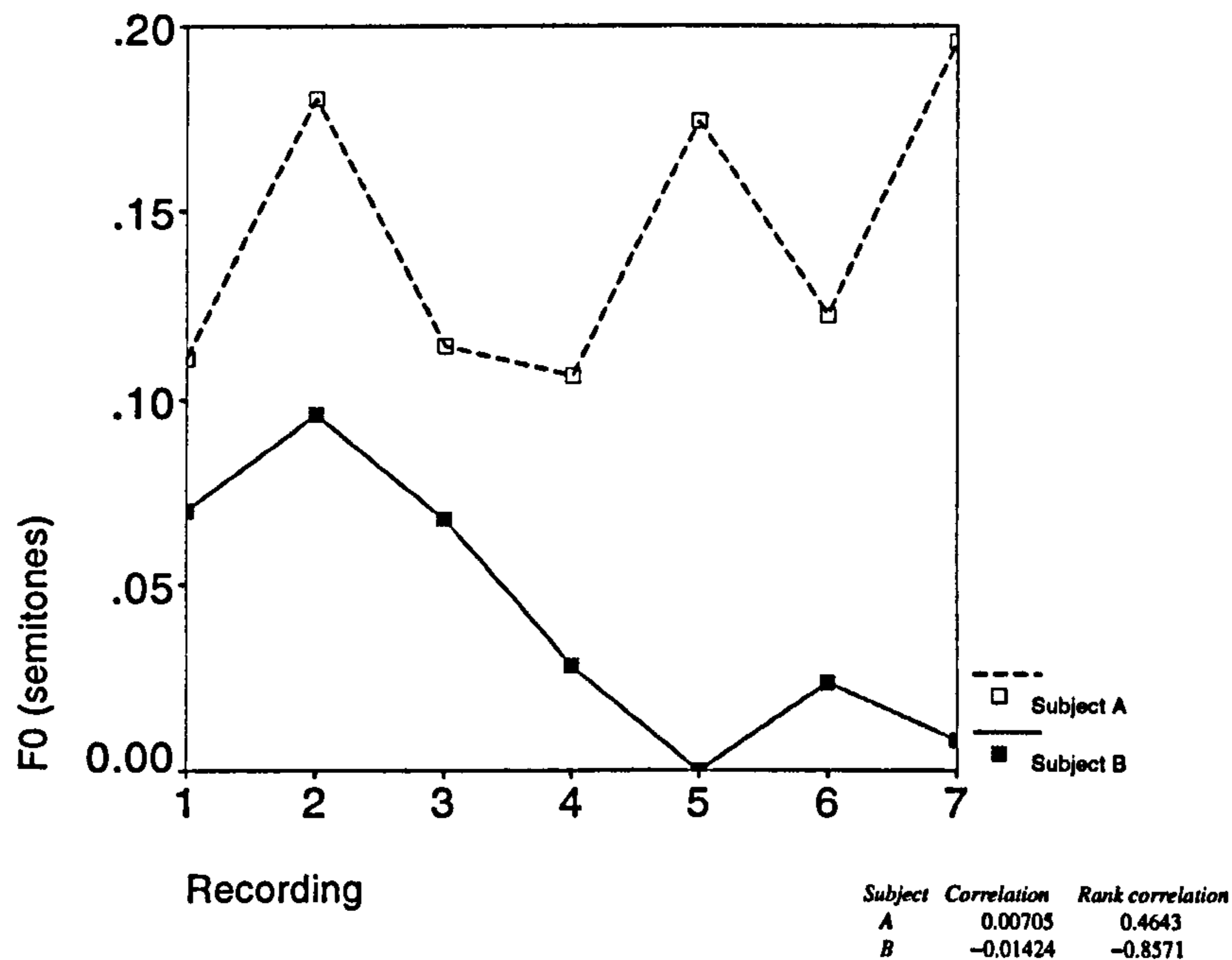


Figure 6.2: Measures of F0 across the recording sessions for the spoken passage relative to the lowest value in the data set

from an electrolaryngograph was recorded on the other channel to monitor the area of vocal fold contact. Subjects were required to maintain a constant distance of approximately 0.3m from the microphone. Prior to recording, subjects were required to provide an indication of the loudest level at which they were to phonate and appropriate recording levels were set to ensure clipping would not occur. These remained unaltered for the duration of the session. Subjects were then required to phonate at a constant volume for a period of several seconds during which the SPL level at the position of the microphone was recorded by means of a DAWE D-1422C digital impulse sound level meter. The assessed level was then used as a reference for subsequent analysis of SPL for the recorded data.

6.4 Results

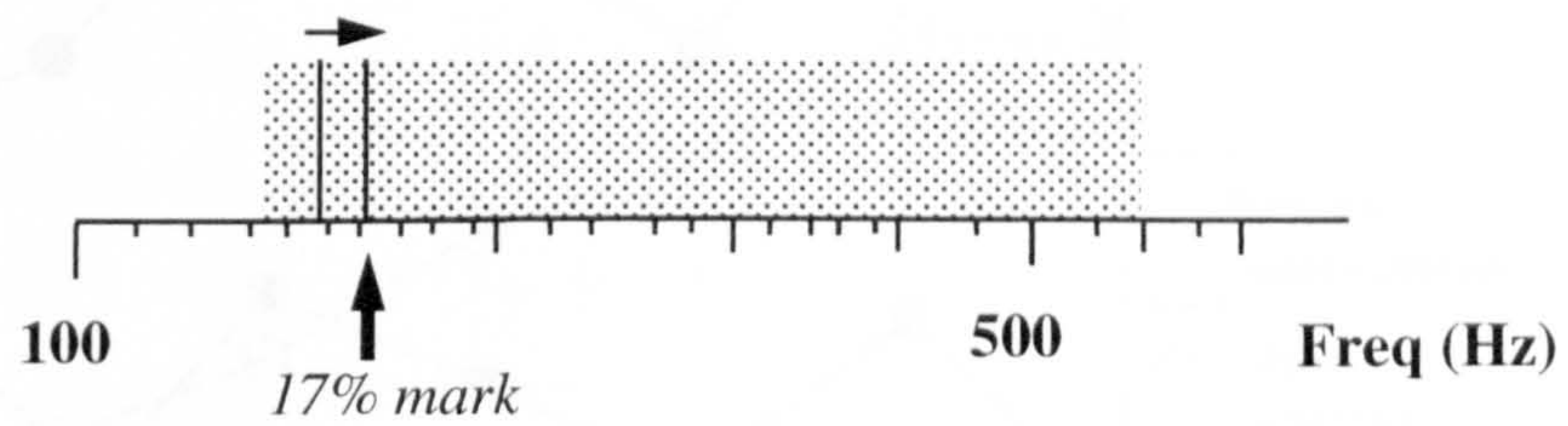
6.4.1 Fundamental frequency

Results for the level of mean spoken passage fundamental frequency (F0) are

shown in figure 6.2. Subject B exhibits a clear, generally decreasing trend in mean F0. Subject A exhibits a relatively higher level of fluctuation.

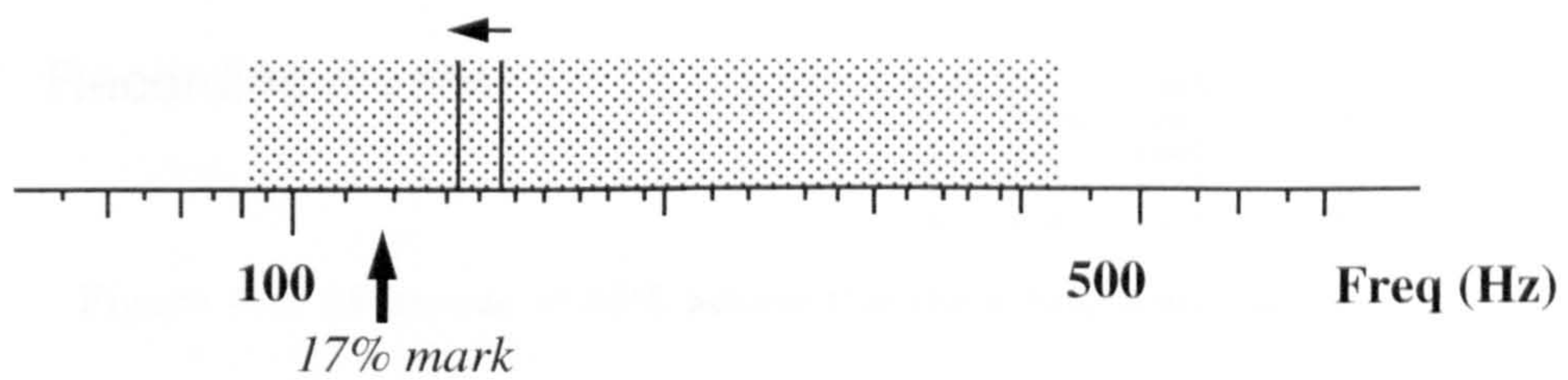
In the analysis of study 1, it was found that there was a statistically significant increase in mean F0 for a group of male and female subjects after a short period of vocal tuition. It would be logical to expect the subjects in this study to exhibit generally increasing mean speaking voice F0. An initial observation of the data does not lend itself to this theory. During the first period of singing tuition in our study, the singing tutor identified a target range for the singing voice of both subjects based upon a professional assessment of their vocal potential. Subsequent tuition was based upon realising the potential of both subject's voice with regard to their target range. However, research by Gramming (1991) into adult voice usage concluded that the mean F0 value for normal speech was at approximately 20% of the total fundamental frequency range in semitones. Awan (1993) reported a similar result of 12-16%. Both authors noted that this level was essentially identical for trained and untrained subjects. In the following assessment, a compromise of 17% is used.

Figure 6.3 illustrates the first and last recorded mean speaking voice F0 data relative to the F0 frequency span after the final singing lesson. This is shown for both subjects. For subject A, the initial mean F0 was approximately 7% of the final F0 range. The pattern of change of F0 illustrated for this subject in figure 6.2 demonstrates a generally unclear trend, although both linear and rank correlations imply a gradual increase in F0 as a product of training. The subject may be regarded as exhibiting a gradual trend towards the 17% level. The indistinct pattern expressed by subject A may be regarded as a consequence of the close proximity of the subject's mean F0 to this target level. For subject B, the initial mean F0 for speech was approximately 30% of the final F0 range. The pattern of change for this subject (figure 6.2) demonstrates a generally decreasing F0. This trend may also be interpreted as implying mean speaking voice F0 is moving towards the 17% level of a target F0 range. The clearer trend expressed by this subject may be a consequence of the greater difference between initial and target mean F0. This result provides strong indications to support Awan's (1993) proposal that *'training to influence aspects of the singing voice may also affect aspects of the speaking voice*



(a) Subject A

Initial speaking voice F0 is at approximately 7% of final F0 range

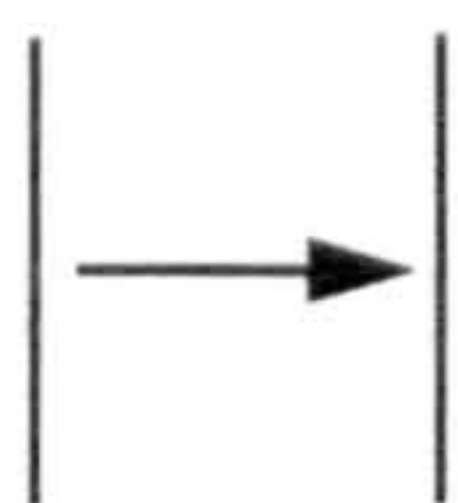


(b) Subject B

Initial speaking voice F0 is at approximately 30% of final F0 range



= Singing pitch range of subject after final singing lesson



= Initial mean F0 for speaking voice, general direction of subsequent change, and final mean F0.

Figure 6.3: A comparison of mean speaking voice F0 with singing voice F0 range assessed after the final tuition session

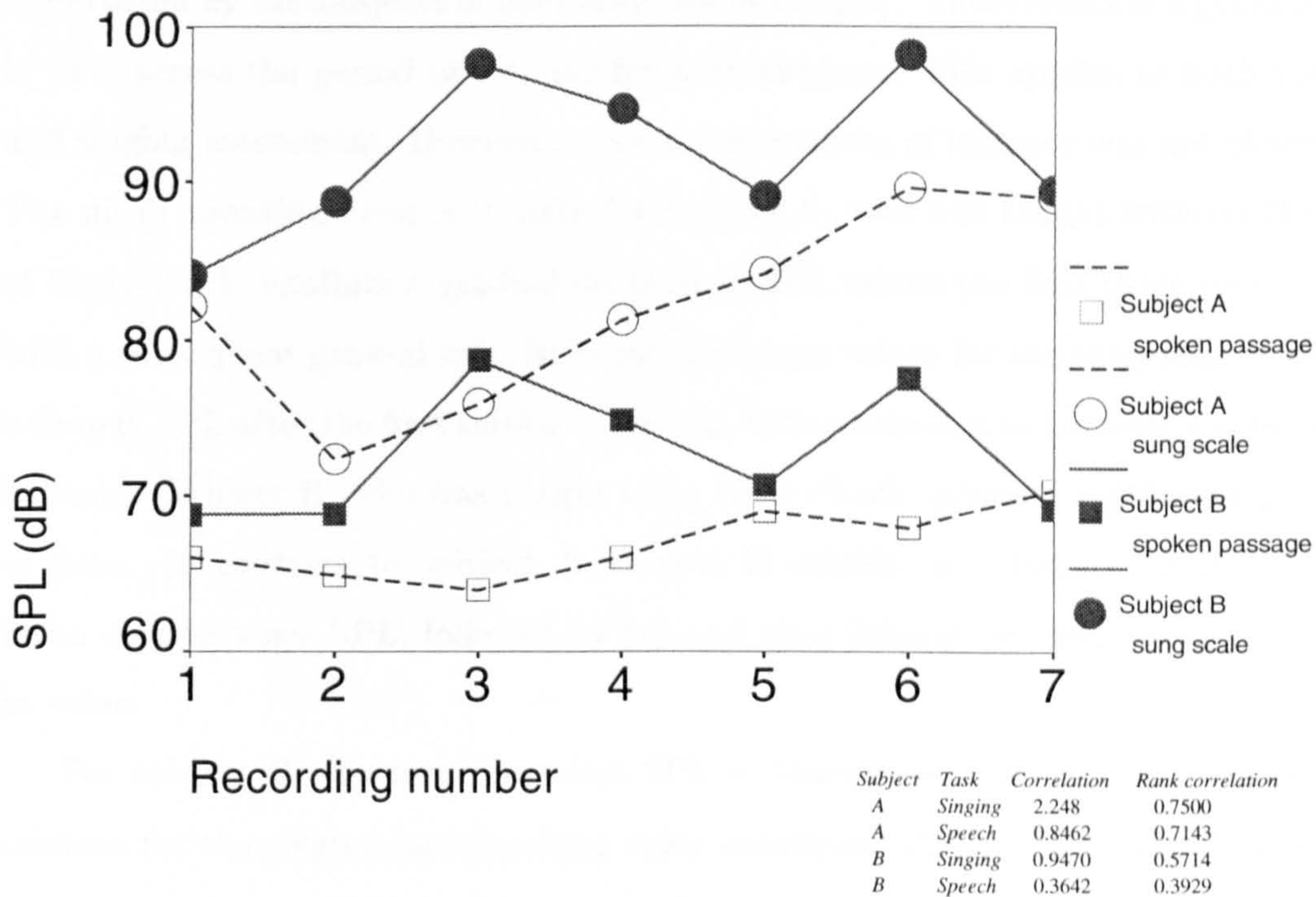


Figure 6.4: Measures of SPL across the recording sessions

in a predictable manner?

However, these observations are subject to the criticism that the observed trends in fundamental frequency may be due to some other cause such as natural variability. The results from study 1 (section 4.2), in which all subjects demonstrated an increase in mean fundamental frequency after a 6 week course of vocal expression may be further regarded as an indicator that the results discussed for the two subjects in this study must be tentatively interpreted.

The role of biofeedback in this trend is hard to identify. F0 is not amongst the parameters presented as biofeedback to subject B. A tentative conclusion is that the F0 trends exhibited by both subjects are expressed as a function of tuition in the manner discussed, and that biofeedback has not had a direct influence.

6.4.2 SPL

Figure 6.4 illustrates the pattern of change for the SPL parameter. There have been previous observations that trained vocalists use significantly greater mean intensity levels in speech than untrained vocalists (i.e., Awan, 1993). Both linear and rank

correlation measures have been applied to the speech and singing assessment tasks undertaken by the subjects in our study (see figure 6.4). These indicate a general rise in SPL across the period of training for both subjects. This applies to both speech and singing assessment. However, a consistent pattern of increase was not observed. The mean speaking voice SPL data for Subject A, who was taught without the aid of biofeedback, exhibits a gradual decrease in SPL across the first three recordings, with a subsequent general rise. However, the mean values for the sung scale exhibits a drop in SPL after the first tuition of session, with a subsequent pattern of consistent increase. Subject B, who was taught using biofeedback, exhibits a different pattern of data. In contrast to subject A, subject B exhibits a pattern of initial rise in mean singing voice SPL, followed by fall and then subsequent rise, with a final fall in value.

For subject B, the trend in mean SPL is almost identical across all recording sessions for the singing and speaking voice assessment data. This is clearly not the case for subject A. However, it may be that this is due to some form of individual difference in the appreciation of tuition, or of physiology or anatomy between the two subjects. For example, the consistency in SPL between speech and singing voice modes for subject A may be due to the usage of two different modes of voice expression, where subject B may only have used one. In addition, the study covers only a relatively small period of training. Differences in trends between subjects must therefore be carefully interpreted. The identical pattern of SPL for both speaking and singing voice expressed by subject B may be a result of the use of biofeedback, in which the subject was taught using objectively measured and displayed parameters. This may have resulted in a higher level of personal confidence in voice usage, resulting in a higher level of influence on the speaking voice.

In terms of the general fluctuations observed, it must be noted that the influence upon SPL of the tuition which the subjects received is hard to assess. It would appear to be reasonable to conclude that fluctuations in SPL would occur as a long term function of tuition, reflective of shifts between different aspects of vocal production during the process of tuition and development.

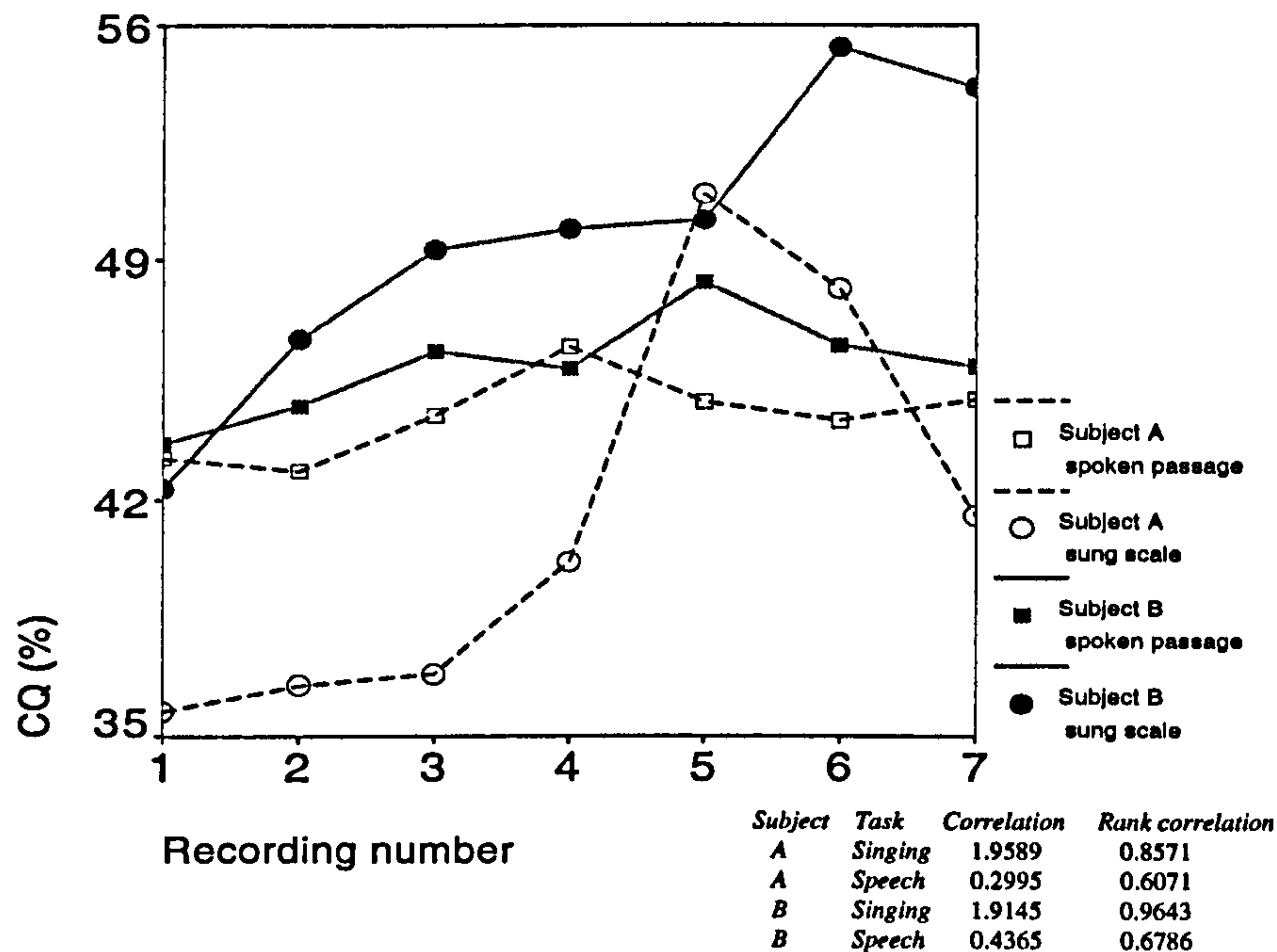


Figure 6.5: Measures of CQ across the recording sessions

6.4.3 CQ and Ratio

Figure 6.5 illustrates the pattern of change for the CQ parameter. For the speech data, the subject taught without the aid of biofeedback (subject A) exhibits small but generally increasing changes, with consistent rise until session 4 whereupon the CQ data dips in value, followed by negligible change. For the singing data, the subject exhibits consistent increases of gradually increasing magnitude until recording 5, whereafter there is consecutive decrease for the two remaining data. The pattern of change for both singing and singing voice data is similar, in that data for both exhibit a general rise followed by fall. However, this is of far greater magnitude for the singing data than for the speech data, and the point at which the pattern of rise changes to fall occurs at different points. These trends contrast with data for the subject taught with the aid of biofeedback (subject B) who exhibits a consistently increasing pattern of CQ for the singing task, with only a small reduction for the final value. The pattern of mean CQ values for the speech task fluctuates up and down to a similar degree of that exhibited by subject A. The pattern of change of mean CQ speaking voice data does not clearly follow that of the singing voice data. Rank correlations (see figure 6.5) assessing the pattern of

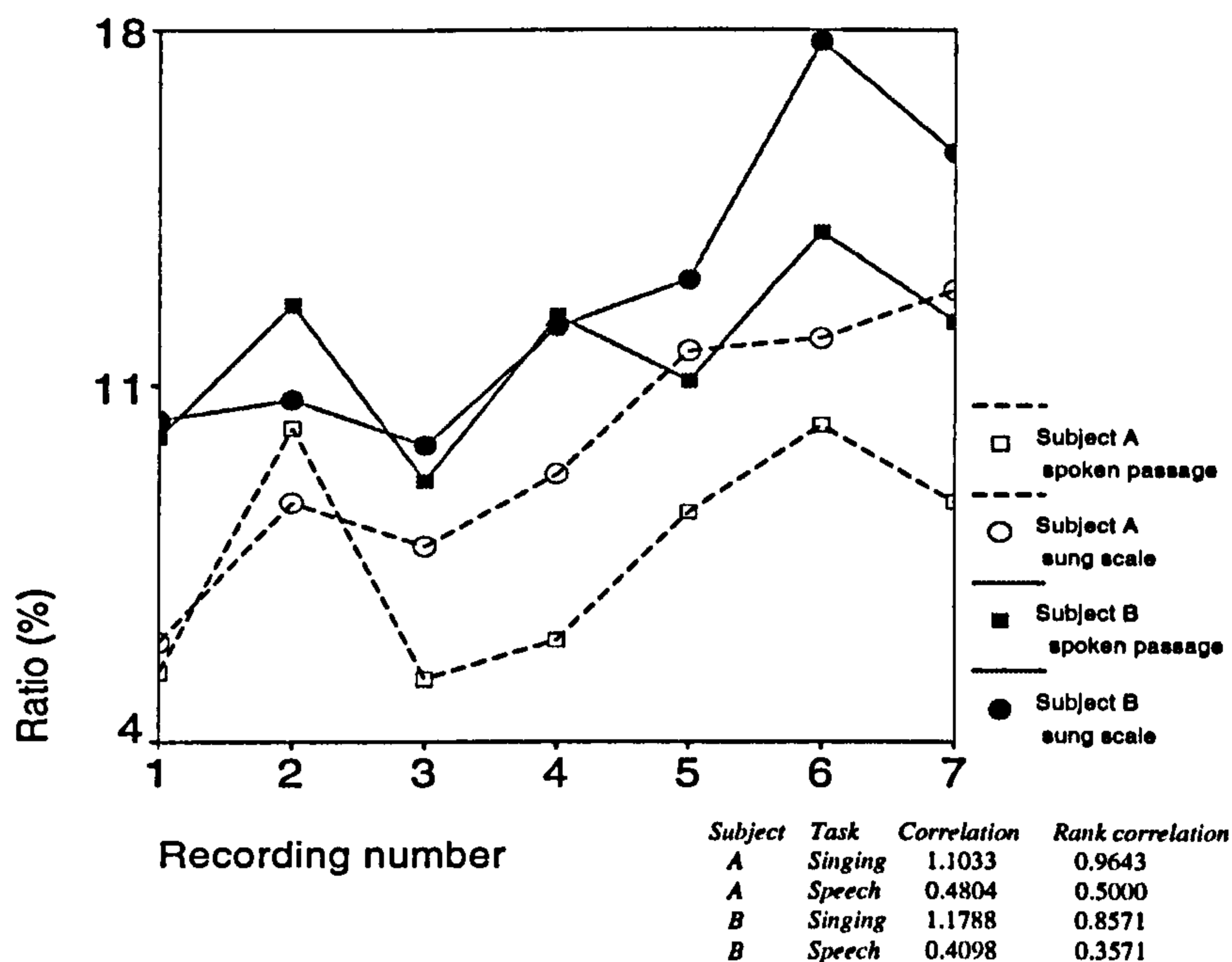


Figure 6.6: Measures of Ratio across the recording sessions.

CQ data with regard to the recording number indicate a higher level of consistent increase for the subject taught using biofeedback. Linear correlations indicate very similar levels of correlation for the two subjects for the singing voice assessment, and a higher level of correlation for the subject taught using biofeedback for the speaking voice assessment.

Figure 6.6 illustrates the pattern of change for the Ratio parameter. For the speech data, the subject taught without the aid of biofeedback (subject A) exhibits a generally consistent pattern of increase. A very similar pattern of change is evident for the singing data, with one major exception after the first singing lesson where a relatively large local increase has been recorded. The subject taught with the aid of biofeedback (subject B) exhibits a fairly consistent general increase in Ratio for the singing task, with a small dip in value for the final recording. For both subjects, the pattern of Ratio values for the speech task follows a pattern which is generally consistent to that of the singing task. Both linear and rank correlations are higher for the singing voice assessment than for the speaking voice assessment, which reflects the nature of the tuition in singing performance received by the subjects. Both forms of longitudinal correlation are similar for the two subjects.

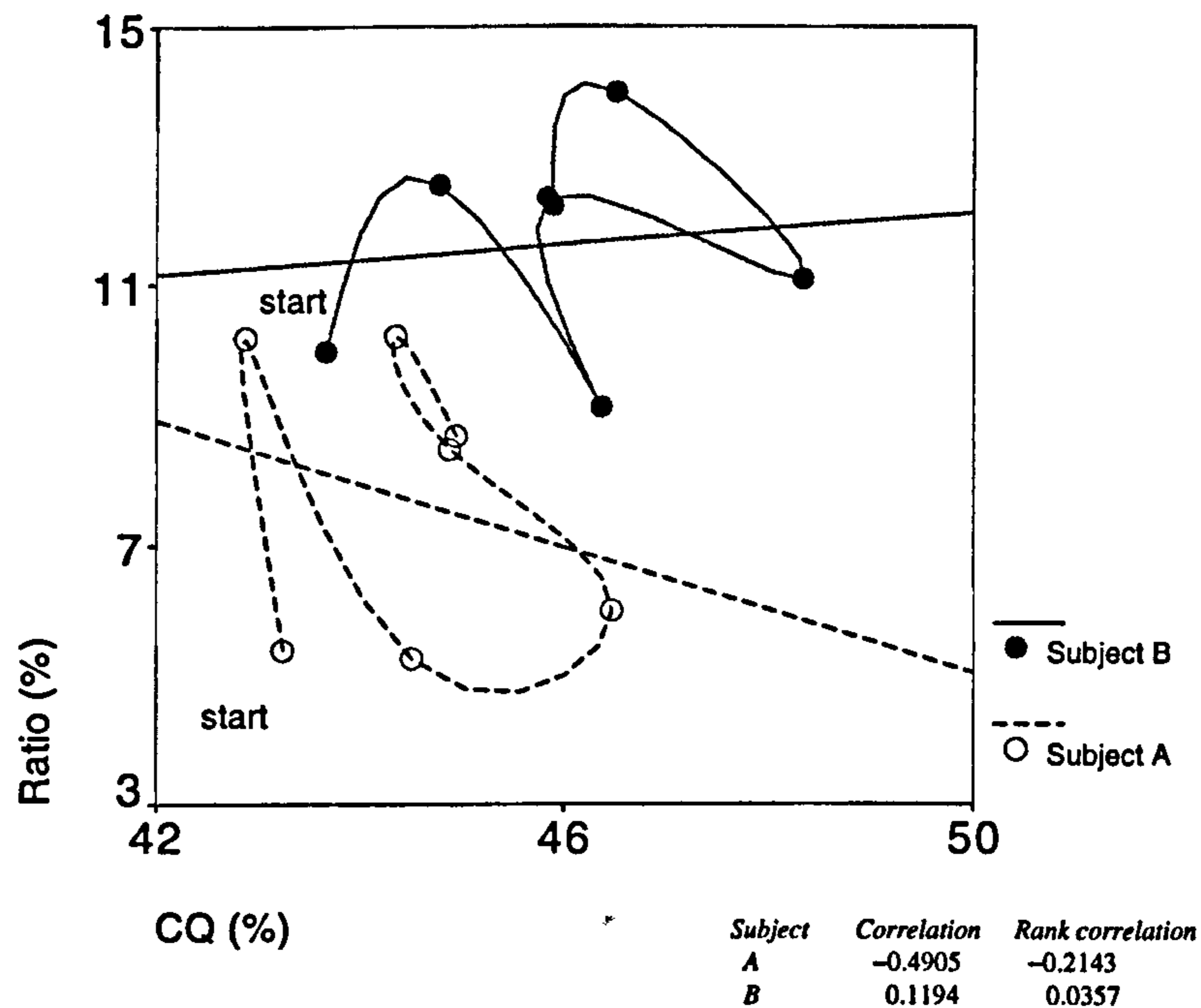


Figure 6.7: Plot of Ratio against CQ for assessment of the spoken passage with line of best fit. Spline interpolation is used to indicate one possible pattern of development between data points.

The CQ and Ratio parameters are shown plotted against each other for the speaking and singing voice assessment in figures 6.7 and 6.8 respectively. For the speaking voice assessment, it can be seen that the slope of the line of best fit for the subject taught without using biofeedback (subject A) is negative. This contrasts with the positive correlation for the subject taught using biofeedback (subject B). It may be that the biofeedback employed by subject B has reinforced the state of vocal development for the subject, resulting in a higher level of confidence in the use of the same mode of vocal usage for speaking voice expression. Alternatively, it may be that the data would have been observed even if biofeedback had not been employed.

To illustrate the way in which the electrolaryngograph and spectrum signals reflect the observed longitudinal change in the CQ and Ratio parameters, representative frames were derived from the recordings for subject B. These are shown in figure 6.9.

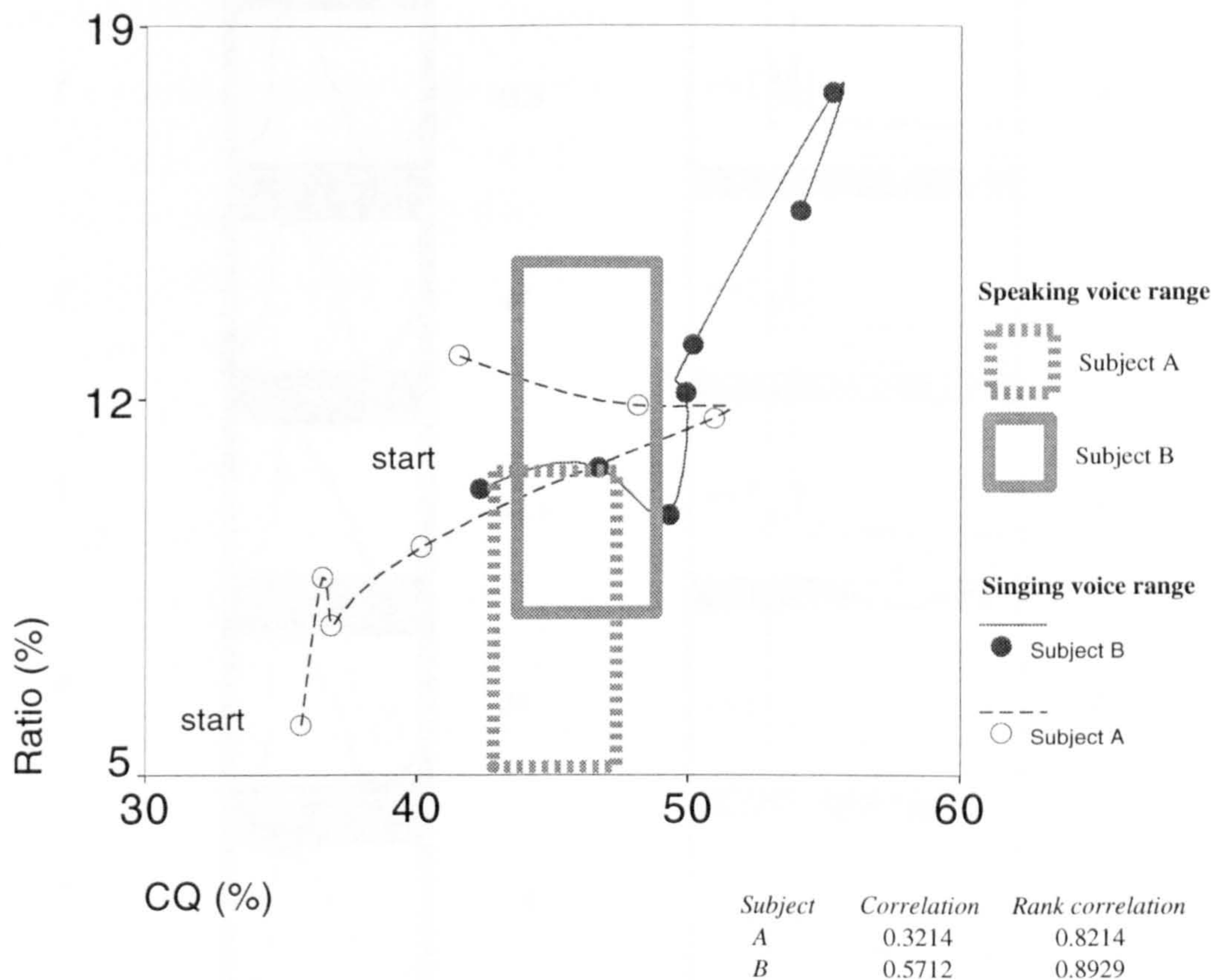


Figure 6.8: Plot of Ratio against CQ for assessment of the sung scale. Spline interpolation is used to indicate one possible pattern of development between data points. The domain of the two parameters for speaking voice assessment shown in the previous figure are superimposed.

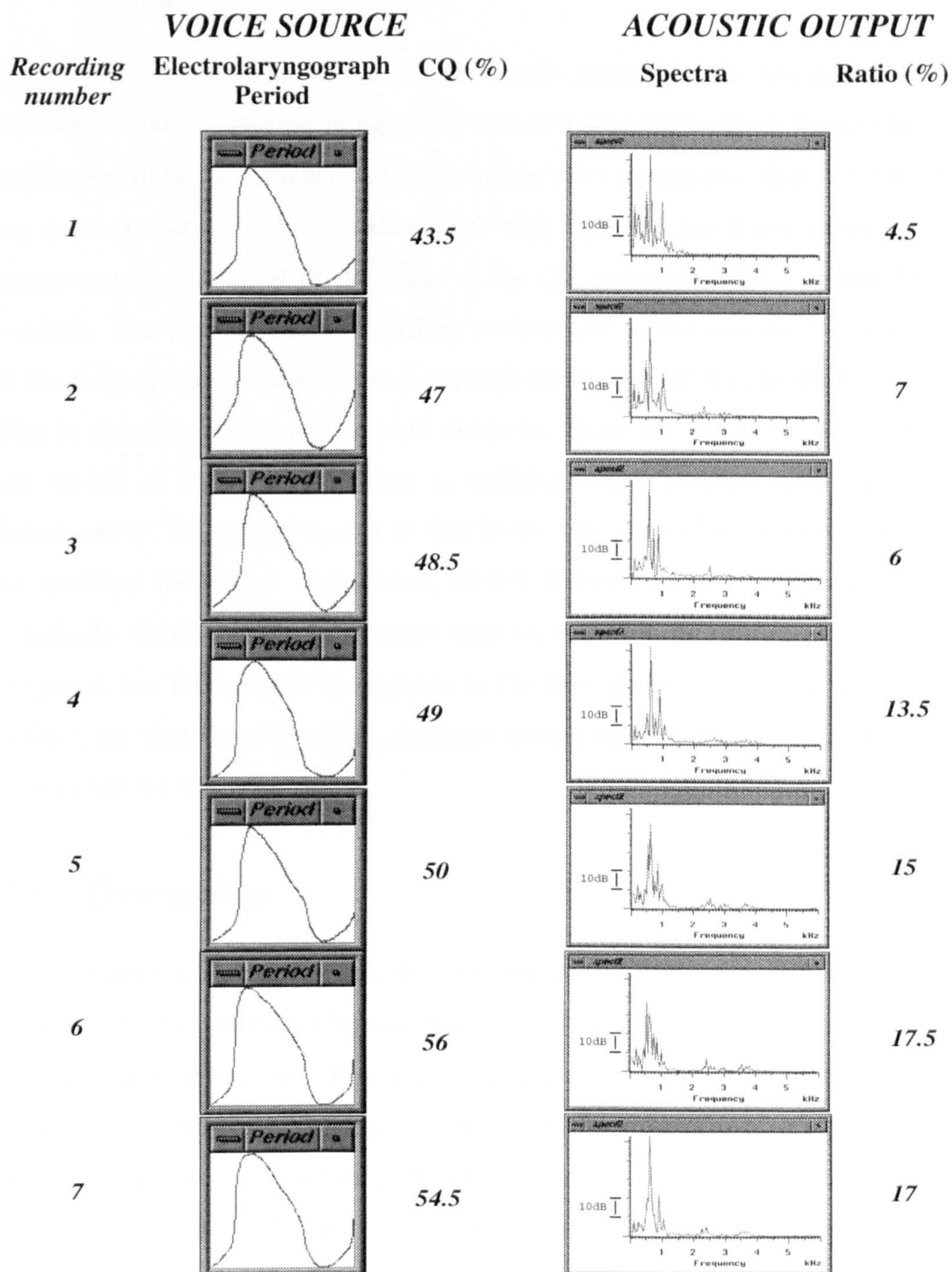


Figure 6.9: Representative images of electrolaryngograph and spectrum activity across the seven recordings for subject B

6.4.4 A comparison between performance of the speaking and singing voice

In figure 6.8 the domain of the CQ and Ratio parameters for the speaking voice assessment data illustrated in figure 6.7 is shown superimposed on the singing voice assessment data. In comparison to the singing voice assessment, data for the speaking voice exhibits a slightly smaller range with regard to the Ratio parameter, and approximately one third of the range of the CQ parameter. This applies to both subjects. The progression of CQ/Ratio performance across successive assessments of the singing voice is shown to pass through the domain of the speaking voice data. This is also shown to apply to both subjects. However, the subject taught without the aid of biofeedback, subject A, exhibits a shift in direction at recording 5. Subsequently, the subject's path of CQ/Ratio data passes back over the domain of his speaking voice data. In contrast, subject B continues in a generally increasing CQ/Ratio direction. This difference may be regarded as further indication that subject A has retrogressed in contrast to the development of subject B. It may be conjectured that the biofeedback employed during the tuition of subject B is largely responsible for this contrast.

6.5 Discussion

In chapter 4, a strong correlation was observed between the CQ and Ratio parameters. In this section, CQ and Ratio data is compared with the second study presented in that chapter. This study has the most comparable time scale of the three studies presented. To recall, study 2 involved analysis across an eight month period of a group of 6 male subjects studying acting skills. The corresponding CQ and Ratio data for both speech and singing tasks have been superimposed on the data derived for subjects A and B. Data for the speaking voice and singing voice assessment are shown in figures 6.10 and 6.11 respectively. For the speaking voice assessment, the span of CQ/Ratio data for the subjects from the previous study is similar to that expressed by subjects A and B. However, the correlation between the two parameters appears stronger for the subjects of the previous study. This is likely to be due primarily to the greater emphasis on speaking voice development in

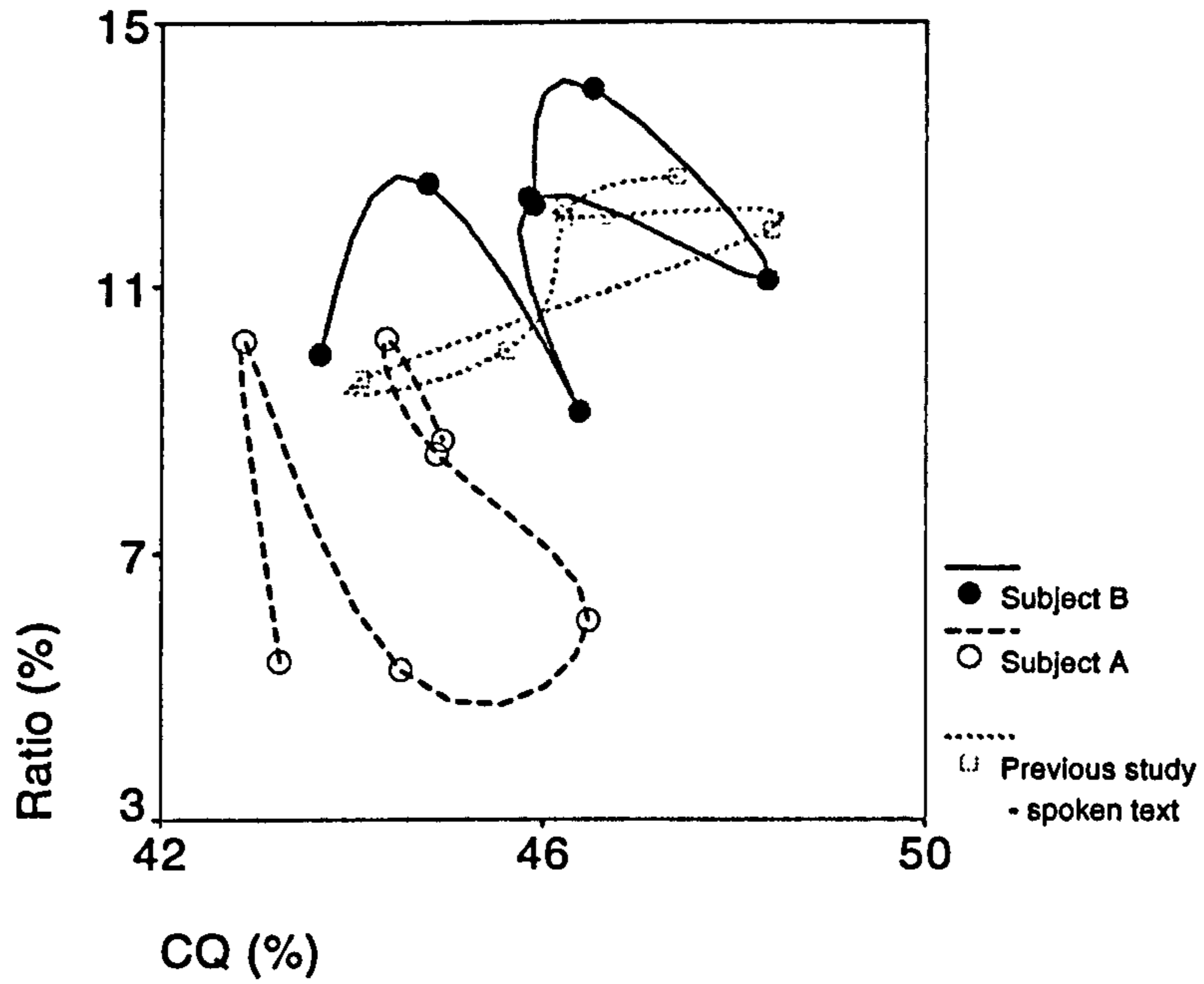


Figure 6.10: Plot of Ratio against CQ for assessment of the spoken passage with data from longitudinal voice study 2 shown superimposed.

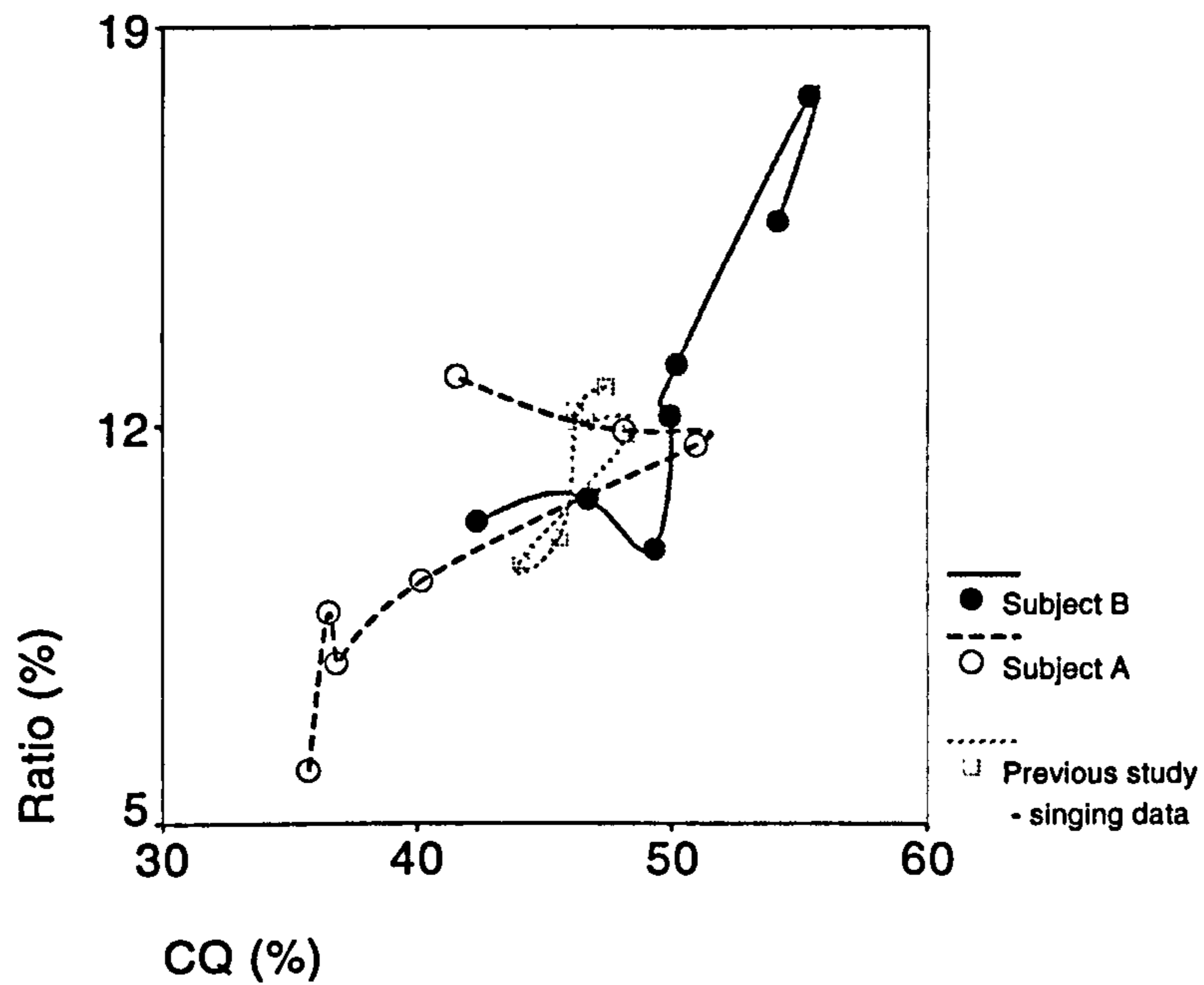


Figure 6.11: Plot of Ratio against CQ for singing assessment with data from a longitudinal voice study 2 shown superimposed

the training of these subjects, and secondly to the relatively greater number of subjects. For the singing voice assessment the span of both CQ and Ratio data for the subjects of the previous study is far smaller than that expressed by subjects A and B. This is also likely to be a reflection of the greater emphasis on singing tuition in the vocal development of subjects A and B. Subjects from the previous study studied primarily acting, rather than singing, skills. They undertook singing lessons at relatively less frequent intervals. Furthermore, the style of singing tuition was not that of the Bel Canto technique adopted by the tutor in the tuition of subjects A and B. In addition, because some subjects were unable to confidently complete a two-octave scale at the start of the recordings made for study 2, the nursery rhyme 'Baa Baa Black Sheep' was used for an assessment of the singing voice instead of a two-octave scale. This form of assessment may be regarded as placing less of a demand on the performance of the singing voice for many subjects, and this may also be reflected in the contrast in assessed CQ/Ratio development.

Figure 6.12 illustrates the trend in data for CQ and Ratio parameters for the subject taught using biofeedback. With regard to the CQ and Ratio parameters, it is interesting to note the small influence on one parameter when biofeedback is presented concerning the other. When biofeedback was presented based on the level of CQ (reflected in figure 6.12, intervals 1 and 2) there was a relatively high increase in the assessed level of CQ. During this period, the Ratio parameter exhibited a slight rise followed by a slight fall in value. When biofeedback was presented based on the level of Ratio (figure 6.12, intervals 3 and 4), the CQ parameter exhibited only very slight increases in value. When biofeedback was presented based on both parameters combined (intervals 5 and 6), there was an initial increase and subsequent decrease in the singing voice assessment for both parameters together. For each of the three occasions in which a previously unobserved parameter is employed as biofeedback there is a large advance in the formal assessment of that parameter for the first tuition session in which it is used, followed by relatively less advance for the second tuition session. This was actually expressed as a reduction in the case of the CQ and Ratio parameters combined (interval numbers 5 and 6). Two conclusions may tentatively be drawn;

1. the visually presented biofeedback in this study has been absorbed as part of

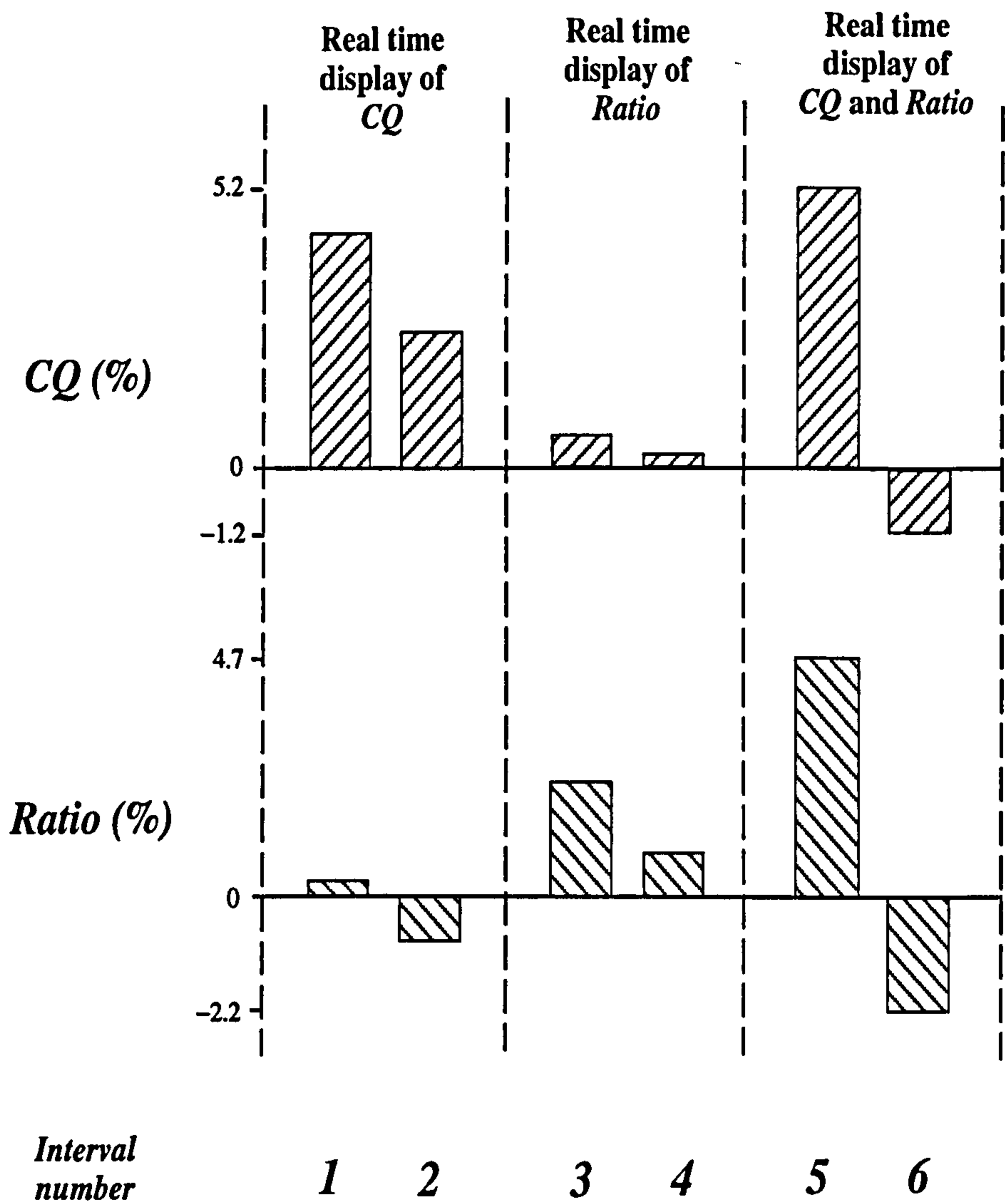


Figure 6.12: Changes in CQ and Ratio from one assessment to the next for subject B

the subject's voice usage technique to an extent and subsequently expressed in formal assessments, and

2. the greatest appreciation of biofeedback based on a particular parameter occurs during initial use of the biofeedback.

Furthermore, it would appear that biofeedback based upon either CQ or Ratio parameters in solitude in the context of this study has a major effect on the visually presented parameter, but small effect with regard to objective assessment of the other parameter.

With regard to the CQ and Ratio parameters considered in combination, the progression of subject A may be regarded as appearing to lose direction along the previously proposed developmental continuum. This is in contrast to subject B, who maintained a relatively steady rate of advance. Both linear and rank correlations (figure 6.7) indicate stronger correlations for the subject taught using biofeedback, both for singing and speech tasks. Two previous observations also support the proposal that biofeedback has resulted in relatively more sustained vocal development; the greater clarity of trend in mean speaking voice F0 for subject B (figure 6.3), and the far greater level of consistency between speaking and singing voice mean SPL (figure 6.4).

However, all of the conclusions presented in this section are not proven. Other factors that are in themselves unrelated to the use of visual feedback, such as natural variability and differences between the two subjects, may be partly or wholly responsible for the observed trends.

6.6 Conclusions

This study has identified clear developmental trends in the level of laryngeal closed quotient and the level of spectral amplitude in the singer's formant region relative to all spectral output. Measures of F0 and SPL have also exhibited trends as a function of the development process. There are indications that the influence of real-time visually presented biofeedback of laryngeal closed quotient and a measure of the 'singer's formant' phenomenon has supported a higher level of consistently sustained increases in parameters relating to vocal performance than an equivalent

tuition context in which biofeedback was not employed. However, the observed trends in data may be partly or wholly due to factors which are unrelated to the use of visual feedback.

Chapter 7

Conclusions

7.1 An overview of research with regard to the original hypothesis

In chapter 1, a hypothesis was proposed which stated:

Computer systems which

- 1. objectively measure a large set of vocal attributes,*
- 2. appropriately translate these attributes to a smaller set of easily interpreted measures relating to vocal efficiency, and*
- 3. dynamically display this information in an appropriate graphical form during the process of vocal tuition*

may be able to provide an useful objective measure of progress for the advancement of vocal development.

In order to satisfy this hypothesis, the following steps were proposed:

- 1. The analyses of developing voice users in order to identify parameters indicative of voice development.*
- 2. The development of a software tool for vocal tuition able to embody these parameters in the process of analysis and subsequent visual display which can operate in real-time.*
- 3. An assessment of the usefulness of the visual feedback system employed in real-time tuition.*

The work that has taken place towards the satisfaction of these goals is considered below:

1. The identification of parameters indicative of voice development

An analysis of the vocal mechanism was presented in chapter 2. This provided the base for an identification of several voice source and acoustic output parameters in chapter 3 which were shown to relate to voice performance. In chapter 4 these parameters were then analysed for their suitability as a measure of vocal *development*. This analysis provided indications to support the proposal that the level of closed quotient of the vocal folds and the level of spectral amplitude in the singer's formant frequency region relative to the level of spectral amplitude in the remainder of the vocal output spectrum were both reflective of vocal development to an extent.

2. The development of an appropriate system

A highly configurable software project was then developed for analysis and visual feedback of voice signals able to operate in real-time as a tool for a voice tutor. As presented in chapter 5, the software is able to assess a number of voice parameters in solitude or in algorithmic combination, and to present the value of these data in an appropriate form of visual display during the phonation process.

3. An assessment of the system using the identified parameters

This software tool was then used to present the two parameters identified as indicative of vocal development as part of a formal assessment of the benefits of their assessment and presentation in real-time. This was presented in chapter 6. The study demonstrated the useability of ALBERT in a training context, where its application in supporting a programme of voice tuition was well received by both instructors and students. There are some indications that use of the system could support a higher level of consistency in improvement of vocal performance than an equivalent tuition context in which the system is not employed.

There are some indications that the original hypothesis has been satisfied, although it has not been proven.

7.2 Directions for future research

There are several areas in which this research may be extended. These are discussed with regard to the three goals:

1. The identification of parameters indicative of voice development

The performance of the voice parameters discussed in chapters 2 and 3 with vocal tuition may be clarified by further studies. These would benefit from assessments over a wide range of contexts of voice usage, over longer periods of time, and with a greater number of subjects. As a result of such research, there may be a clearer appreciation of parameters whose role is largely known, such as the CQ and Ratio parameters, in addition to others whose role is less clear, such as the rate of opening and closing parameters considered in voice studies 1 and 2. The nature of the developing female voice in particular warrants further research.

2. The development of an appropriate system

The ALBERT system was designed to be highly flexible. Changes to the software are required only if a currently unsupported parameter is identified as useful for monitoring in a tuition context.¹

3. An assessment of the system using the identified parameters

The usefulness of any changes to ALBERT could then be assessed. There are many different tuition contexts including the field of linguistics and the treatment of dysfunctional voice subjects in which the benefits of biofeedback are largely untested. Formal tests with large number of subjects would provide indications of the usefulness of the ALBERT system in these contexts.

¹If alterations are required, the use of the world-wide standard programming language called 'C' will considerably facilitate any required alterations. The graphics library which ALBERT employs in the visual display of information (called 'GL') is also rapidly becoming adopted as a world-wide standard for graphics programming. Furthermore, the cost of hardware on which the ALBERT system will run without alteration has dropped over the past 3 years from £7000 to £3000, with an associated increase in speed of a factor of 3. This trend looks set to continue, bringing affordable tools for voice tuition within reach of the average voice tutor and user.

Appendix A

Study 1 - programme of work

Over a period of six weeks subjects received vocal tuition involving the following techniques.

1. Physical balance of posture and alignment, relaxation (including support from Alexander Technique specialist)

Free-up physical constriction and nervous tension.

Locate areas of tension and contrast stiff limbs with relaxed areas.

Increase ability to achieve stillness, calm control, hold attention.

Check balance and carriage of head, neck, shoulders, and support for the larynx within neck, and any other areas needing release.

Check flexibility and ease of vocal tract jaw, lips, tongue, soft palate and throat.

Increase ability and ease in deepening breath capacity and control.

Convey effective visual message to listeners, readiness to interact.

2. Breathing

Extend effective breath capacity to power the voice.

Centralise breathing, involve lower rib expansion and diaphragm and abdominal muscles to support easy breath and voice flow.

Select adequate breath supply to support vibration of vocal folds.

Employ greater pressure of breath to achieve greater volume of voice.

3. Resonance

Balance modification of the voice in throat, mouth and nose.

Maximise use of open jaw, mouth shaping and resonance for carrying power (projection).

Balance mouth resonance emphasising front for greater flexibility of inflexion.

Make sure voice "rooted", supported by the body.

4. Pitch

Each subject tests to select optimum pitch within their unique voice range.

Increase range and flexibility.

5. Muscular energy of organs of speech (lips tongue, etc) in forming words

Check precision of shape and closure in forming vowels and consonants (for clarity only, not to adopt any specific accent).

Increase awareness and sensitivity of sound qualities in words that enhance meaning, visual or sound 'image'.

6. Reading text and speaking

Adapt to the 'intention' or mood in spoken text or in speaking.

Establish appropriate pauses, rhythm, phrasing and pace.

Assess the role, status or tone the speaker transmits.

Size up need to adapt to the listening group.

Estimate effect of voice with reverberation from windows, walls, or absorbed by carpets, ceiling height, etc.

Appendix B

Study 1 - assessment procedure

1. Opening lines from the short story 'The Voyage' by Katherine Mansfield.

"It was a beautiful night, mild, starry, only when they got out of the cab and started to walk down the Old Wharf that jutted out into the harbour, a faint wind blowing off the water ruffled under Fenella's hat, and she put up her hand to keep it on. It was dark on the Old Wharf, very dark: the wool sheds, the cattle trucks, the cranes standing up so high, the little squat railway engine, the huge coils of rope, all seemed carved out of solid darkness. Here and there on a rounded wood-pile that was like the stalk of a huge black mushroom, there hung a lantern, but it seemed afraid to unfurl its timid, quivering light in all that blackness; it burned softly, as if for itself..."

2. Excerpt from 'Androcles and the Lion' by Bernard Shaw.

Subjects were instructed not to read the parenthesised actions, which were left in order to keep the context clear. Subjects were given the following context prior to reading : The centurion addresses the Christians with whom he, and the other soldiers, have been fraternizing on the journey to Rome.

"Halt! Orders from the Captain. *[They halt and wait]*."

Now then you Christians, none of your larks. The Captain's coming. Mind you behave yourselves. No singing. Look respectful. Look serious, if you're capable of it. See that big building over there? That's the Coliseum. That's where you'll be thrown to the lions or sent to fight the gladiators presently.

Think of that; and it'll help you to behave properly before the captain. [*The Captain arrives*].

Attention! Salute!..."

3. Consonant-vowel-consonant word sequences.

BOOD BEAD BAD BUD BED BIRD BARD BOARD
MOON MEAN MAN MUN MEN MURN MARN MORN

Appendix C

Study 2 - assessment procedure

1. Please say your age and experience
2. Please say your name and today's date
3. Please read the following passage:

The Story of Arthur the Rat

There once was a young rat named Arthur, who would never take the trouble to make up his mind. Whenever his friends asked him if he would like to go out with them he would only answer, "I don't know." He wouldn't say "Yes" and he wouldn't say "No" either. He could never learn to make a choice.

His aunt Helen said to him, "No-one will ever care for you if you carry on like this. You have no more mind than a blade of grass." Arthur looked wise, but said nothing.

One rainy day the rats heard a great noise in the loft where they lived. The pine rafters were all rotten, and at last one of the joists had given way and fallen to the ground. The walls shook and the rats' hair stood on end with fear and horror. "This won't do," said the old rat who was chief, "I'll send out scouts to search for a new home."

Three hours later the seven scouts came back and said, "We've found a stone house which is just what we wanted. There's room and good food for us all. There's a kindly horse named Nelly, a cow, a calf and a garden with an Elm tree." Just then the old rat caught sight of young Arthur. "Are you coming

with us?" he asked. "I don't know," Arthur sighed, "the roof may not come down just yet." "Well," said the old rat angrily, "we can't wait all day for you to make up your mind. Right about face! March!" And they went off.

Arthur stood and watched the other rats hurry away. The idea of an immediate decision was too much for him. "I'll go back to my hole for a bit," he said to himself, "just to make up my mind."

That night there was a great crash that shook the earth, and down came the whole roof. Next day some men rode up and looked at the ruins. One of them moved a board, and under it they saw a young rat lying on his side, quite dead, half in, and half out of his hole.

4. Please think of a passage you can recite for at least 30 seconds. Please recite it -
 - a) in your normal speaking voice
 - b) in a quiet, withdrawn voice (minimal projection)
 - c) as if you are on a stage in front of an audience (maximum projection)
5. Please sing the following verse.

Baa-Baa Black Sheep

Baa-Baa Black Sheep have you any wool?

Yes sir, Yes sir, Three bags full.

One for the master, and one for the dame,

And one for the little boy who lives down the lane.

Appendix D

Study 3 - assessment procedure

1. Please say your age and experience
2. Please say your name and today's date
3. Please read the following passage:

The Story of Arthur the Rat

There once was a young rat named Arthur, who would never take the trouble to make up his mind. Whenever his friends asked him if he would like to go out with them he would only answer, "I don't know." He wouldn't say "Yes" and he wouldn't say "No" either. He could never learn to make a choice.

His aunt Helen said to him, "No-one will ever care for you if you carry on like this. You have no more mind than a blade of grass." Arthur looked wise, but said nothing.

One rainy day the rats heard a great noise in the loft where they lived. The pine rafters were all rotten, and at last one of the joists had given way and fallen to the ground. The walls shook and the rats' hair stood on end with fear and horror. "This won't do," said the old rat who was chief, "I'll send out scouts to search for a new home."

Three hours later the seven scouts came back and said, "We've found a stone house which is just what we wanted. There's room and good food for us all. There's a kindly horse named Nelly, a cow, a calf and a garden with an Elm tree." Just then the old rat caught sight of young Arthur. "Are you coming

with us?" he asked. "I don't know," Arthur sighed, "the roof may not come down just yet." "Well," said the old rat angrily, "we can't wait all day for you to make up your mind. Right about face! March!" And they went off.

Arthur stood and watched the other rats hurry away. The idea of an immediate decision was too much for him. "I'll go back to my hole for a bit," he said to himself, "just to make up my mind."

That night there was a great crash that shook the earth, and down came the whole roof. Next day some men rode up and looked at the ruins. One of them moved a board, and under it they saw a young rat lying on his side, quite dead, half in, and half out of his hole.

4. Please sing a two octave scale up and down. You may wish to take a breath between each octave.

Appendix E

ALBERT User Guide

ALBERT

Acoustic and Laryngeal Biofeedback

Enhancement in Real-Time

—

USER GUIDE

—

David Rossiter

**Electronics Department, University of York,
Heslington, York, YO1 5DD, ENGLAND**

December 1994

Contents

1	Introduction	208
1.1	The ALBERT program	208
1.2	Reading this document on the computer	208
1.3	Terminology used in this manual	210
1.4	Organisation of this manual	210
1.5	Hardware required	211
1.5.1	Introduction	211
1.5.2	Voice source analysis	211
1.5.3	Acoustic analysis	212
1.5.4	Both combined	212
1.6	Setting up the environment	212
1.6.1	Introduction	212
1.6.2	The ALBERT_SFS variable	212
1.6.3	The ALBERT_INFO variable	213
1.6.4	Other programs	213
1.7	Summary	214
2	How to get going	215
2.1	How to run ALBERT	215
2.2	Recording and playback	215
2.3	The voice source signal	216
2.3.1	Introduction	216
2.3.2	Correct use of the electrolaryngograph	216
2.3.3	Ensuring ALBERT is analysing correctly	216

2.3.4	Analysis modes	218
2.3.5	Period display options	218
2.3.6	Animation	218
2.4	The acoustic signal	219
2.5	Summary	219
3	Parameter and visualization control	221
3.1	Introduction	221
3.2	Voice parameters	221
3.2.1	Extending the list of voice parameters	222
3.3	Parameter visualization	222
4	Functional units	226
4.1	Introduction	226
4.2	Header form	226
4.3	Audio form	227
4.4	Acoustic form	227
4.5	Acoustic and source form	228
4.6	File record form	228
4.7	Memory record form	229
4.8	Voice source form	229
4.9	Misc buttons form	231
4.10	Visualization form	231
4.11	Threshold form	234
4.12	Refractory time form	234
4.13	Ratio form	234
4.14	Params form	235
4.15	Period form	237
4.16	Save and load form	239
4.17	Forms ON OFF control	239
4.18	Forms size control	240

5	Command line control	241
5.1	Command line options	241
5.2	Preparing aliases	242
6	Functions keys	243

List of Figures

1.1	An overview of the ALBERT system process	209
1.2	Use of the electrolaryngograph	211
2.1	One period of the electrolaryngograph waveform with normalised amplitude and frequency	216
2.2	Voice source control form	217
2.3	Voice period control form	219
2.4	Acoustic analysis control form	220
3.1	ALBERT parameter control window	223
3.2	ALBERT visualization control window	224

Chapter 1

Introduction

1.1 The ALBERT program

ALBERT was created as part of a SERC sponsored project entitled *Real-time visual displays for professional voice development* between 1992 and 1994. It was designed to satisfy the need for a comprehensive analysis system for voice source and acoustic output signals of the human voice able to operate in real-time. An illustration of the way in which ALBERT can be employed for real-time visual feedback is shown in figure 1.1. In this mode of operation, ALBERT is used for real-time visual feedback.

The software runs on a Silicon Graphics machine equipped with audio hardware (such as an Indigo or Indy computer). It was written entirely using the 'C' programming language. Graphics were controlled using the 'GL' graphics library. The FORMS graphical user interface (GUI) system developed by Mark Overmars was used in the creation and handling of the graphical user interface. David Rossiter, the Research Associate employed for the project, wrote the software and this manual.

1.2 Reading this document on the computer

This User Guide is available both as a postscript document and as an interactive hypertext document at the following [www](#) address.¹

¹This document was written using the 'latex' formatting system. A version suitable for viewing on the world wide web ([www](#)) was obtained by converting the document using the 'latex2html' program created by Nik Dracos.

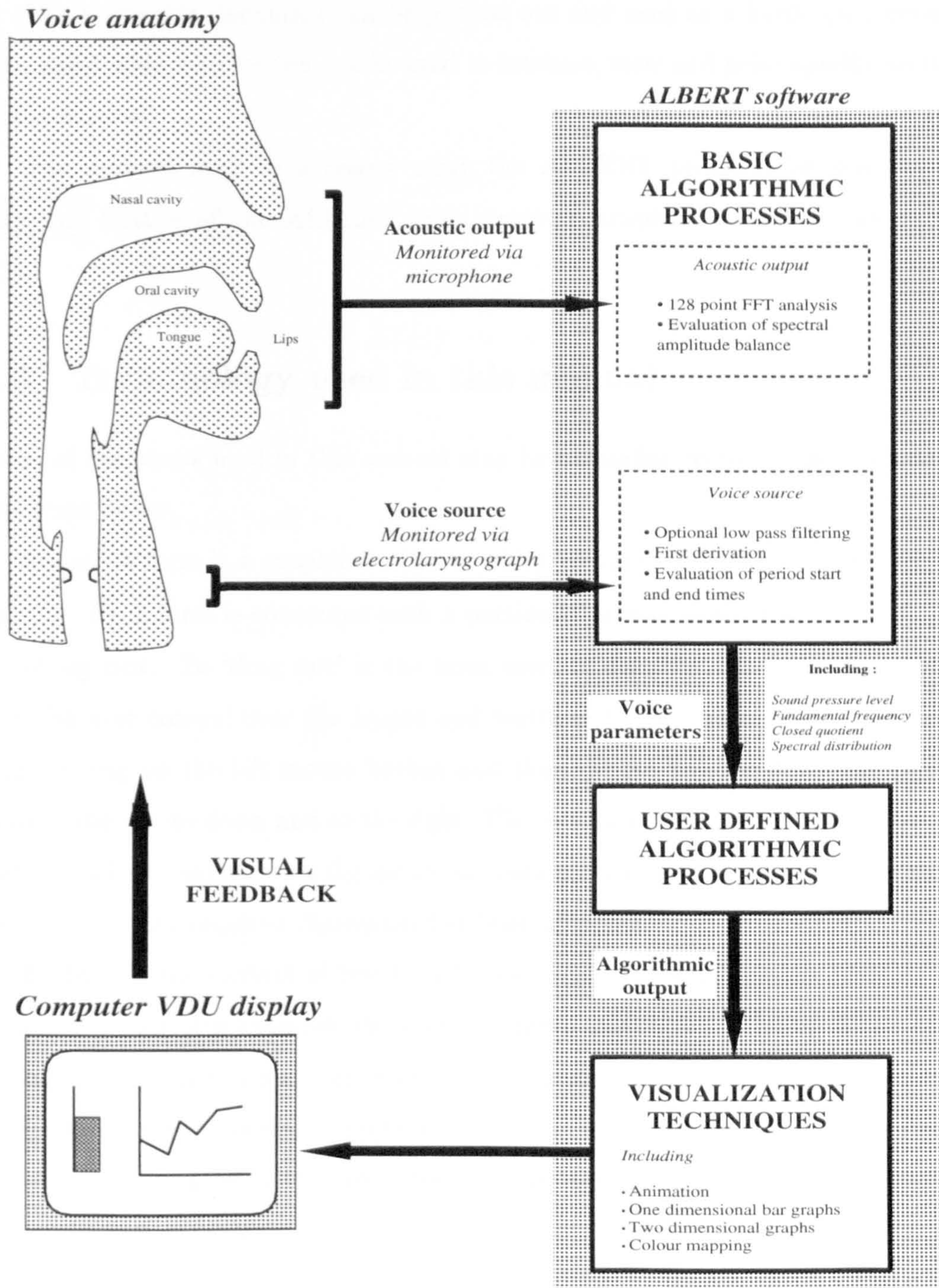


Figure 1.1: An overview of the ALBERT system process

http://www.york.ac.uk/~elec10/albert_manual.html

The postscript document can be printed out and used as a hardcopy reference. The world wide web version can be used to traverse, view and print specific sections of the manual.

The manual may be accessed using the ALBERT program by pressing the button of the ALBERT main program graphical interface (see section 4.2).

1.3 Terminology used in this manual

Some of the terms used in this manual may be unfamiliar to the reader. These are described below.

Form. A form is a graphical window which contains controls that the user can operate. Each form is concerned with a particular area of operation.

Drag out. To 'drag out' is the term used to describe the technique used to give the user control over the height and width of the window. This involves the user clicking on the left mouse button and then, whilst holding down the button, moving the mouse down and to the right. The outline of a window will be shown. It will expand and shrink with the mouse movement. The user may release the mouse button when the required dimension has been achieved.

Fields. In the context of this User Guide, *fields* represent certain qualities of a voice parameter. For example, the legal minimum and legal maximum values of the 'Frequency' parameter are both fields of the parameter.

In the course of the text, a rectangular outline is used to indicate the description of a button presented in the user interface. For example, and buttons appear in several places.

1.4 Organisation of this manual

In section 1, an overview of the requirements and aims of the ALBERT software is provided. In section 2 the basic controls required for the user to start using ALBERT for voice source and acoustic analysis are discussed. In section 3, control

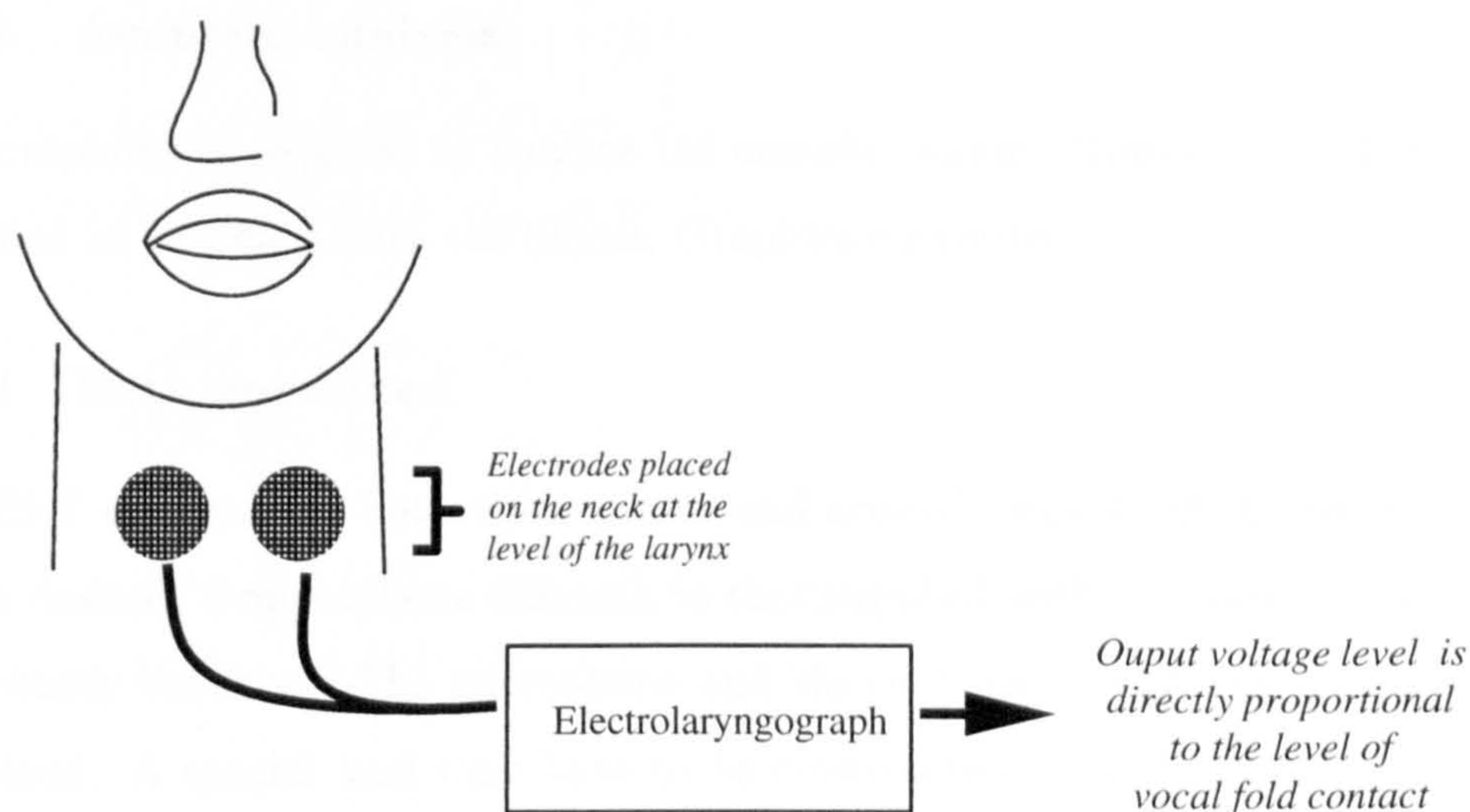


Figure 1.2: Use of the electrolaryngograph

over voice parameter fields and visualization is described. In section 4, a complete description of the use of all ALBERT forms is provided. Finally, in section 5 and 6 command line control and Function key operations are described.

1.5 Hardware required

1.5.1 Introduction

The ALBERT system can be used to analyse the voice source, or the acoustic output, or both combined. There are certain hardware requirements (in addition to the computer itself) depending on the desired function.

1.5.2 Voice source analysis

An external piece of hardware is required in order to use the system for voice source analysis. This is called an *electrolaryngograph*. In order to use it, two electrodes need to be placed on the surface of the neck at the level of the thyroid cartilage. A high frequency current is applied. When correctly used, the level of impedance to this current is proportional to the area of vocal fold contact. Use of the electrolaryngograph is illustrated in figure 1.2.

1.5.3 Acoustic analysis

A microphone is required to analyse the acoustic signal. However, one is normally provided as standard with the Silicon Graphics computer.

1.5.4 Both combined

ALBERT can process both voice source and acoustic signals at the same time. If this is desired, a microphone different to that supplied with the computer hardware is probably required. The microphone and electrolaryngograph signals need to be combined. A special lead may have to be constructed for this. Alternatively, one method is to pass microphone and electrolaryngograph outputs into a DAT machine, and then pass the output from the DAT machine into the Silicon Graphics computer. This has the benefit that any input to the computer can be recorded for future reference.

1.6 Setting up the environment

1.6.1 Introduction

It is advisable to set up the software environment in which ALBERT operates before running the program. Some ALBERT functionality relies on setting these environment variables, as discussed below.

1.6.2 The ALBERT_SFS variable

The audio and/or voice source signals recorded to file within ALBERT can be interactively viewed and printed. To achieve this, ALBERT uses the speech filing system (SFS) suite of programs.² A temporary file is created during the process of transferring the recorded signals from the file in which ALBERT originally recorded the signals into the SFS domain. ALBERT relies on one environment variable to point to a directory where this temporary file is to be placed. This variable is called `$ALBERT_SFS`.

²This requires that the SFS programs are properly set up in the normal way - see section 1.6.4.

It is usually appropriate to point this variable to the temporary file space area of the computer, such as `/usr/tmp`. For this purpose, it is recommended that you have the following lines in your `.cshrc` file in your home directory.

```
set ALBERT_SFS = /usr/tmp
setenv ALBERT_SFS $ALBERT_SFS
```

Before using this utility, you must be sure that the directory (in this case, `/usr/tmp`) exists and that the directory can be read from and written to by all users.

If required, the relevant directory can be created by typing

```
mkdir /usr/tmp
```

and can be made readable and writable by everyone by entering

```
chmod a+rw /usr/tmp
```

1.6.3 The ALBERT_INFO variable

ALBERT relies on another environment variable to tell it how to present the manual. This variable is called `$ALBERT_INFO`. It is recommended that you have the following lines in your `.cshrc` file in your home directory.

```
set ALBERT_INFO = 'mosaic
http://www.york.ac.uk/~elec10/albert_manual.html'
setenv ALBERT_INFO $ALBERT_INFO
```

With this configuration, upon clicking the `manual` button ALBERT will call up the world wide web (www) viewer program called 'mosaic' and point it to the interactive hypertext version of the manual.

1.6.4 Other programs

In order to achieve its power, ALBERT uses several other programs for certain functions. Separate documentation exists for each of them, to which the reader is referred for further information. They are;

1. **icol.** This is an editor for interactively creating a colour scheme against which to map a range of data values.
2. **bob.** This is a 3D interactive viewer program which is able to use icol to produce visualization displays of any section of a range of data.
3. **sfs programs.** These are used when the user requests that the recorded audio and/or electrolaryngograph signals be viewed. This is done by pressing the , , or buttons (see section 1.6.2).

You must ensure that these programs are properly set up in the normal way, with the user path set so that the programs can be executed when ALBERT needs to do so. Please refer to the manuals for information on how to do this.

1.7 Summary

This chapter has introduced the ALBERT system, discussed the hardware requirements, and provided information on how to access the User Guide and arrange the environment so that ALBERT can operate properly. In the next chapter the user is introduced to the ways in which ALBERT is used to analyse voice source and acoustic signals.

Chapter 2

How to get going

2.1 How to run ALBERT

To run the program, just type

```
albert
```

from the normal shell window.

Some command line options exist. These are discussed later, in section 5.1.

2.2 Recording and playback

Voice source and acoustic signals can be recorded within ALBERT. Recordings can be made to memory or to file. Two forms are provided for control over each of these. They are essentially identical. To begin recording a sequence, just press the `record` button on the form. Press it again to finish recording. Press `play` to play it back. During playback, the signals will be analysed by ALBERT if the algorithms have been turned on. The user can examine the soundfile by pressing the `view` button, which transfers the audio information into SFS format and calls the SFS viewer program.

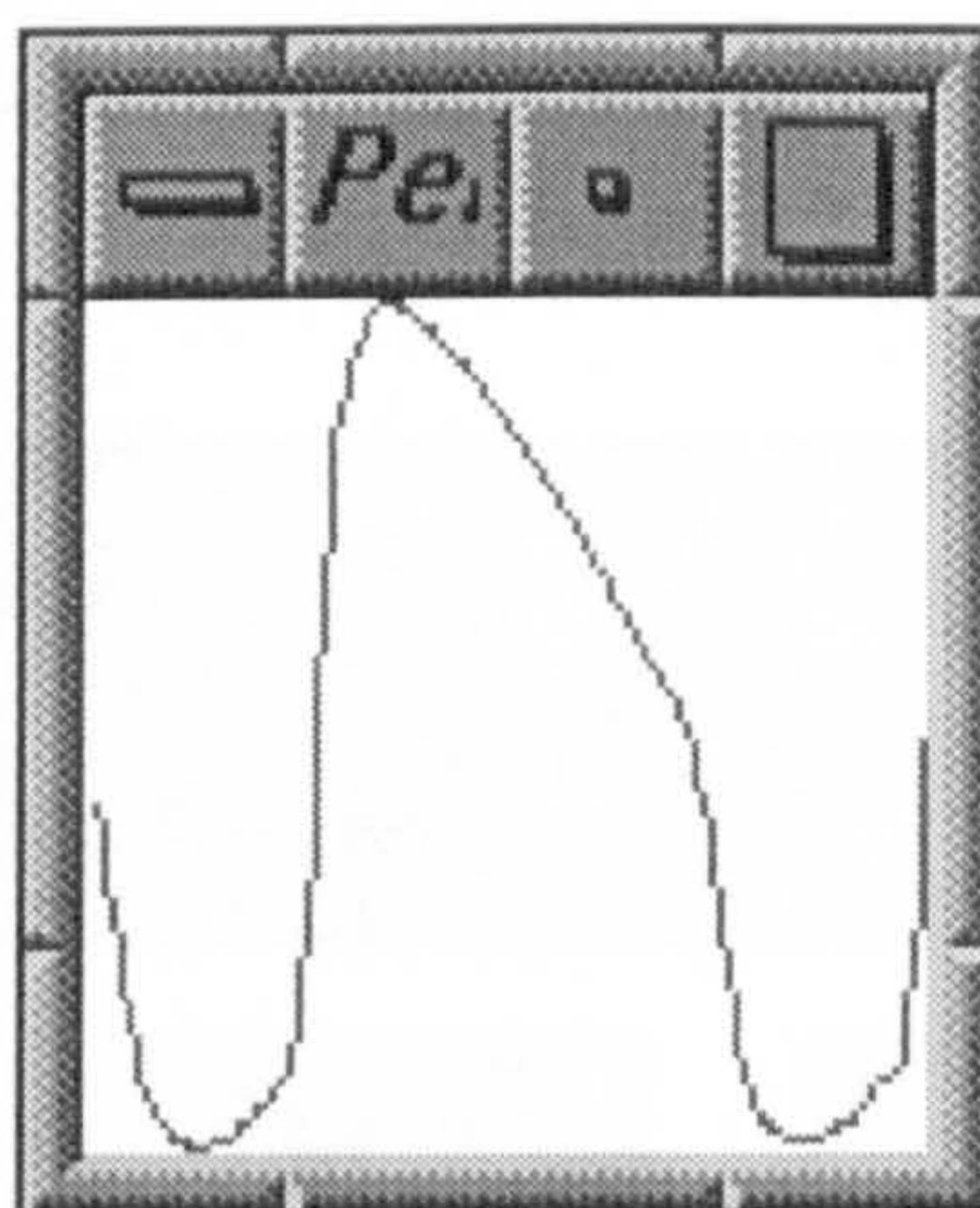


Figure 2.1: One period of the electrolyngograph waveform with normalised amplitude and frequency

2.3 The voice source signal

2.3.1 Introduction

There are some basic steps before ALBERT can be used to analyse an electrolyngograph signal.

2.3.2 Correct use of the electrolyngograph

Ensure that the electrolyngograph is used in a manner appropriate for getting the best possible signal. This requires the careful positioning of both electrodes on either side of the neck at the level of the larynx. The electrolyngograph signal should be viewed on an oscilloscope. Once a basic signal has been obtained, an improved signal can often be obtained by trying a few different electrode positions whilst phonating. A good clear signal on an oscilloscope display should look similar to that shown in the window displayed in figure 2.1.

2.3.3 Ensuring ALBERT is analysing correctly

The user must ensure that ALBERT is set to interpret the information from the electrolyngograph in the best way. This requires ensuring that the program treats this signal appropriately. The main controls for this are shown in figure 2.2.

The primary selections that need to be made are;

Figure 2.2: Voice source control form

1. whether to Invert the signal or not. If the signal on the CRO looks like an upside-down version of figure 2.1, press the invert ON button.
2. the channel (*left* or *Right*) on which the signal is incoming.

There is an option to let you see what signal the program is receiving. Press the button *Period* on the voice source form (section 4.8). A window will come up. An example is shown in figure 2.1. In this window the computer will display what it regards as one period of the electrolaryngograph signal. It scales the signal to fit the window regardless of the frequency or amplitude of the signal. This display is very useful in ensuring that the program is treating the incoming electrolaryngograph signal correctly.

The voice source assessment algorithm needs to be ‘tuned’ to ensure correct operation. An appropriate method for this is;

1. Get the ‘apanel’ program display on the screen by pressing the *Audio control* button at the top left of the main screen display.
2. Say ‘ah’ aloud for several seconds. During this time perform actions (3) and (4) below.

3. Alter the input level meters on the apanel display so the level is close to the maximum height possible *without* the red area lighting up.
4. Press the threshold button. ALBERT will now calibrate itself.

A slider control is also provided so that the user may directly control the performance of the electrolaryngograph signal analysis algorithm. This is present on the Threshold form (section 4.11).

2.3.4 Analysis modes

There are two modes for deriving information from the electrolaryngograph signal. The first examines the incoming signal at periodic intervals and uses only the first legitimate signal it encounters in each buffer of incoming signal. The second option requires ALBERT to extract every single legitimate period from the incoming signal. The first option is suitable for real-time use; the second is not, because it takes a lot more time to process all the periods in each buffer instead of just the one. To select the first option, press the attribute derivation button. To select the latter option, press the button.

2.3.5 Period display options

The user is able to control the display of the period signal in several ways. The control form for this is shown in figure 2.3. For example, the user can select how much of the previous period to display. If the user wants to see only the current period, he can would move the slider shown on the form to the far right. If the user wishes to see all of the previous period in addition to the current period, he/she would move the slider to the far left. An intermediate selection of 40% is good for positioning the main body of the electrolaryngograph period signal approximately in the centre of the window.

2.3.6 Animation

Using the same form (figure 2.3), the user has access to a simple animated display. This infers the action of the larynx from the last period display. The user should first obtain a good period signal, and then press the button. Animation

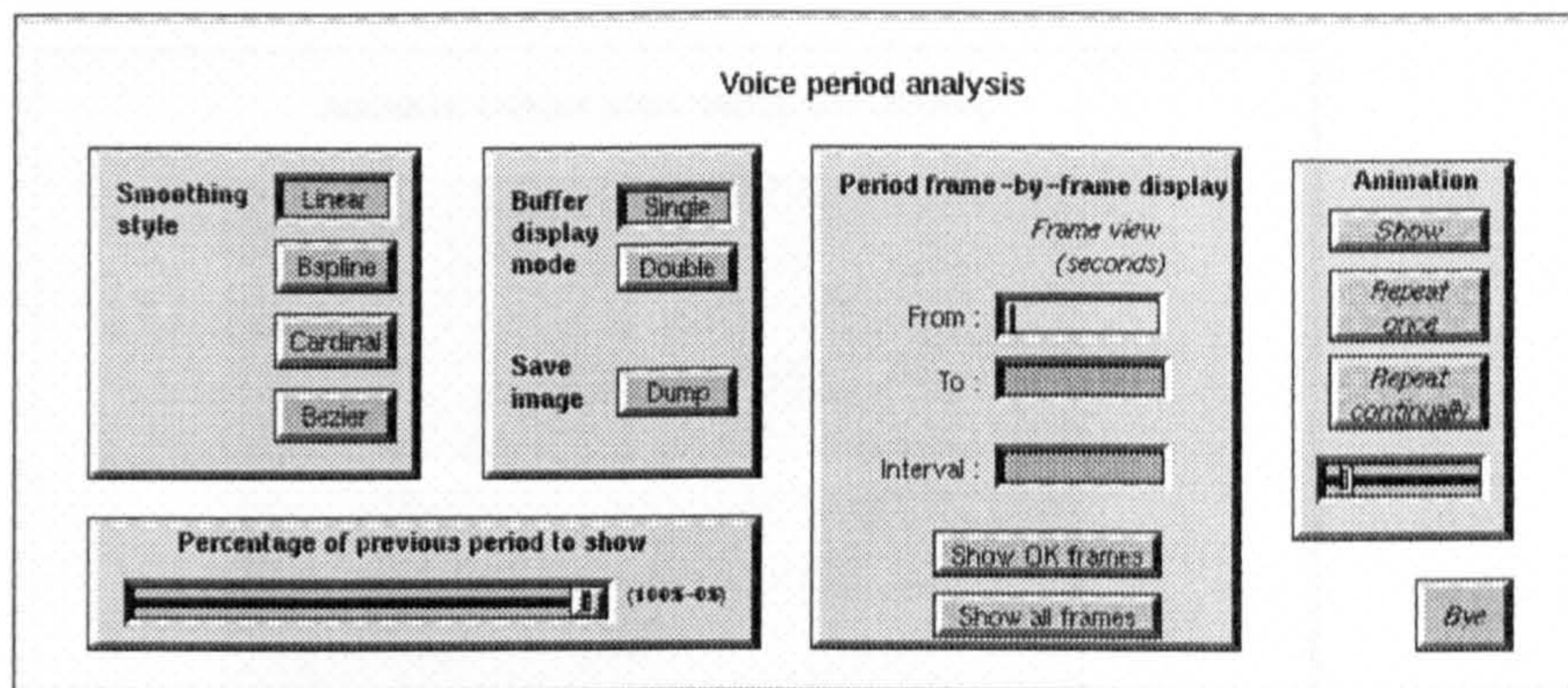


Figure 2.3: Voice period control form

will be shown in a scaleable window. Repeated action is available by pressing the comma key, or by selecting one of the appropriate buttons.

2.4 The acoustic signal

The control for handling the processing of the acoustic signal is shown in figure 2.4. In order to analyse the acoustic signal, the user must first select the relevant channel (left/Right). The selected channel will usually be the opposite to the electrolaryngograph signal channel. However, it is possible to analyse the spectrum of the electrolaryngograph signal instead of the acoustic signal simply by selecting the channel of the electrolaryngograph signal.

In order for the acoustic signal to be analysed, the algorithm used to process the signal must be turned on. This is achieved by pressing the FFT algorithm ON button. The user can then select whether to view the acoustic information as a plot of spectral amplitude across frequency, or as a spectrogram. To display the former, the user would select Amp/Freq Show. To display the latter, the user would select Freq/time Show.

2.5 Summary

This chapter has introduced the basic techniques required to use ALBERT for voice source and acoustic signal analysis. In the next chapter advanced techniques are

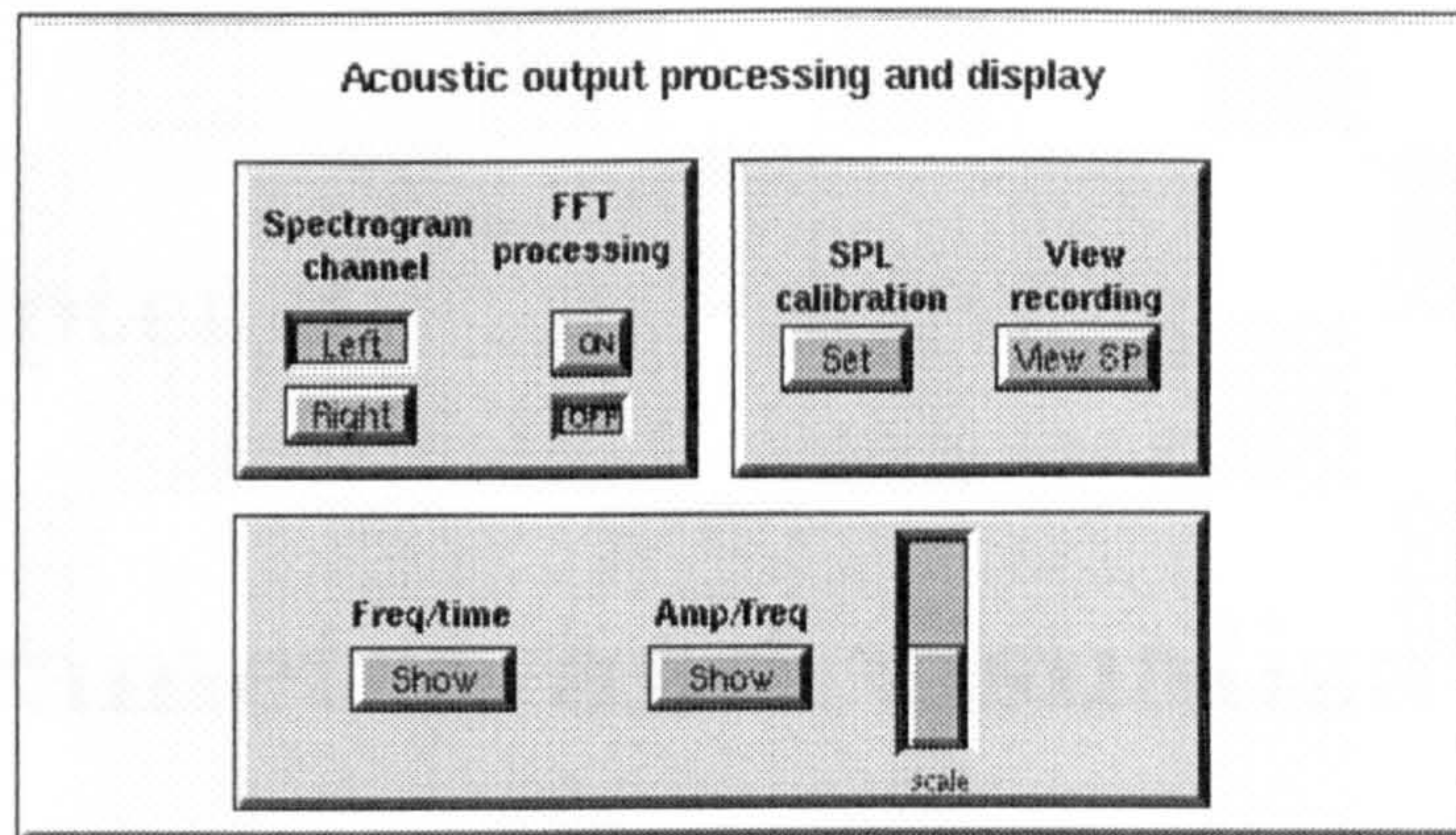


Figure 2.4: Acoustic analysis control form

presented for controlling and displaying the state of the voice parameters which ALBERT derives from the two signals.

Chapter 3

Parameter and visualization control

3.1 Introduction

From the voice source and acoustic signals several voice parameters are derived. This chapter describes how these parameters may be controlled and visualized.

3.2 Voice parameters

If both voice source and electrolaryngograph signals are being processed correctly, ALBERT updates a number of parameters which the user can analyse. These are:

1. the fundamental frequency of the voice (F0)
2. the closed quotient of the voice (CQ)
3. the rate of opening of the vocal folds (oa)
4. the rate of closing of the vocal folds (ca)
5. the spectral distribution of the voice output (Ratio)
6. the loudness of the voice (SPL)
7. the change in loudness from one time instant to the next (shimmer)
8. the period-to-period change in fundamental frequency (jitter)

9. time
10. the peak-to-peak amplitude of the electrolaryngograph period (lx amp)
11. the change in amplitude of the electrolaryngograph period from one time instant to the next (lx shimmer)

The user can display each of these parameters using visualization techniques described in the next section. Prior to this, however, the user may wish to examine or alter some of the properties of the parameters. The control for this is shown in figure 3.1.

The parameters are listed in the window shown in the top-left hand corner of the form. To see the fields of a particular parameter, the user needs to select the parameter from this list by clicking on it. The fields will then be shown in the small windows to the right and below the parameter list. The user can alter the fields by clicking on the field and altering the text.

3.2.1 Extending the list of voice parameters

A new parameter may be created by the user by pressing the *add new parameter*. A 'skeleton' parameter is then created by ALBERT with dummy fields.

The user may then alter these fields according to his/her requirements. The user is able to define a new parameter based on an algorithmic treatment of other parameters. For example, to provide a new parameter which gives an indication of the performance of CQ with regard to frequency, the user can request a new parameter button, provide a new name for the parameter (for example, CQ_freq), and enter the equation for the new parameter (in this case, CQ/Freq). Once the other fields have also been entered, the new parameter may be used for visualization or for reference by a further algorithm.

3.3 Parameter visualization

The state of any of the parameters supported by ALBERT can be displayed using the Visualization controls shown in figure 3.2. For example, to display the state of the fundamental frequency parameter as a bar graph, the user would typically

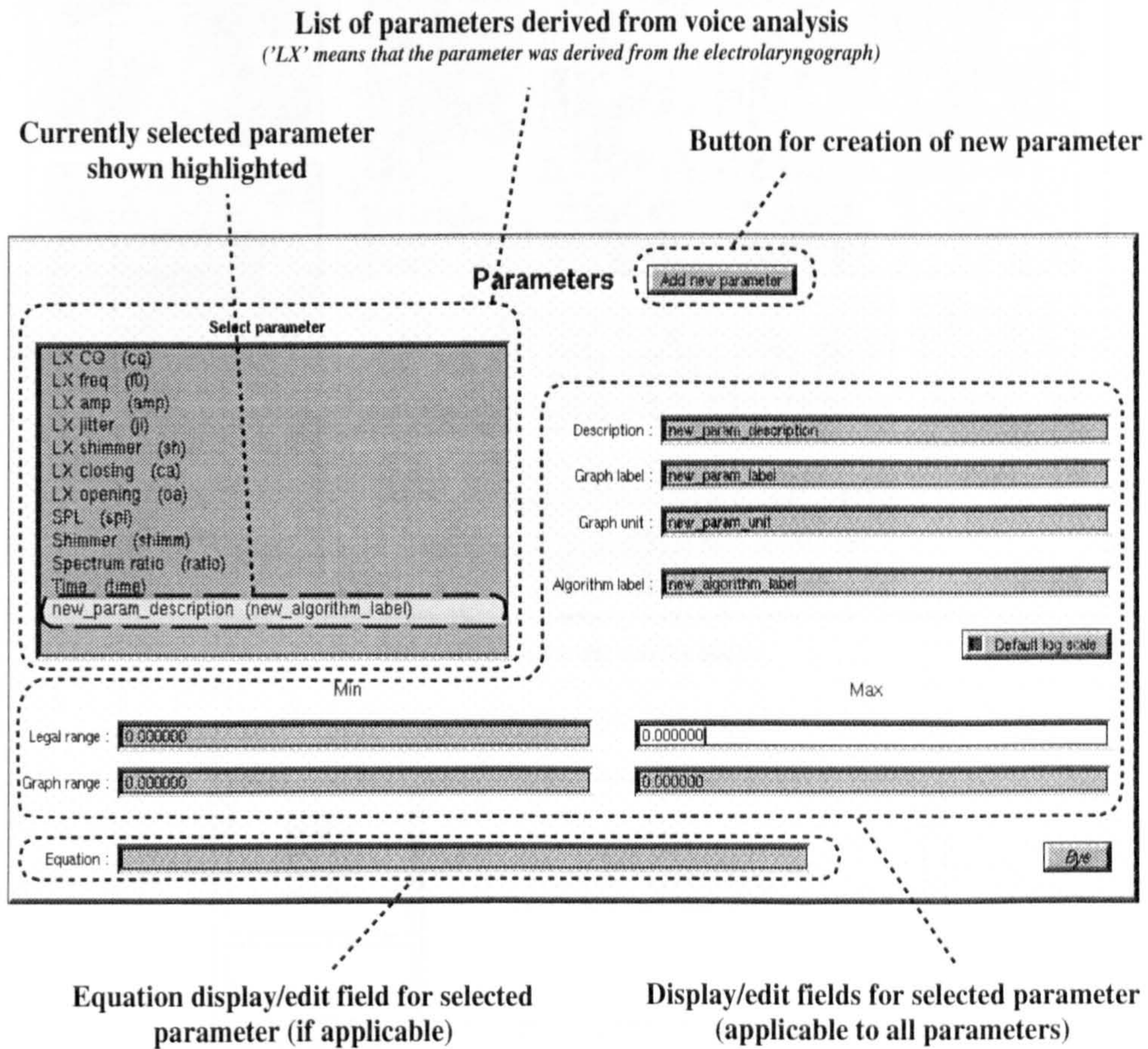
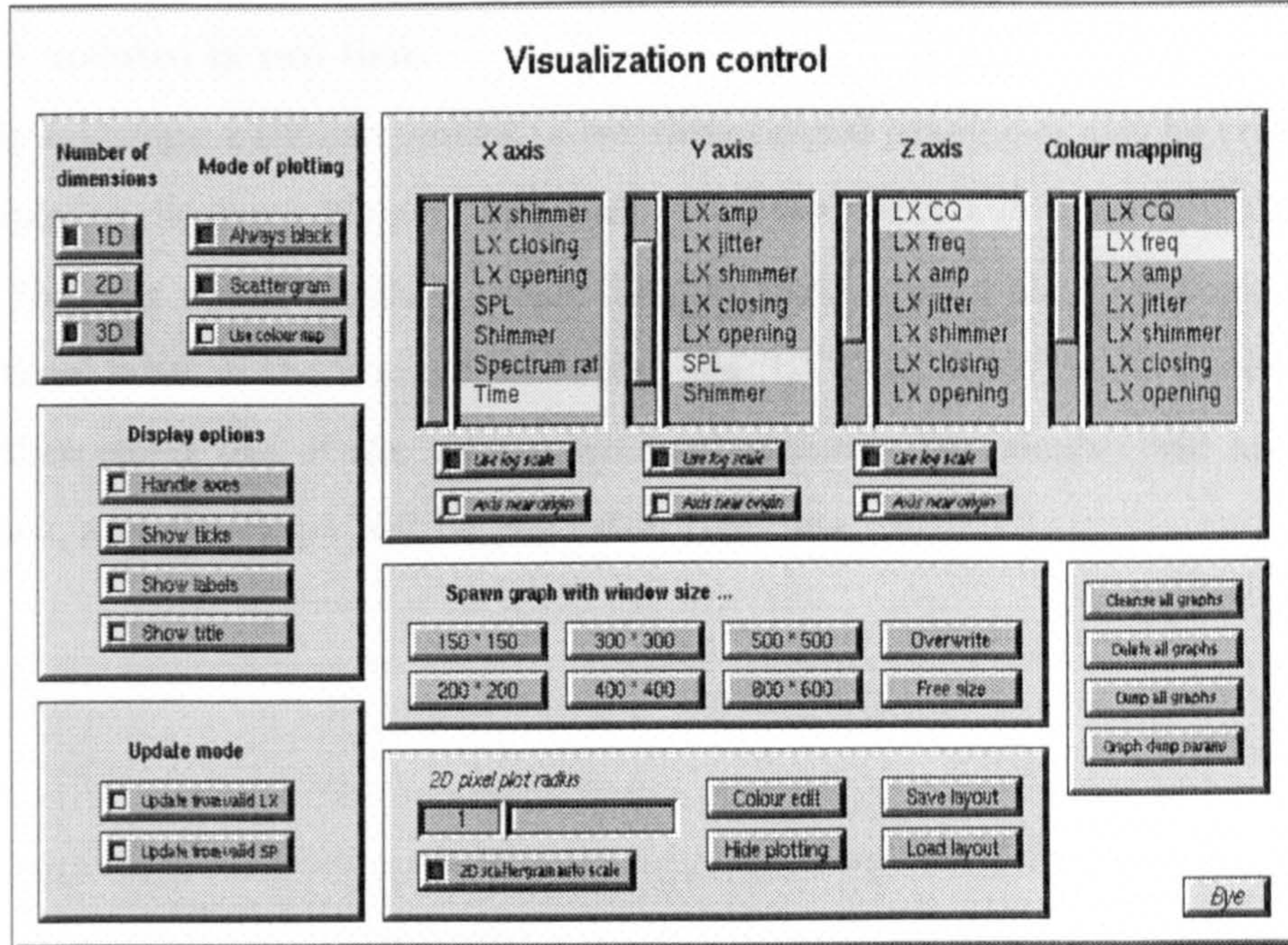
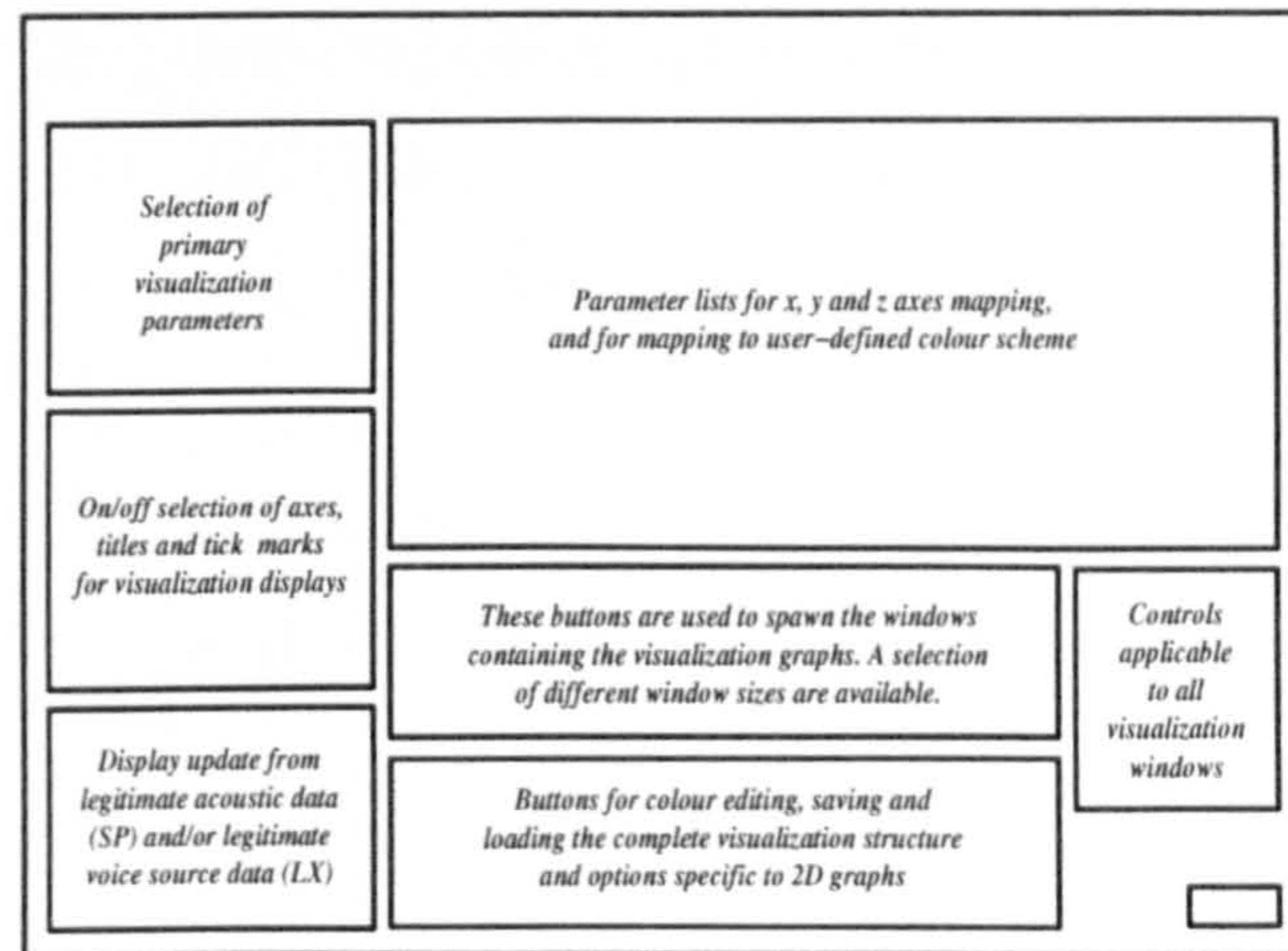


Figure 3.1: ALBERT parameter control window



(a) visualization control window display



(b) key to display

Figure 3.2: ALBERT visualization control window

select the 1D and Always black buttons, click on the 'LX Freq' label in the y axis column, and then select one of the 'Spawn window' options in the centre of the form. The display will appear in a window, and the state of the parameter will be shown updated in real-time.

By following a similar process, a two dimensional graph can also be created. For example, to display a 2D plot with time along the x axis and frequency along the y axis, the user would typically select the 2D and Always black buttons, click on the 'time' label in the x axis column and the 'LX Freq' label in the y axis column, and then select one of the 'Spawn window' options. The display will appear in a window, and the graph will be plotted in real-time.

Chapter 4

Functional units

4.1 Introduction

ALBERT has many functional components. For example, there are some functions concerned with only voice source analysis, and there are some functions concerned with acoustic analysis. The controls for the various ALBERT functions are grouped together and presented to the user on separate units, called 'forms'. In this section the functionality and control of each of the 16 forms is described in detail.

The order in which the forms are described in the following sections is the same as that in which they are listed on the Visibility control form, which is usually shown at the top right hand corner of the screen. If you are reading this section next to a, or on, a computer running ALBERT, you may wish to click on the form being described so that you can relate the descriptions with the actual form display.

4.2 Header form

This form provides a banner for the display of the name and version number of the software. It additionally provides the user with easy access to the ALBERT manual in electronic form.

4.3 Audio form

This form provides the user with an easy method to start the program used for audio input/output control. This is currently the Silicon Graphics 'apanel' program.

4.4 Acoustic form

This form contains the main controls for acoustic processing. It is shown in figure 2.4. It contains controls for :

1. left / right channel selection
2. turning on / off the fft algorithm, which must be enabled for acoustic analysis and SPL assessment
3. SPL calibration. For this, the fft process must be engaged. Phonation is required at a constant output level and constant distance from subject to microphone. During this period of stable phonation this button should be pressed. ALBERT will then prompt for the SPL level in dB's (as determined by an SPL meter at the position of the microphone). Assuming the subject maintains the same constant position relative to the microphone, ALBERT will subsequently supply a correct SPL level reading when the fft process is turned on. This is expressed as the 'spl' variable.
4. View recording. This button transfers the acoustic channel of a recorded file sequence (section 4.6) to the Speech Filing System (SFS) domain. The SFS editor program called 'es' will appear. The user may then zoom in and play or print part or all of the signal sequence.
5. Freq/time. Provides the user with a spectrogram display.
6. Amp/freq. Provides the user with a display of spectral amplitude across frequency.
7. Scale. The display of spectral amplitudes in the Amp/freq window can be scaled vertically using the slider provided. Use of this slider does not otherwise affect processing and display of the acoustic signal.

4.5 Acoustic and source form

This form provides the user with the ability to change the processing and display update rate. Three sliders are provided; one displays the number of updates per second, another displays the number of samples per buffer, and a third displays the length of time of each buffer. All three slider values are inter-related. A change in one of the three will result in automatic changes to the other two.

There is additionally a button to transfer the file recording of *both* voice source and acoustic signals into SFS form for interaction and display in the manner described previously (section 4.4).

4.6 File record form

This form provides the user with easy control over the recording and playback of acoustic and laryngeal signals. Using this form, both signals can be recorded to file. The name of the file is 'default.bin' by default. The user may select a new file name by clicking on the box.

To start the recording of the incoming signal, the user clicks on the button. The recording light will turn on. To finish recording, the user presses the button again. The recording light will turn off and the length of the file shown at the top of the form will be updated. This is expressed in seconds.

To play the file, the user clicks on the button. The play indicator light will be turned on and the recorded file will be played back. If the user wishes to stop playback before the end of the file has been reached, he/she needs only to click on the PLAY button a second time. The play indicator light will then be turned off and playback will cease. Alternatively, if the user does not interrupt playback, the play indicator light will turn off and play back will cease automatically when the end of the file is reached.

A button is provided. When this is not turned on, the recorded file is read and played via the internal speakers. If ALBERT has a heavy work load, there will be a delay between the playback of each buffer. Although this does not affect the performance of ALBERT whatsoever, it can be very disconcerting to the human ear to hear a disjointed stream of audio and electrolaryngograph signals. For

this reason, the silence mode was implemented. When engaged, all buffers of audio and electrolaryngograph data that are read from file are silently processed.

4.7 Memory record form

This form duplicates the function and display of the File record form discussed in the last section, but with recording and playback made to/from the computers memory rather than a file.

All the functions described in the last section are present, with the exception of the button. Clearly, as this form does not involve recording to file, this button is not required. Instead, it has been replaced by a button. This enables the user to transfer the contents of the memory recording to a file, whose name they may determine. If provision was not made for this function then any memory recordings would be lost when ALBERT ceased operation.

4.8 Voice source form

This form is the primary form for controlling the processing and display of voice source signals. It is shown in figure 2.2. The controls are individually discussed below.

1. LX channel. This enables the user to select the channel of the incoming electrolaryngograph signal; or .
2. LX processing. This enables the user to turn the entire voice source processing mechanism or . For example, if the user is only interested in processing the acoustic signal, then this button would sensibly be switched off.
3. Invert. This control permits the user to invert the incoming electrolaryngograph signal. For example, this option is useful because often there are differences between DAT machines made by different manufacturers which may mean that a signal recorded on one DAT machine will played back on another machine in an inverted form.

4. Attribute derivation. This controls the extent to which each buffer of incoming electrolaryngograph signal is processed and graphical displays based on analysis of the signal are updated. There are two options. The first is labelled **First in buffer**. If this option is selected, only the first electrolaryngograph period present in each buffer will be processed. The second option is labelled **All in buffer**. In this case, all periods are processed. The former option is typically used for real-time function; the latter for analysis when real-time performance does not matter, but a thorough analysis of the signal does.
5. **Assess**. When the user presses this button, the electrolaryngograph signal is analysed and the 'threshold' parameter of the algorithm used to assess the electrolaryngograph signal is assessed and set. This parameter may also be directly controlled by the use of the Threshold form discussed below (section 4.11).
6. **Show period**. When pressed, the outline of a window displaying one normalised period of the electrolaryngograph signal appears. The user may then use the mouse and 'drag out' the window to the required width and height, as he/she feels is appropriate. The signal is normalised so that regardless of signal amplitude and frequency, the period is scaled to fit within the window dimensions. To remove the window, press the **Show period** button again.
7. **View LX**. The voice source channel of the signals which have been recorded to file (using the form discussed previously in section 4.6) are transferred into Speech Filing System (SFS) domain and displayed using the interactive viewer program called 'es'. The user can zoom in to examine and playback sections of the signal. Press 'q' or click on **quit** to exit the viewer. See the 'es' manual page for further information on how to use the program.
8. Low pass filter control. This enables the user to engage a low pass filter. This can be useful in processing noisy electrolaryngograph signals. The cut-off frequency, which has a default of 5000 Hz, is shown in a small window. If required, the user can click on the window and alter this value.

4.9 Misc buttons form

This form contains buttons which were judged as not appropriate for inclusion on other functional units. These are;

1. . This calls the 'soundeditor' program supplied by Silicon Graphics. It is useful for editing sound. See the soundeditor manual page for information on how to use the program.
2. . This calls the 'soundfiler' program supplied by Silicon Graphics. It is useful for converting audio files between different formats.
3. . This can be used to exit ALBERT. The user is prompted with the message 'Are you sure?'. Clicking on causes the program to terminate. Clicking on ensures the program does not terminate, and that ALBERT continues to perform as it did before the button was pressed.

4.10 Visualization form

This is the main form used to display voice source and acoustic parameters. It is shown in figure 3.2. There are seven sections in which elements of the GUI which are functionally similar have been grouped. These are discussed in order of their appearance from left to right and top to bottom.

1. **Number of dimensions/ Mode of plotting.** This section enables the user to control the highest level attributes of the visualization. These are the selection of , , or graph configurations, and the mode of colour display. Options for the latter are; , , and .
2. **Parameter mapping.** This section provides the user with the means to map the voice parameters assessed by ALBERT to the axes of the display configuration. If the user has selected a 1D display, then only the 'Y axis' column is applicable. If the user has selected a 2D display, then only the 'X axis' and 'Y axis' columns are applicable. If the user has selected a 3D display, then the 'X axis', 'Y axis' and 'Z axis' columns are all applicable. The 'Colour

mapping' column is applicable for either 1D, 2D or 3D configurations if the user has selected the **Use colour map** option, as discussed above.

3. **Display options.** The user is able to select which portions of axes display he/she desires. The first option, **Handle axes** determines whether any axes or portions of axes are displayed at all. If this is flagged on, then the remaining three controls apply. These are; **Show ticks**, which controls the display of axes ticks, **Show labels**, which controls the display of axes labels, and **Show title**, which controls the display of axes titles.
4. **Spawn graph.** This section presents the user with a list of possible window sizes from which an appropriate selection may be made. The list of possible window sizes varies from very small to very large. In addition, the user may spawn a window of his/her own specific height and length by selecting the **Free size** window option. The outline of a window will then appear. The user can then click on the left mouse button and 'drag out' the outline of the desired window size. It is also possible to 'overlay' a new graph on top of a graph of the same format which is already present on the screen. This is achieved by clicking on the **Overwrite** button, and then selecting the graph which the current configuration is to be positioned over by clicking on the top of the graph.
5. **General graph control.** Four buttons are available for high-level graph control. These are;
 - (a) **Cleanse all graphs** All graphs are cleared of data and reset to original state. All statistics associated with the graphs (such as mean and standard deviation values) are also reset.
 - (b) **Delete all graphs** All graphs, and the windows containing the graphs, are cleared of data and removed.
 - (c) **Dump all graphs** All graphs, are dumped to file in the Iris 'rgb' file format. They are given the suffix '.rgb'. This includes any statistics windows associated with the graphs. Any statistics data displayed in these windows are also output in text format. These have the suffix

'*.txt*'. In each case, the first few characters of the file name are selected by the user (see (d) in this list). If the graph under consideration is a 1D graph, the filename is then extended by an underscore character and the label text of the y axis parameter. If the graph under consideration is a 2D graph, the filename is extended by an underscore character, the label text of the x axis parameter, another underscore character and the label text of the y axis parameter. If the graph under consideration is a 3D graph, the filename is extended by an underscore character, the label text of the x axis parameter, another underscore character, the label text of the y axis parameter, another underscore character and the label text of the z axis parameter.

- (d) **Graph dump parameters** If the user presses this button, he is prompted with a string entry field. This contains the first few letters of the filename used in the writing to file of graph images. Any sequence of appropriate characters may be entered. This enables the user to easily alter the filename of all graph images so that, for example, a previous set of graph images is not overwritten. If the user clicks on the **Use graph index in filename** option, then the filename of each graph image transferred to file (as described in the last list item) is further extended by an underscore character and a unique number. This is useful if, for example, two graphs are present on the screen with the same x and y axis, but with otherwise different configurations. If this option was not enabled, the transfer of the second graph image to file will overwrite the image of the first graph, because the graphs will have identical filenames.
6. **Update mode** The graph data and display may be optionally controlled to react only to an update of either voice source parameters, or acoustic parameters, or both. For example, it is often required that displays are updated only when voiced activity is detected. In this case it would be appropriate to select the **Update from valid LX** button, and to ensure the **Update from valid SP** button was deselected. The visual display of any parameter (even if it is based on the acoustic signal) will then only take place if voicing action is detected.

4.11 Threshold form

This form provides easy access to a major parameter used in the algorithmic processing of voice source signals. It is often used in conjunction with the refractory time form and the voice source form (sections 4.12 and 4.8 respectively).

This parameter represents the smallest signal level that the voice source processing algorithm will judge to be a rising edge of the electrolaryngograph signal. If this value is too low, then the algorithm may be too eager to interpret the incoming signal as a rising edge. If this value is too high, then the algorithm will never detect a rising edge. A sensible intermediate value is required. This parameter is automatically set by the use of the button on the voice source form (section 4.8). Alternatively, the user may wish to override the automatic setting. This can be achieved by simply clicking on the slider and selecting the desired level.

4.12 Refractory time form

This form provides easy access to a major parameter used in the algorithmic processing of voice source signals. It is often used in conjunction with the threshold form and the voice source form (sections 4.11 and 4.8 respectively).

The parameter represents the smallest interval within which the voice source analysis algorithm is not permitted to detect the rising edge of an electrolaryngograph period. This parameter can be used to strengthen the performance of the voice source processing algorithm. For example, if it is known that the vocal task being analysed will not require phonation higher than 500Hz, then this slider can be set at 2ms; if it is known that the vocal task will not require phonation higher than 250Hz, then the slider may be set at 4ms. This may be achieved by simply clicking on the slider and selecting the desired level.

4.13 Ratio form

This form enables the user to control the parameters used in evaluating the 'Ratio' parameter. This parameter is derived by assessing the total spectral power within one frequency range with regard to a larger frequency range of which the former is

a subset. Two sections are present in the form;

1. **Total frequency band.** Two sliders are presented with which the user can control the lower and upper frequency limits of the larger band. These default to 0Hz and 5000Hz respectively.
2. **SF frequency band.** Similarly, two sliders are presented with which the user can control the lower and upper frequency band within which the phenomenon of the singer's formant is known to occur. These default to 2000Hz and 4000Hz respectively.

4.14 Params form

This form enables the user to view and alter ALBERT voice parameter fields. It is shown in figure 3.1.

The form is divided into two sections. The first is a list of the descriptions and labels of all voice parameters.

The second is a list of the parameter fields for the currently selected parameter (shown highlighted in the list). To look at the fields of a particular parameter, the user needs only to click on the name of the relevant parameter in the list. To adjust any of the parameter fields, the user needs only to click on the appropriate field and alter the value as appropriate. The following parameter fields are displayed. All of them may be altered by the user.

1. *Description.* Verbose description of the parameter.
2. *Graph label.* The parameter label (optionally) displayed when the parameter is mapped to a graph axis. For example, the default label used for the 'frequency' parameter is 'Freq'.
3. *Graph unit.* The parameter unit (optionally) displayed when the parameter is mapped to a graph axis. For example, the default unit used for the 'frequency' parameter is 'Hz'.
4. *Algorithm label.* This is the label used to refer to the parameter in an algorithm reference (discussed below). For example, the default label used for the

'frequency' parameter is 'f0'.

5. *Legal range - minimum.* This is the minimum legal value of the parameter. For example, the default legal minimum value of the 'frequency' parameter is 30.
6. *Legal range - maximum.* This is the maximum legal value of the parameter. For example, the default legal maximum value of the 'frequency' parameter is 1200.
7. *Graph range - minimum.* This is the minimum parameter value to be plotted on a graph. For example, the default minimum graph range for the 'frequency' parameter is 30.
8. *Graph range - maximum.* This is the maximum parameter value to be plotted on a graph. For example, the default maximum graph range for the 'frequency' parameter is 1200.

In addition, a **Default log scale** button is provided. If this is true, then any visualization of this parameter will be mapped logarithmically by default. If this is not true, then any visualization of this parameter will be mapped linearly by default. For example, this is usually true for the 'frequency' parameter, but is not true for the 'time' parameter.

A new parameter may be created by the user by clicking on the button. A skeleton parameter is then created by ALBERT with dummy fields which the user then needs to tailor to his/her own requirements. A message will appear explaining that a new parameter has been created, and that the user should select the parameter from the parameter list in order to edit the parameter fields. This new parameter must be defined in the 'Equation' field shown at the bottom of the form. For example, if the user has created a new parameter to assess the performance of the CQ parameter with regard to the F0 parameter, he/she may feel it is appropriate to enter the following;

cq/f0

Once the other fields have also been entered, the new parameter can be accessed via the list of parameters presented on the Visualization form (section 4.10). It may then be used in the same manner as that used for the other parameters for axes mapping or colour display.

4.15 Period form

This form is used for analysis and animation of the electrolaryngograph period. It is shown in figure 2.3. It is usually used in conjunction with the voice source form (section 4.8).

1. **Smoothing style.** This enables the user to select the graphical style of smoothing used in the normalised period display. Four options are available; **Linear**, **Bspline**, **Cardinal**, and **Bezier**. The selection of these options do not otherwise alter the way in which the electrolaryngograph signal is processed.
2. **Miscellaneous.** In this section the user is able to control the form of buffering used in the normalised period display. **Single** or **Double** buffering may be selected. In the latter mode, the sequence of electrolaryngograph periods appears visually smoother, but at the cost of a small time delay which may be inappropriate for real-time operation with a heavy work load.

A button is also provided for the transfer of the normalised period window image to file. The image is transferred to a file named 'period.rgb'.

3. **Period frame-by-frame display.** This facility is provided to enable the user to examine a sequence of voice source activity that has been recorded to file. The user needs to enter the start time, end time and interval of analysis. For example, to see electrolaryngograph periods at one second intervals, starting at 2 seconds and ending at 5 seconds, the user needs to enter '2', '5', and '1' respectively. Finally, to start the analysis sequence, the user must choose between two options. These are **Show OK frames** and **Show all frames**. The former option instructs ALBERT to only display a period if one is encountered at the appropriate time interval. The latter instructs ALBERT to display the

nearest subsequent period if a period is not encountered at the required time interval. For both options, timings and frame numbers are shown underneath each displayed period.

4. **Animation.** This facility was developed to help in an explanation of the origin of the electrolaryngograph period.

Click on the **Show** button. A window outline will appear. Click on the left mouse button and drag out the required window size. An artificial pair of vocal folds will appear, and they will immediately move together and apart in a manner similar to that of real vocal folds.

The user can click on **Repeat once** to repeat the sequence. Usually, however, it is more appropriate to select the **Repeat continuously** button. The animation sequence will then continue until the user clicks on the **Repeat continuously** button a second time. The slider provided can be used to alter the speed of the animation.

For correct operation, the user needs to ensure one valid period of electrolaryngograph activity is recorded. Use of the normalised period display (section 4.8) will help with this. This is because the last closed quotient (CQ) parameter value derived by ALBERT is used in the animation sequence. For example, during the animation sequence the vocal folds will be shown in contact for a relatively greater period of time if a large CQ value was previously assessed than if a small CQ value was assessed.

5. **Voice period analysis.** By default, the normalised period display discussed previously (section 4.8) will only show one complete period. However, it is often useful to be able to see some of the previous period as well. For example, this is useful to ensure that the voice source assessment algorithm is operating correctly. To this end a slider is provided, labelled 'Percentage of previous period to show'. The user may then change the slider value to the desired level. The normalised display will change to reflect the new value, in real-time.

4.16 Save and load form

This form exists purely to give the user access over the saving and loading of several ALBERT domains. The following domains may be saved and loaded. They are listed here in the order in which they appear on the Save and load form.

1. *Visualization.* This covers all 1D, 2D and 3D graphs, and any associated statistics windows. The mapping of parameters to graphs, in addition to the position of graphs on the screen and within windows are all included.
2. *Items.* This includes the parameter fields of all voice parameters, including any new parameters that may have been created by the user.
3. *Forms.* This includes the position, size and visibility of every form.
4. *Acoustic display + flags.* This includes the position and size of both types of acoustic display and all flags associated with acoustic processing. This does not include any information that may be held within the 'Items' and 'Visualization' domains (discussed above).
5. *Voice source display + flags.* This includes the position and size of the normalized period display and all flags associated with electrolaryngograph processing. This does not include any information that may be held within the 'Items' and 'Visualization' domains (discussed above).
6. *All the above.* This covers all of the domains listed, which in aggregate completely define a state of ALBERT.

Note that the loading functions for these domains are also accessible via ALBERT command line options (section 5.1).

4.17 Forms ON OFF control

This form provides a means for the user to control the visibility of each and every form. A simple toggle button is provided for each form. Each button has an associated indicator light. If a form is currently displayed, the light is on. If the form

is currently not displayed, the light is off. The user may click on the appropriate button to instruct ALBERT to display or hide a form as he/she requires.

4.18 Forms size control

This form provides a means for the user to control the width and height of each form.

Three sliders are provided. Two are provided for the control of the x axis and y axis dimensions. The third is provided for easy control of both in combination. This slider enables the aspect ratio of the form to remain constant during the change in size. In order to use these controls, the user should click the toggle button corresponding to the relevant form once (if the form to be changed is not visible) or twice (if the form to be changed is visible). This is because changes made using the size control form are addressed only to the last form displayed. Slider values can then be altered. The form under alteration will immediately reflect the changes in dimension.

Chapter 5

Command line control

5.1 Command line options

To see these, just type

```
albert -I
```

ALBERT will just list the command line options, and will not start up in the normal way.

The following sequence (or similar) should appear:

```
This is albert version 2.8
```

```
See on-line manual for details
```

```
By David Rossiter 1993, 1994
```

Syntax :

```
albert          -I - show this information
                -P <filename> - play binary audio file on start up
                -v <filename> - load visualization config file
                -i <filename> - load items config file
                -f <filename> - load forms config file
                -a <filename> - load acoustic config file
                -p <filename> - load voice source config file
                -A <filename> - load complete setup config file
```

The second option, `-P <filename>`, enables ALBERT to point to a particular voice source and acoustic file recording by default. This is useful if, for example, the user is working on the analysis of a particular file.

The remaining options all provide command line access to the controls discussed in section 4.16.

5.2 Preparing aliases

This section describes one way in which users may find it easier to use ALBERT. It does not contain essential information, and may be skipped.

Normally, ALBERT is started simply by typing `albert`. ALBERT will then start in its default configuration. It is possible to configure ALBERT, and then to save the configuration information to file. This configuration can then be used when ALBERT is started up, if required. As discussed, the command structure for this would be;

```
albert -A config.all
```

If a particular configuration is regularly used, it is a good idea to create an alias for the command, to aid memory and reduce the amount of typing required. For example, if a configuration of ALBERT intended for voice source analysis is saved to a file called `'source.all'`, then it would be useful to have the following line in the `.cshrc` file in your home directory.

```
alias sbert 'albert -A source.all'
```

Thereafter, only the word `'sbert'` needs to be entered to start ALBERT in that particular configuration.

Other `'alias'` directives can be used to create a portfolio of easily labelled configurations of ALBERT.

Chapter 6

Functions keys

There are some special functions which are accessed by pressing one of the function keys. These are;

For any visualization display:

1. **F2** Move the mouse to a window. Press this button. The window image will be transferred to file. The name of the file will be derived according to the procedure outlined in section 4.10.

Concerning 2D graphs:

1. **F4** Point the mouse at a 2D window. Then press this button. The two dimensional graph is then translated to a three dimensional 'wafer' and passed to the three dimensional visualization program 'bob' for interactive viewing. The name of the intermediate file is *bob.XxYxZ*, where X, Y, and Z are the lengths of the relevant dimension. In this case, the Z dimension will be 1. Note that the current version of 'bob' is not able to *interactively* visualize a data slice of only one data item thick.
2. **F7** Point the mouse at a 2D window. Press this button. A new window is spawned. This shows mean and standard deviation, and total number of samples statistics for the parameters mapped to the graph.
3. **F9** Point the mouse at a 2D window. Press this button. The graph is then calibrated to use the entire range of the colour map. This is intended for use with scattergram displays after a sequence of data has been analysed.

Concerning 3D graphs:

1. **F10** Point the mouse at a 3D window. Press this button. The graph is then translated to the three dimensional visualization program 'bob' for interactive visualization. While 'bob' is running, normal program operation is temporarily halted. To stop the program, select **quit** from the 'bob' menus. The name of the intermediate file is *bob.XxYxZ*, where X, Y, and Z are the lengths of the relevant dimension.
2. **F11** Point the mouse at a 3D window. Press this button. The three dimensional graph is then saved in binary format, but the 'bob' program is not automatically executed. The name of the file is *bob.XxYxZ*, where X, Y, and Z are the lengths of the relevant dimension.

Appendix F

ALBERT parameter testing

F.1 Introduction

This Appendix describes tests made to establish the accuracy of the ALBERT software in assessing the voice parameters it is capable of analysing.

F.2 A computer program for testing ALBERT

In order to maintain precise control over the signal sequences with which ALBERT was to be tested, a computer program was written. This program was designed to generate a signal sequence and store it in a file which could then be read from ALBERT for assessment.

Several features were implemented so that a comprehensive range of test signal sequences could be generated. The user is provided with control over the following parameters.

1. The number of periods to output.
2. The (initial) period fundamental frequency.
3. The (initial) peak-to-peak amplitude of the period.
4. The mark phase percentage. For example, a value of 50 would be entered for a period of equal mark to space ratio.
5. Sine or square wave output.

6. The fundamental frequency increment to be applied to each successive period.
7. The peak-to-peak amplitude increment to be applied to each successive period.

The software code, which contains a description of the command line parameters, is listed at the end of this Appendix.

F.3 Closed quotient

To assess ALBERT analysis of closed quotient (CQ), a square wave signal was generated for a range of frequencies with constant mark:space ratio. This was performed for levels of CQ of 5%, 50%, and 95%. On each occasion, the signal sequence consisted of 1000 signal periods. The mean data are presented in table F.1. On each occasion, the standard deviation was assessed as zero.

<i>Frequency</i>	<i>5%</i>	<i>50%</i>	<i>95%</i>
50.0	5.000	50.000	95.000
100.0	5.000	50.000	95.000
200.0	5.000	50.000	95.000
400.0	5.000	50.000	95.000
500.0	5.000	50.000	95.000
750.0	5.000	50.000	95.000
1000.0	4.255	50.000	95.744

Table F.1: Assessment data for CQ values at input CQ levels of 5%, 50% and 95% across a range of input F0 values

From the table it can be seen that a precise representation of the input mark:space ratio is attained except at the highest frequency. In this instance, errors were evident for the minimum and maximum CQ levels. It is conjectured that these are due to quantisation errors, and that this error was not evident for the 50% level because the quotient was a factor of the number of samples representing one period. Consequently, a precise evaluation occurred at which quantisation error was not evident.

F.4 Fundamental frequency

In order to test ALBERT's analysis of fundamental frequency, a square wave signal was generated for a range of frequencies. On each occasion, the signal sequence consisted of 1000 signal periods. The mean data are presented in table F.2. On each occasion, the standard deviation was assessed as zero.

In each case, the signal is accurately analysed.

<i>Input F0</i>	<i>Assessed F0</i>
50	50.000
100	100.000
200	200.000
400	400.000
500	500.000
750	750.000
1000	1000.000

Table F.2: Assessment data for F0. Mean values are shown across a range of input F0 values

From the sequence of test signals presented to ALBERT, it would appear that no error exists in the assessment. However, this is because the periods assessed by ALBERT are of a length which are appropriate for accurate assessment. If an alternative value is generated, such as a frequency of 1025Hz, then a value of 1043.478 is returned and the error becomes apparent.

A discussion of the effects of quantisation is given in chapter 5.

F.5 Rate of opening and closing

In order to test ALBERT's analysis of the rate of opening and the rate of closing, square and sine wave signals were generated for a range of frequencies. Each signal sequence consisted of 1000 signal periods. The mean data are presented in table F.3. On each occasion, the value assessed for the rate of opening and the value for

the rate of closing were identical, and the standard deviation was assessed as zero.

The results are shown in figure F.3.

<i>F0</i>	<i>Square input</i>	<i>Sine input</i>
50	1.000	0.0033
100	1.000	0.0066
200	1.000	0.0131
400	1.000	0.0261
500	1.000	0.0327
750	1.000	0.0490
1000	1.000	0.0652

Table F.3: Assessment of rates of opening and closing for a selection of frequency values

Results for the square wave signal are consistent, and concord with theoretical expectations.

Results for the sine wave test signal increase with the fundamental frequency of the period. An examination of the data indicates that the assessed values are directly proportional to the input frequency. The equation for deriving the value of these parameters (labelled *ca* for the rate of closing, and *oa* for the rate of opening) for sine wave input can be expressed as follows.

$$ca = oa = n.F0$$

where *F0* is the fundamental frequency of the input signal and *n* is a constant of value $n = 0.000652$.

A discrete unit is interpreted by the computer as representative of one instant of the periodic signal. The value is proportional to frequency because the number of samples representing one period of the incoming signal is itself directly proportional to the frequency of the signal.

F.6 Peak to peak amplitude of the signal

In order to test ALBERT's analysis of the peak-to-peak amplitude of the period, a square wave was generated at a frequency of 200Hz at varying levels of peak-to-peak amplitude. Each signal sequence consisted of 1000 signal periods. The mean data are presented in table F.4. On each occasion, the standard deviation was assessed as zero.

The results are shown in figure F.4. Values are presented in terms of the number of quantisation levels from peak to peak. There is clearly a precise correlate.

<i>Input</i>	<i>Pk-to-pk amplitude</i>
30	30
300	300
3000	3000
30000	30000

Table F.4: Assessment of the peak-to-peak amplitude for a selection of input peak-to-peak amplitude values

F.7 Ratio

In order to test ALBERT's analysis of the 'Ratio' parameter, two sine waves were generated. One sine wave was generated at a frequency within the superset frequency range but outside the subset range. Another was generated at a frequency within the subset range. Both sine waves were of equal amplitude. ALBERT was then used to assess the value of the Ratio parameter. The amplitude of the sine wave signal within the subset range was then doubled, and the parameter assessed again. The amplitude of the sine wave was then set to be three times the amplitude of the other partial, at which the parameter was assessed, and again at four times the amplitude.

The values are shown in table F.5. There is a clear correlation between the theoretical results and the actual results.

<i>Ratio between partials</i>	<i>Parameter value</i>	<i>Theoretical value</i>
1:1	49.825	50.0
1:2	66.369	66.66
1:3	73.472	75.00
1:4	78.390	80.00

Table F.5: Assessment of the Ratio parameter

F.8 SPL

In order to test ALBERT's analysis of sound pressure level, one 200Hz sine wave was generated at a known peak-to-peak amplitude. ALBERT was then calibrated with this signal as a reference level of 50dB. The amplitude of the partial was then successively increased. On each occasion, the new SPL value was observed.

The results are shown in figure F.6 in addition to the theoretical equivalent.

<i>Relative increase</i>	<i>Parameter value</i>	<i>Theoretical value</i>
1:1	50.000	50+0.000
1:2	56.047	50+6.020
1:3	59.590	50+9.542
1:4	62.091	50+12.041
1:5	64.096	50+13.979
1:6	65.636	50+15.563
1:7	67.034	50+16.901
1:8	68.190	50+18.061
1:9	69.231	50+19.084
1:10	70.185	50+20.000

Table F.6: Assessment of the sound pressure level

F.9 Jitter

In order to test ALBERT's analysis of the 'jitter' parameter, a sequence of square waves was generated with each period one Hertz higher than the previous period. For example, if the initial fundamental frequency is 100Hz and the increment parameter is 5Hz then a sequence of periods will be generated with in the following fundamental frequency order.

100Hz

105Hz

110Hz

115Hz

120Hz

125Hz

...

Several sequences were generated. Each was generated with 101 periods. This ensures that the jitter value is calculated by ALBERT 100 times. The first column of table F.7 lists the initial fundamental frequency of the sequence. A frequency increment value of 1Hz was requested in each case. The mean and standard deviation of the analysis of the sequence are listed in the remaining two columns.

<i>Initial FO</i>	<i>Jitter mn</i>	<i>Jitter sd</i>
100	0.993	0.07
500	0.995	0.92
1000	0.910	1.95

Table F.7: Assessment of the Jitter parameter

Theoretically, there would be no deviation from the mean in each case. However, in practice a deviation from the mean is evident. The level of this error is proportional to the fundamental frequency. This is because of the effects of quantisation.

ALBERT is not always able to accurately determine the F0 value of each and every period (discussed previously in section F.4). This problem is particularly evident in the analysis of relatively high fundamental frequency signals. Because the accuracy of the jitter parameter is dependent on the accuracy of the F0 parameter, an error is present in evaluating the jitter parameter as a consequence of the error due to quantisation implicit in the derivation of the F0 parameter.

F.10 Electrolaryngograph period shimmer

In order to test ALBERT's analysis of the electrolaryngograph 'shimmer' parameter, a series of test signals were generated that employed signals of regularly increasing peak-to-peak amplitude. For example, if the initial peak-to-peak amplitude is 1000 quantisation levels, and the increment is 10 samples, then a sequence of periods will be generated with the following peak-to-peak amplitude values.

1000

1010

1020

1030

1040

1050

...

Several sequences were generated. Each sequence consisted of 1000 periods. The first period in each sequence had the peak-to-peak amplitude value listed in the first column in table F.8. An amplitude increment value of 10 was requested in each case. The mean data is presented in the second column in the table. In each case the standard deviation was zero. There is clearly a precise correlate.

<i>Initial amp</i>	<i>Amp mn</i>
100	10
1000	10
10000	10

Table F.8: Assessment of the electrolaryngograph Shimmer parameter

F.11 Acoustic shimmer

In order to test ALBERT's analysis of the 'shimmer' parameter derived from the acoustic signal a test sequence of increasing, known, peak-to-peak amplitude was generated. In order to create such a sequence, sine wave signals were generated. These sequences were then concatenated in order to produce a single test signal sequence. Each element of the concatenated sequence was identical in terms of fundamental frequency and the total number of periods in the element. However, each element of the sequence had a unique peak-to-peak signal amplitude. For each successive element of the concatenated sequence, the peak-to-peak amplitude was one order higher than that of the previous element.

An expression for the total sequence (labelled S_T) can be derived in terms of the sequence elements (each labelled S_x , where x is the peak-to-peak amplitude of the element, in samples) as follows.

$$S_T = S_{10} + S_{100} + S_{1000} + S_{10000}$$

The symbol $+$ is used to indicate the process of concatenation and not addition. The length of each element in the concatenated sequence (S_x) was set equal to the buffer length employed by ALBERT for the test. It is known that a order of magnitude increase in peak-to-peak amplitude of a sine wave is theoretically equivalent to a relative increase of 20dB. Therefore, the measure of shimmer under test should express a mean value of 20dB, with zero standard deviation.

A total of three different signal sequences were presented to ALBERT. Each employed a different peak-to-peak amplitude starting value. On each occasion,

ALBERT was first calibrated using the first element in the concatenated sequence as a reference of 10dB. All three sequences employed the same structure and timing of logarithmic increments. The mean and standard deviation results are presented in table F.9.

<i>Initial pk-to-pk amp</i>	<i>Shimmer mn</i>	<i>Shimmer sd</i>
10	20.064	0.061
30	20.074	0.063
60	20.097	0.070

Table F.9: Assessment of the Shimmer parameter

F.12 Signal generator software code

The code of the software program used to generate the signal sequences used for testing ALBERT is listed below.


```
/******
```

C source code for 'signal' program.

This program writes a sequence of square or sine waves to a binary soundfile in ALBERT format.

By David Rossiter 1995

To compile:

```
cc signal.c -lc_s -lgl_s -lm -laudio -o signal
```

Usage:

```
signal <output filename> <number of periods> <frequency>  
<peak-to-peak amplitude in samples> <mark phase percentage>  
<wavetype> <frequency increment> <peak-to-peak amp increment>
```

where :

<output filename> is the name of the binary file to which the sequence is to be written.

<number of periods> is the total number of periods to be written to file.

<frequency> is the (initial) period frequency.

<peak-to-peak amplitude> is the (initial) peak-to-peak amplitude of the signal, in samples.

<mark phase percentage> is the percentage of the period to be output as 'mark' state. For example, this value would be 50 to create a period of equal mark to space ratio.

<wavetype> Type of waveform. 1 for sinewave, 2 for squarewave.

<frequency increment> The increase in frequency for each successive

period, in Hz. A value of 0 ensures output of a constant frequency.

<peak-to-peak amp increment> The increase in peak-to-peak period amplitude for each successive period, in samples. A value of 0 ensures output of a constant peak-to-peak amplitude.

Note that the <mark phase percentage> option is not applicable if a sinewave is being generated.

[Further work: implement buffering operation]

*****/

/* Include files */

#include <fcntl.h>

#include <math.h>

#include <stdlib.h>

#include <malloc.h>

#include <stdio.h>

#include <gl.h>

#include <gl/gl.h>

#include <gl/device.h>

#include <audio.h>

#include <sys/file.h>

#include <sys/types.h>

#include <sys/ctype.h>

#include <sys/stat.h>

/* Defines */

#define SAMPLE_RATE 48000 /* Fixed sampling rate */

#define SINEWAVE 1 /* Flag for sine wave */

#define SQUAREWAVE 2 /* Flag for square wave */

/* Variables */

char filename[30]; /* Destination filename */


```

FILE *filepointer;      /* Pointer to destination file */

char signal_type;      /* Type of signal; sine or square */

float  freq_inc,        /* Increment in frequency each period */
      amp_inc;         /* Increment in amplitude each period */

long   loop,           /* Variables used in for() loops */
      period_number;

long   number_of_periods, /* Total number of periods to be output */
      samples_per_period; /* Number of samples for each period */

float  frequency,      /* Current period frequency */
      mark,            /* Current period frequency */
      mark_perc,      /* Current period frequency */
      amp;            /* Signal peak-to-peak amplitude */

short  mark_value,    /* Value of output signal during mark phase */
      space_value;   /* Value of output signal during space phase */

long   mark_length,   /* Number of samples for mark phase */
      space_length;  /* Number of samples for space phase */

/* Functions */
void update_sample_timing_values()
{
    /* Calculate number of samples per period */
    samples_per_period=((1.0f/frequency)*(float)SAMPLE_RATE);

    /* Calculate number of samples per period */
    mark=mark_perc/100.0f; /* Convert to fraction */

    /* Calculate number of samples per quotient */
    mark_length=(float)samples_per_period*mark;
    space_length=(float)samples_per_period*(float)(1.0f-mark);
}

```

```

void update_mark_space_values()
{
    /* Calculate sample values for mark and space quotients */

    mark_value=amp/2.0f;
    space_value=-1.0f*(amp/2.0f);
}

/***** Main program *****/

main(char *argc, char **argv)
{
    /* Obtain parameters from command line string */
    strcpy(filename, argv[1]);          /* Output filename */
    number_of_periods=(long)atof(argv[2]); /* Number of periods to be written */
    frequency=atof(argv[3]);           /* (Initial) fundamental frequency */
    amp=atof(argv[4]);                 /* (Initial) pk-to-pk amplitude */
    mark_perc=atof(argv[5]);           /* Mark quotient as perc of period */
    signal_type=atof(argv[6]);         /* Type of signal */
    freq_inc=atof(argv[7]);            /* Frequency increment each period */
    amp_inc=atof(argv[8]);             /* Amplitude increment each period */

    /* Update parameters */
    update_mark_space_values();
    update_sample_timing_values();

    /* Open file for output */
    if ((filepointer=fopen(filename,"w")) == NULL)
        printf("unable to open output file %s",filename);
    else

        /***** Period output loop *****/
        for (period_number=0;period_number<number_of_periods;period_number++) {

            if (signal_type==SQUAREWAVE) {

```



```

/***** MARK signal section *****/
for (loop=0;loop<mark_length;loop++) {
    fwrite(&mark_value, sizeof(short), 1, filepointer); /* Left channel */
    fwrite(&mark_value, sizeof(short), 1, filepointer); /* Right channel */
}

/***** SPACE signal section *****/
for (loop=0;loop<space_length;loop++) {
    fwrite(&space_value, sizeof(short), 1, filepointer); /* Left channel */
    fwrite(&space_value, sizeof(short), 1, filepointer); /* Right channel */
}

} /* end of 'if (signal_type==SQUAREWAVE)' */
else

if (signal_type==SINEWAVE) {

for (loop=0;loop<(mark_length+space_length);loop++) {

    mark_value=amp*0.5f          /* Calculate sinewave value (uses radians) */
        (float) sinf( (float) ((float)loop
            /(float)(mark_length+space_length)) *2.0f *3.141592f);

    fwrite(&mark_value, sizeof(short), 1, filepointer); /* Left channel */
    fwrite(&mark_value, sizeof(short), 1, filepointer); /* Right channel */
}

} /* end of 'if (signal_type==SINEWAVE)' */

/* Handle incremental frequency */
if (freq_inc) {
    frequency+=freq_inc; /* Update value */
    update_sample_timing_values();
}

/* Handle incremental peak-to-peak amplitude */
if (amp_inc) {
    amp+=amp_inc;          /* Update value */
}

```

```
        update_mark_space_values();
    }

}

fclose(filepointer); /* Close file */
}
```

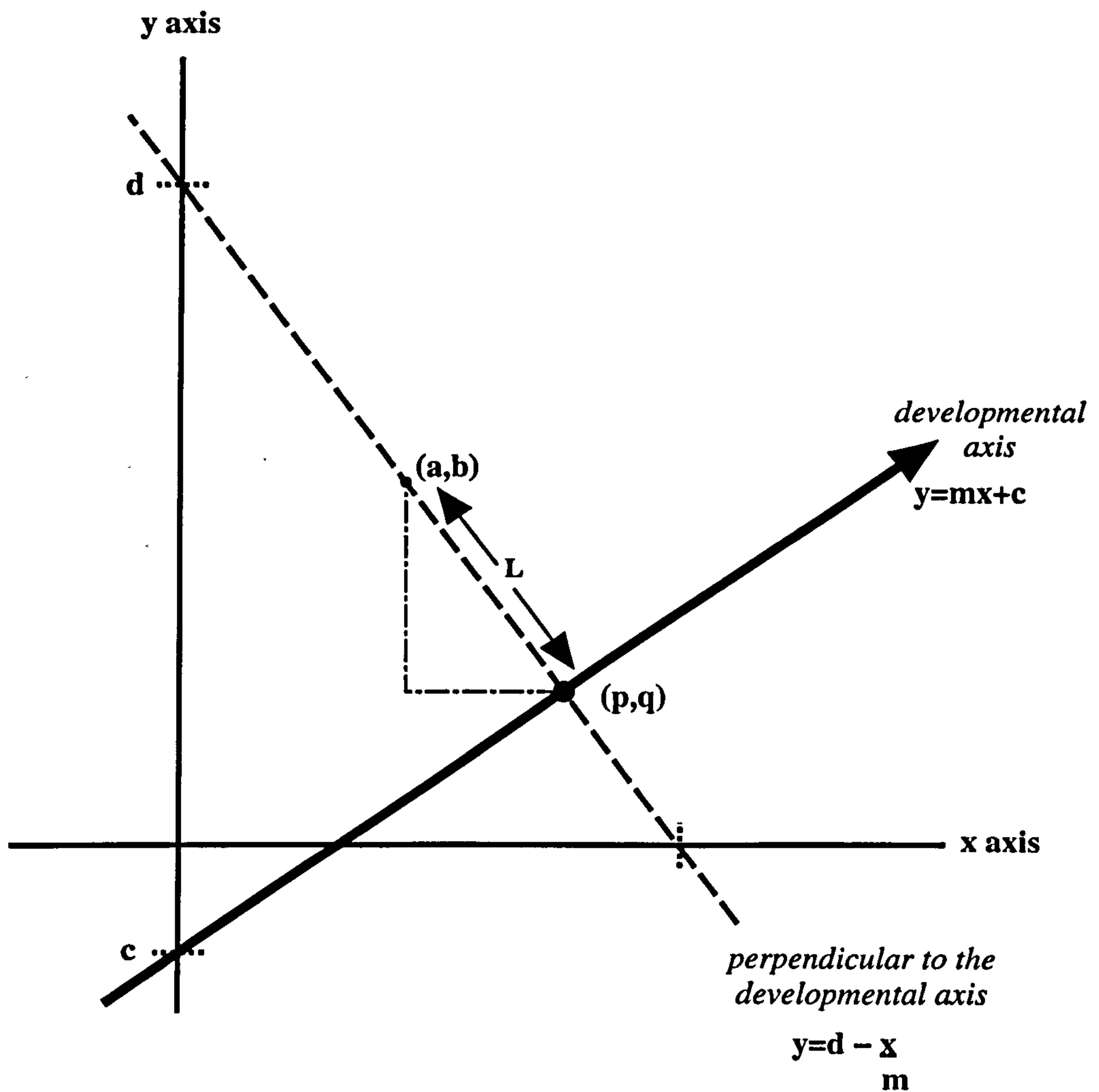

Appendix G

Progress along a linear correlation

G.1 Introduction

This appendix presents two equations which may be used to derive and present information in the ALBERT system described in chapter 5. Both equations may be used to present information relative to a linear relationship between two variables. The first (labelled G.4 below) provides a single measure representative of the relative progress of the user along this line. The second (labelled G.5 below) provides a single measure representative of the relative distance from the line.

Please refer to figure G.1. Both equations require the equation of a line (labelled the *developmental axis* with formula $y = mx + c$) against which the relative position of the data pair (labelled (a, b)) is to be compared. For example, in chapter 4 a continuum was proposed between the level of closed quotient (labelled CQ) of the vocal folds and the level of spectral amplitude in the singer's formant frequency band relative to all voice spectra output (labelled Ratio). The equation of the correlation between these two variables was determined as $y = 1.132x - 2.5$, with the x axis mapped to CQ and the y axis mapped to Ratio. The values for m (1.132) and c (-2.5) from this equation, in addition to new CQ and Ratio values derived in real time (entered as a and b respectively), may be entered into equation G.4 to produce a measure indicative of the progress of the voice user along this continuum. Thus,



- (a,b)** data pair derived in real-time
- (p,q)** point of intersection between developmental axis and perpendicular
- L** a measure of how far (a,b) is away from the developmental axis
- c** y offset of developmental continuum line
- d** y offset of perpendicular line

Figure G.1: Deriving a measure of advancement along a line in 2D space

the user needs only to enter this equation into a new parameter (detailed in section 5.10.3) and spawn a display of this parameter (section 5.12) to create an appropriate tool for real-time visual feedback.

In the following explanation, the 2D linear relationship against which progress is measured is referred to as the *developmental axis*, although this may be regarded simply as a convenient label.

G.2 Derivation

G.2.1 Measurement of position along a continuum

The first step is the derivation of a point along the developmental axis (labelled (p, q)) at which a perpendicular line will intersect the data pair (a, b) .

The equation of the developmental axis is known;

$$y = mx + c \quad (\text{G.1})$$

The formula for the perpendicular line is

$$y = d - \frac{x}{m} \quad (\text{G.2})$$

Rearranging G.2 to produce the offset for the perpendicular, labelled d ;

$$d = y + \frac{x}{m}$$

$$d = b + \frac{a}{m} \quad (\text{G.3})$$

At the point where the perpendicular intersects the developmental axis both line equations are equivalent. For example;

$$G.1 = G.2$$

$$mx + c = d - \frac{x}{m}$$

$$mp + c = d - \frac{p}{m}$$

$$p\left(m + \frac{1}{m}\right) = d - c$$

$$\frac{p(m^2 + 1)}{m} = d - c$$

$$p = \frac{m(d - c)}{1 + m^2}$$

Substituting G.3;

$$p = \frac{m\left(b + \frac{a}{m} - c\right)}{1 + m^2}$$

$$p = \frac{bm + a - cm}{1 + m^2} \tag{G.4}$$

G.2.2 Measurement of the distance of the data pair from the continuum

Using pythagoras' rule, a measure (L) may be derived for expressing how far away the data pair (a, b) is relative to the developmental axis (see figure G.1);

$$L^2 = ((b - q)^2 + (p - a)^2)$$

In terms of L ;

$$L = \sqrt{(b - q)^2 + (p - a)^2} \tag{G.5}$$

Bibliography

- [1] ABBERTON, E., & FOURCIN, A.J. (1978) Intonation and speaker identification. *Language and Speech*, vol. 21, no. 4, pp. 305-318
- [2] ABBERTON, E.R.M., HOWARD, D.M., & FOURCIN, A.J. (1989) Laryngographic assessment of normal voice: a tutorial. *Clinical Linguistics and Phonetics*, vol. 3, no. 3, pp. 281-296
- [3] ACKER, B.F. (1987) Vocal tract adjustments for the projected voice. *Journal of Voice*, vol. 1, no. 1, pp. 77-82
- [4] ALLEN, K.D., BERNSTEIN, B., & CHAIT, D.H. (1991) EMG biofeedback treatment of pediatric hyperfunctional dysphonia. *Journal of Behavioral Therapy & Experimental Psychiatry*, vol. 22, no. 2, pp. 97-101
- [5] ANASTAPLO, A. & KARNELL, M.P. (1988) Synchronized videostroboscopic and electroglottographic examination of glottal opening. *Journal of the Acoustical Society of America*, vol. 83, no. 5, pp. 1883-1890
- [6] ANDERSON, R.E. & HELSTRUP, T. (1993) Visual discovery on mind and on paper. *Memory & Cognition*, vol. 21, no. 3, pp. 283-293
- [7] ARDRAN, G.M., KEMP, F.H., & MANEN, L. (1953) Closure of the larynx. *The British Journal of Radiology*, vol. 26, pp. 497-509
- [8] ARENDS, N., POVEL, D.J, VAN OS, E., & MICHIELSEN, S. (1991) An evaluation of the Visual Speech Apparatus. *Speech Communication*, vol. 10, no. 4, pp. 405-414

- [9] ARONSON, A.E. (1985) *Clinical voice disorders: an interdisciplinary approach*, 2nd Ed., New York: Thieme Inc.
- [10] ASKENFELT, A., GAUFFIN, J. SUNDBERG, J., & KITZING, P. (1980) A comparison of contact microphone and electroglottograph for the measurement of vocal fundamental frequency. *Journal of Speech and Hearing Research*, vol. 23, no. 2, pp. 258-273
- [11] AWAN, S.N. (1993) Superimposition of speaking voice characteristics and phonetograms in untrained and trained vocal groups. *Journal of Voice*, vol. 7, no. 1, pp. 30-37
- [12] BAKEN, R.J. (1987) *Clinical measurement of speech and voice*. College-Hill Press: New York
- [13] BAKEN, R.J. & ORLIKOFF, R.F. (1987) The effect of articulation on fundamental frequency in singers and speakers. *Journal of Voice*, vol. 1, no. 1, pp. 68-76
- [14] BARRY, W.J., GOLDSMITH, M., FOURCIN, A.J. & FULLER, H. (1990) Larynx analyses on normative reference data. Project report, Alvey project MMI, London: University College
- [15] BEIJK, C. (1992) A comparative study of the fundamental frequency and frequency range of deaf and normally-hearing boys and girls, aged 4 to 20 years. *Transcripts of the Laryngograph User Group Meeting, held in London, 4 June*
- [16] BERKE, G.S. & GERATT, B.R. (1993) Laryngeal biomechanics: an overview of mucosal wave mechanics. *Journal of Voice*, vol. 7, no. 2, pp. 123-128
- [17] BLITZ, J. (1964) *Elements of acoustics*. London: Butterworths
- [18] BOYLE, E.A., ANDERSON, A.H., & NEWLANDS, A. (1994) The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, vol. 37, no. 1, pp. 1-20

- [19] BROOKS, S., FALLSIDE, F., GULIAN, E, & HINDS, P. (1981) Teaching vowel articulation with the computer vowel trainer: methodology and results. *British Journal of Audiology*, vol. 15, pp. 151-163
- [20] CARR, P.B., & TRILL, D. (1964) Long-term larynx excitation spectra. *Journal of the Acoustical Society of America*, vol. 36, pp. 575-82
- [21] CHAMBERS, D. & REISBERG, D. (1985) Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance*, vol. 11, pp. 317-328
- [22] CHENG, Y.M., & GUERIN, B. (1985) Control parameters in male and female glottal sources. In: *Laryngeal function in phonation and respiration*, Ed. Baer, T., Sasaki, C., & Harris, K.S., College-Hill: Toronto, pp. 219-238
- [23] CHILDERS, D.G., ALSAKA, Y.A., HICKS, D.M., & MOORE, G.P. (1987) Vocal fold vibrations: an EGG model. In: *Laryngeal function in phonation and respiration*, Ed. Baer, T., Sasaki, C., & Harris, K.S., Toronto: College-Hill, pp. 181-202
- [24] CHILDERS, D.G., & KRISHNAMURTHY, A.K. (1985) A critical review of electroglottography. *CRC Critical Reviews in Bioengineering*, vol. 12, pp. 131-161
- [25] CLARK, M.J.O. (1992) Acoustics of the singing voice. Unpublished final year project submitted as partial requirement for the degree of Physics, University of Edinburgh.
- [26] CLEVELAND, T.F. (1976) The acoustic properties of voice timbre types and their importance in the determination of voice classification in male singers. Ph.D. dissertation, University of Southern California
- [27] CLEVELAND, T., & SUNDBERG, J. (1983) Acoustic analysis of three male voices of different quality. *Proceedings of the Stockholm Music Acoustics Conference*, pp. 143-156
- [28] COLEMAN, R.F. (1987) Performance demands and the performer's vocal capabilities. *Journal of Voice*, vol. 1, no. 3, pp. 209-216

- [29] COMINS, R. (1992) Voice in the curriculum. *Voice*, vol. 1, no. 2, pp. 67-73
- [30] COOKSEY, J.M. (1993) Do adolescent voices 'break' or do they 'transform'? *Voice*, vol. 2, no. 1, pp. 15-39
- [31] COOLEY, J.W., & TUKEY, J.W. (1965) An algorithm for the machine computation of complex fourier series, *Mathematics of Computation*, vol. 19, pp. 297-301
- [32] COTT, R. (1994) No current title; paper in preparation. Speech and Language Department, South Tyneside District Hospital, South Shields, Tyne & Wear, NE34 OPL
- [33] CRYSTAL, D. (1988) *The Cambridge Encyclopedia of English Language*. Cambridge: Cambridge University Press
- [34] DAVIES, P., LINDSEY, G.A., FULLER, H. & FOURCIN, A.J. (1986) Variation in glottal open and closed phase for speakers of English. *Proceedings of the Institute of Acoustics*, vol. 8, pp. 539-546
- [35] DODGE, C., & JERSE, T.A. (1985) *Computer Music*. London: Schirmer books
- [36] DRAKOS, N. (1993) Text to hypertext conversion with LaTeX2HTML. *Baskerville*, vol. 3, no. 2, pp. 12-15
- [37] EDGINGTON, M.D., BARNES, C.M., STRINGER, P.D., & HOWARD, D.M. (1992) The speech filing system: a tool for cooperative speech research. *Proceedings of the Institute of Acoustics*, vol. 14, no. 6, pp. 79-86
- [38] ESTILL, J. (1988) Belting and classic voice quality: some physiological differences. In: *Medical problems of performing artists*, Philadelphia: Hanley and Belfus, pp. 37-43, March
- [39] EVANS, M., & HOWARD, D.M. (1993) Larynx closed quotient in female belt and opera qualities: a case study. *Voice*, vol. 2, no. 1, pp. 7-14
- [40] FANT, G. (1960) *Acoustic theory of speech production*. Mouton & co.: The Hague, Netherlands

- [41] FANT, G. (1983) Acoustic parameters of the voice source. Proceedings of the Stockholm Music Acoustics Conference, pp. 11-26
- [42] FOURCIN, A.J., & ABBERTON, E.R.M (1971) First applications of a new laryngograph. Medical and Biological Review, vol. 21, pp. 172-182
- [43] FRY, D.B. (1979) The physics of speech. Cambridge: Cambridge University Press
- [44] GARNER, P. (1993) Real-time vocal competence: assessment and development system. Unpublished transfer report to DPhil, Electronics Department, York University, England
- [45] GAUFFIN, J. & SUNDBERG, J. (1978) Pharyngeal constrictions. *Phonetica*, vol. 35, pp. 157-168
- [46] GAUFFIN, J. & SUNDBERG, J. (1989) Spectral correlates of glottal voice source waveform characteristics. *Journal of Speech and Hearing Research*, vol. 32, pp. 556-565
- [47] GELFER, M.P., ANDREWS, M.L., & SCHMIDT, C.P. (1991) Effects of prolonged loud reading on selected measures of vocal function in trained and untrained singers. *Journal of Voice*, vol. 5, no. 2, pp. 158-167
- [48] GILBERT, H.R., POTTER, C.R., & HOODIN, R. (1984) Laryngograph as a measure of vocal fold contact area. *Journal of Speech and Hearing Research*, vol. 27, pp. 178-182
- [49] GOLDSTEIN, M.H., & STARK, R.E. (1976) Modification of vocalizations of preschool deaf children by vibrotactile and visual displays. *Journal of the Acoustical Society of America*, vol. 59, pp. 1477-1481
- [50] GRAMMING, P. (1991) Vocal loudness and frequency capabilities of the voice. *Journal of Voice*, vol. 5, no. 2, pp. 144-157
- [51] GRAMMING, P., SUNDBERG, J., TERNSTROM, S., LEANDERSON, R., & PERKINS, W.H. (1988) Relationship between changes in voice pitch and loudness. *Journal of Voice*, vol. 2, no. 2, pp. 118-126

- [52] HADFIELD, P. (1994) How to make a robot smile. *New Scientist*, vol. 143, no. 1934, 16 July
- [53] HAJI, T., HORIGUCHI, S., BAER, T., & GOULD, W.J. (1986) Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation. *Journal of the Acoustical Society of America*, vol. 80, no.1, pp. 58-62
- [54] HARDCASTLE, W.J. (1981) Experimental studies in lingual coarticulation. In: *Towards a history of phonetics*, ed. Asher, R.E. & Henderson, E.J.A., Edinburgh: University Press
- [55] HENDEE, W.R., & WELLS, P.N.T (1993) *The perception of visual information*. London: Springer-Verlag
- [56] HERTEGARD, S., GAUFFIN, J., & SUNDBERG, J. (1990) Open and covered singing as studied by means of fiberoptics, inverse filtering, and spectral analysis. *Journal of Voice*, vol. 4, no. 3, pp. 220-230
- [57] HILLENBRAND, J., & GAYVERT, R.T. (1993) Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech and Hearing Research*, vol. 35, pp. 694-700
- [58] HIRANO, M. (1975) *Phonosurgery: Basic and clinical investigations*. *Otologia Fukuoka*, vol. 21, Sup. 1 (in Japanese)
- [59] HIRSON, A., & ROE, S. (1993) Stability of voice and periodic fluctuations in voice quality through the menstrual cycle. *Voice*, vol. 2, no. 2, pp. 78-88
- [60] HIRSON, A., FRENCH, J.P., & HOWARD, D.M. (1993). Forensic aspects of telephone speech: preliminary findings. In: Windsor Lewis, J. (Editor) *Studies in general and English phonetics: Essays in honour of Professor JD O'Connor*, London: Routledge
- [61] HIXON, T.J & COLLABORATORS. (1987) *Respiratory function in speech and song*. London: Taylor & Francis Ltd.
- [62] HOLLIEN, H. (1960) Vocal fold variation related to changes in vocal fold length. *Journal of Speech and Hearing Research*, vol. 3, no. 2, pp. 150-156

- [63] HOLLIEN, H. (1993) That golden voice - talent or training? *Journal of Voice*, vol. 7, no. 3, pp. 195-205
- [64] HOLLIEN, H., & SHIPP, T. (1972) Speaking fundamental frequency and chronological age in males. *Journal of Speech and Hearing Research*, vol. 15, pp. 155-159
- [65] HORII, Y. (1983) Automatic analysis of voice fundamental frequency and intensity using a visi-pitch. *Journal of Speech and Hearing Research*, vol. 26, pp. 467-471
- [66] HOWARD, D.M. (1991) Speech: measurements. In: *Concise encyclopedia of biological and biomedical measurement systems*, Ed. Peter A. Payne, Oxford: Pergamon Press
- [67] HOWARD, D.M. (1992) Quantifiable aspects of different singing styles - a case study. *Voice*, vol. 1, no. 1, pp. 47-62
- [68] HOWARD, D.M. (1994) Variation of electrolyngographically derived closed quotient for trained and untrained adult female singers. Accepted for publication by the *Journal of Voice*.
- [69] HOWARD, D.M., & GARNER, P. (1992) DSP-56000 based real-time electrolyngographically derived closed quotient. *Proceedings of the Institute of Acoustics*, vol. 14, no. 6, pp. 375-381
- [70] HOWARD, D.M., & LINDSEY, G.A. (1987) New laryngograms of the singing voice. *Proceedings of the 11th International Congress of Phonetic Sciences*, vol. 5, pp. 166-169
- [71] HOWARD, D.M., & LINDSEY, G. (1989) Spectral features of renowned tenors in CD recordings. *Proceedings of the International Conference on Speech Research*, pp. 17-20
- [72] HOWARD, D.M., LINDSEY, G.A., & ALLEN, B. (1990a) Toward the quantification of vocal efficiency. *Journal of Voice*, vol. 4, no. 3, pp. 205-212 {see also *Errata: Journal of Voice*, vol. 5, no. 1, pp. 93-95 (1991)}.

- [73] HOWARD, D.M., LINDSEY, G., & PALMER, S. (1990b) Voice quality measures for female singers. *Speech, Hearing and Language: Work in Progress*, University College London, No. 4, pp. 139-149
- [74] HOWARD, D.M., LINDSEY, G. & PALMER, S. (1991) Larynx closed quotient for the female voice. *Proceedings of the 12th International Congress of Phonetic Sciences*, No. 5, pp. 10-13
- [75] HOWARD, D.M., & WELCH, G.F. (1989) Microcomputer based singing ability assessment and development. *Applied Acoustics*, vol. 27, no. 2, pp. 89-102
- [76] HOWARD, D.M., & WELCH, G.F. (1993) Visual displays for the assessment of vocal pitch matching development. *Applied Acoustics*, vol. 39, pp. 235-252
- [77] HEARN, D., & BAKER, M.P. (1986) *Computer graphics*. London: Prentice-Hall
- [78] HUCKVALE, M.A., BROOKES, D.M., DWORKIN, L.T., JOHNSON, M.E., PEARCE, D.J., & WHITTAKER, L. (1987) The SPAR speech filing system. *Proceedings of the European Conference on Speech Technology*, pp. 305-317
- [79] HUGGINS, A.W.F. (1980) Better spectrograms from children's speech: a research note. *Journal of Speech and Hearing Research*, vol. 23, pp. 19-27, March
- [80] HUNT, R.W.G. (1987) *The reproduction of colour* (4th edn). London: Fountain Press
- [81] IMMOOS, M. (1993) The application of cognitive aspects of resonance therapy in voice teaching: imagery stimulation, unpublished document [A supplement to a poster presentation given at the Voice Foundation conference, Philadelphia, 1993: Immoos-Langlois, M. Application of resonance therapeutic principles in the area of vocal pedagogy]
- [82] JOHANSSON, C, SUNDBERG, J, & WILBRAND, H. (1983) X-ray study of articulation and formant frequencies in two female singers. *Proceedings of the Stockholm Music Acoustics Conference*, pp. 203-218

- [83] JONES, D. (1994) Catching a still in slow motion. The Guardian 'On Line' section. Thursday 23 June
- [84] KAY (1993) 'Computerized speech lab' promotional literature. Kay Elemetrics Corp, 12 Maple Avenue, P.O. Box 2025, Pine Brook, NJ 07058-2025, USA
- [85] KELLEY, A. (1985) Fundamental frequency measurements of female voices from twenty to ninety years of age. In: Aronson, A., Clinical voice disorders, (2nd ed.), New York: Thieme, pp. 51
- [86] KELMAN, A.W. (1981) Vibratory patterns of the vocal folds. *Folia Phoniatrica*, vol. 33, pp. 73-99
- [87] KENT, R.D. (1993) Vocal tract acoustics. *Journal of Voice*, vol. 7, no. 2, pp. 97-117
- [88] KENT, R.D., & BURKARD, R. (1981) Changes in the acoustic correlates of speech production. In: D.S.Beasley & G.A.David (editors), *Aging: communication processes and disorders*. New York: Grune & Stratton, pp. 47-62
- [89] KLINGHOLZ, F. (1990) Acoustic representation of speaking-voice quality. *Journal of Voice*, vol. 4, no. 3, pp. 413-219
- [90] LADEFOGED, P. (1973) The features of the larynx. *Journal of phonetics*, vol. 11, pp. 73-83
- [91] LAMPORT, L. (1986) *Latex: a document preparation system*. Wokingham: Addison-Wesley
- [92] LAVENDER, J. (1994) Back to basics. *The singer*, October/November, pp. 13-15
- [93] LAVER, J., & HANSON, R.J. (1981) Describing the normal voice. In: Darby, J.K., (ed.) *Speech evaluation in psychiatry*, New York: Grune and Stratton, pp. 57-78
- [94] LEONARD, R.J., RINGEL, R.L., DANILOFF, R.G., & HORII, Y. (1987) Voice frequency change in singers and nonsingers. *Journal of Voice*, vol. 1, no. 3, pp. 234-239

- [95] LIEBERMAN, P. (1977) *Speech physiology and acoustic phonetics*. London: Macmillan
- [96] LIEBERMAN, P., & BLUMSTEIN, S.E. (1988) *Speech physiology, speech perception, and acoustic phonetics*. pp. 3-15, Cambridge: Cambridge University Press
- [97] LINDSEY, G.A., & HOWARD, D.M. (1989) Larynx excitation in the singing voice. *Proceedings of the professional voice (UK) 3rd international symposium*, pp. 6-13
- [98] LINDSEY, G., BREEN, A., & FOURCIN, A. (1988) Glottal closed time as a function of prosody, style and sex in English. *Proceedings of Speech 1988: 7th FASE symposium*, pp. 1101-1106
- [99] LOFQVIST, A. (1986) The long-term average spectrum as a tool in voice research. *Journal of Phonetics*, vol. 14, pp. 471-5
- [100] LUDLOW, C.L., BASSICH, C.J., CONNOR, N.P., COULTER, D.C., & LEE, Y.J. (1987) The validity of using phonatory jitter and shimmer to detect laryngeal pathology. In: Baer, T., Sasaki, C., Harris, K.S., (Ed.s) *Laryngeal pathology in phonation and respiration*. Boston: College-Hill, pp. 492-508
- [101] MACDONALD, L.W. (1990) Using colour effectively in displays for human-computer interface. *Displays*, vol. 11, no. 3, pp. 129-141
- [102] MACCURTAIN, F., METZ, D.E., & REED, V.W. (1981) Panel discussion: Measurement of voice quality at the source. In: *Transcripts of the tenth symposium, Care of the professional voice. Part 2: Instrumentation in voice quality, teaching, medicine*. New York: The Voice Foundation
- [103] MACNEILAGE, P.F., LADEFOGED, P., & SAWASHIMA, M. (1979) Speech production. *Proceedings of the 9th International Congress of Phonetic Sciences*, no. 1, pp. 11-40
- [104] McCRONE, J. (1994) Trainers for your brain. *New Scientist*, 3 September, pp. 35-39

- [105] MICHI, K., YAMASHITA, Y., IMAI, S., SUZUKI, N., & YOSHIDA, H. (1993) Role of visual feedback treatment for defective /s/ sounds in patients with cleft palate. *Journal of Speech and Hearing Research*, vol. 36, pp. 277-285
- [106] MILLER, D.G., & SCHUTTE, H.K. (1990) Feedback from spectrum analysis applied to the voice. *Journal of voice*, vol. 4, no. 4, pp. 329-334
- [107] MORGAN, N. (1994) Sound perception. *Future Music*, November, pp. 67
- [108] MOORE, R.S. (1994) Selected research on children's singing skills. In: Welch, G., & Murao, T. (Ed.s) *Onchi and singing development*, pp. 41-48, London: David Fulton
- [109] MORROW, A. (1994) Lost for words. *The Guardian*, Education supplement. October 4, pp. 4-5
- [110] NILSONNE, A., & SUNDBERG, J. (1985) Differences in ability of musicians and nonmusicians to judge emotional state from the fundamental frequency of voice samples. *Music Perception*, vol. 2, no. 4, pp. 507-516
- [111] ORLIKOFF, R.F., & KAHANE, J.C. (1991) Influence of mean sound pressure level on jitter and shimmer measures. *Journal of Voice*, vol. 5, no. 2, pp. 113-119
- [112] ORTON, J.L. (1938) *Voice culture made easy*. London: Thorsons
- [113] OSTER, A.M. (1988) Computer-based speech training. *Proceedings of Speech 1988*, no. 2, pp. 645-651
- [114] PAYNE, P.A. (1991) *Concise encyclopedia of biological and biomedical measurement systems*. Permagon Press: London
- [115] PENDERSON, M.F., MOLLER, S., KRABBE, S., BENNET, P. & KITZING, P. (1984) Change of voice in puberty in choir girls. *Acta Otolaryngology, Sup.* 412, pp. 46-49
- [116] PENDERSON, M.F., ZKITZING, P., KRABBE, S. & HERAMB, SD. (1982) The change of voice during puberty in 11 to 16 years old choir singers measured with electrolaryngographic fundamental frequency analysis and compared to other phenomena of puberty. *Acta Otolaryngology, Sup.* 386, pp. 189-192

- [117] PERKINS, W.H. & KENT, R.D. (1986) Textbook of functional anatomy of speech, language, and hearing. London: Taylor & Francis
- [118] PICKETT, J.M. (1991) The sounds of speech communication: a primer of acoustic phonetics and speech perception. Chicago: Allyn & Bacon
- [119] PRINGLE, M.B. (1993) Practical aspects of establishing a combined voice clinic in Wexham Park Hospital. *Voice*, vol. 2, no. 2, pp. 97-111
- [120] PROCTOR, D.F. (1980) Breathing, Speech and Song. New York: Springer-Verlag
- [121] RAPHAEL, B.N., & SCHERER, R.C. (1987) Voice modifications of stage actors: acoustic analyses. *Journal of Voice*, vol. 1, no. 1, pp. 83-87
- [122] REED, S.K., & JOHNSEN, J.A. (1975) Detection of parts in patterns and images. *Memory and Cognition*, vol. 3, pp. 569-575
- [123] ROSSITER, D., HOWARD, D.M., DOWNES, M., & GARNER, P. (1993) Real time visual displays for singing development. Proceedings of the 11th Eurographics Conference, pp. 63-72
- [124] ROSSING, T.D. (1990) The science of sound. Wokingham: Addison-Wesley
- [125] ROTHENBERG, M. (1973) A new inverse filtering technique for deriving the glottal airflow waveform during voicing. *Journal of the Acoustical Society of America*, vol. 53, pp. 1632-1645
- [126] ROTHENBERG, M., & MAHSHIE, J.J (1988) Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research*, vol. 31, pp. 338-351
- [127] ROUBEAU, B., CHEVRIE-MULLER, C., & ARABIA-GUIDET, C. (1987) Electrolottographic study of the changes of voice registers. *Folia Phoniatica*, vol. 39, pp. 280-289
- [128] RUBIN, H.J., LECOVER, M., & VENNARD, W. (1967) Vocal intensity, subglottic pressure and air flow relationships in singers. *Folia Phoniatica*, vol. 19, pp. 393-413

- [129] SAKAKURA, A., & TAKAHASHI, H. (1988) Body wall vibration in trained and untrained voices. In: *Vocal Physiology: Voice production, mechanisms and functions*, edited by Fujimura, O., New York: Raven Press, pp. 391-401
- [130] SAPIR, S. (1993) Vocal attrition in voice students: survey findings. *Journal of Voice*, vol. 7, no. 1, pp. 69-745
- [131] SATALOFF, R. T. (1992) The Human Voice. *Scientific American*. December, vol. 267, no. 6, pp. 108-115
- [132] SATALOFF, R.T., SPIEGEL, J.R., CARROLL, L.M., DARBY, K.S., HAWKSHAW, M.J., & RULNICK, R.K. (1990) The clinical voice laboratory: practical design and clinical application. *Journal of Voice*, vol. 4, no. 3, pp. 264-279
- [133] SAWASHIMA, M., & HIROSE, H. (1981) Abduction-adduction of the glottis in speech and voice production. In: *Vocal fold physiology*, pp. 329-346
- [134] SCHUTTE, H.K., & MILLER, D.G. (1993) Belting and pop, nonclassical approaches to the female middle voice: some preliminary considerations. *Journal of Voice*, vol. 7, no. 2, pp. 142-150
- [135] SCHERER, R.C., & TITZE, I.R. (1987) The abduction quotient related to vocal quality. *Journal of Voice*, vol. 1, no. 3, pp. 246-251
- [136] SCRIPTURE, E.W. (1902) *The elements of experimental phonetics*. New York: Charles Scribner's Sons. Reprinted by New York: AMS Press, 1973
- [137] SEIDNER, W., SCHUTTE, H., WENDLER, J., & RAUHUT, A. (1983) Dependence of the high singing formant on pitch and vowel in different voice types. In: *Proceedings of the Stockholm Music Acoustics Conference*, ed. A. Askenfelt, S. Felicetti, E. Jansson, & J. Sundberg, pp. 261-268
- [138] SIMONSON, D. (1992) Harmonic relationships between sung pitch, the first vowel formant and the singing formant: a study in enhancing acoustical efficiency in singing. *Voice*, vol. 1, no. 2, pp. 103-124
- [139] SLAVIT, D.H., & MARAGOS, N.E. (1994) Arytenoid adduction and type I thyroplasty in the treatment of aphonia. *Journal of Voice*, vol. 8, no. 1, pp. 84-91

- [140] SMITH, L. & MONK, T. (1994) SST: a program for teaching sight singing. A demonstration presented at the International Computer Music Conference. Listed in the Proceedings, pp. xiii
- [141] SPAAI, G.W.G., STORM, A., & HERMES, D.J. (1993) A visual display for the teaching of intonation to deaf persons: some preliminary findings. *Journal of Microcomputer Applications*, vol. 16, pp. 277-286
- [142] STABLES, A. (1992) Speaking and listening at key stage 3: some problems of teacher assessment. *Educational Research*, vol. 34, no. 2, pp. 107-115
- [143] STARK, R.E. (1972) Teaching /ba/ and /pa/ to deaf children using real-time spectral display. *Language and Speech*, vol. 15, pp. 14-29
- [144] STATHOPOULOS, E.T., & SAPIENZA, C. (1993) Respiratory and laryngeal function of women and men during vocal intensity variation. *Journal of Speech and Hearing Research*, vol. 36, pp. 64-75
- [145] STEWART, L.C., LARKIN, W.D., & HOUDE, R.A. (1976) A real-time sound spectrograph with implications for speech training for the deaf. *IEEE international conference on acoustics, speech and signal processing*, vol. ASSP-33, no. 1, pp. 164-173
- [146] STEVENS, K.N. (1988) Modes of vocal fold vibration based on a two-section model. *Vocal physiology: Voice production, mechanisms and functions*. Ed. Fujimura, O., New York: Raven Press
- [147] STRADLING, P. (1968) *Diagnostic bronchoscopy: an introduction*. London: E & S Livingstone, Ltd.
- [148] SUMBY, W.H., & POLLACK, I. (1954) Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, vol. 26, pp. 212
- [149] SUNDBERG, J. (1970) Formant structure and articulation of spoken and sung vowels. *Folia Phoniatica*, vol. 22, pp. 28-48

- [150] SUNDBERG, J. (1974) Articulatory interpretation of the "singing formant". *Journal of the Acoustical Society of America*, vol. 55, pp. 838-844
- [151] SUNDBERG, J. (1977) The acoustics of the singing voice. *Scientific American*. April, pp. 82-91
- [152] SUNDBERG, J. (1987) *The science of the singing voice*. Illinois: Dekalb Press
- [153] SUNDBERG, J., & GAUFFIN, J. (1979) Waveform and spectrum of the glottal voice source. In: Lindblom & Ohmau, (Editors), *Frontiers of Speech communication research*, pp. 301-320
- [154] TEACHEY, J.C., KAHANE, J.C., & BECKFORD, N.S. (1991) Vocal mechanics in untrained professional singers. *Journal of Voice*, vol. 5, no. 1, pp. 51-56
- [155] TITZE, I.R. (1990) Interpretation of the electroglottographic signal. *Journal of Voice*, vol. 4, no. 1, pp. 1-9
- [156] TITZE, I.R., & STRONG, W.J. (1976) Normal modes in vocal cord tissues. *Journal of the Acoustical Society of America*, vol. 67, no. 3, pp. 736-743
- [157] TITZE, I.R., LUSCHEI, E.S., & HIRANO, M. (1989) Role of the thyroarytenoid muscle in regulation of fundamental frequency. *Journal of Voice*, vol. 3, no. 3, pp. 213-224
- [158] TUFTE, E. (1983) *The visual display of quantitative information*. Cheshire, Connecticut: Graphics Press
- [159] WALDEN, B.E., BUSACCO, D.A., & MONTGOMERY, A.A. (1993) Benefit from visual cues in auditory-visual speech recognition by middle-aged and elderly persons. *Journal of Speech and Hearing Research*, vol. 36, pp. 431-436
- [160] WALKER, R. (1992) Formants in singers: stylistic and cultural perspectives. *Voice*, vol. 1, no. 1, pp. 19-46
- [161] WANG, S. (1983) Singing voice: bright timbre, singer's formant, and larynx position. In: *Proceedings of the Stockholm Music Acoustics Conference*, Editors: A. Askenfelt, S. Felicetti, E. Jansson, and J. Sundberg, pp. 313-322

- [162] WARD, D., & BURNS, E. (1978) Singing without auditory feedback. *Journal of Research in Singing and Applied Vocal Pedagogy*, vol. 1, no. 2, pp. 24-44
- [163] WATANABE, A., UEDA, Y., & SHIGENAGA, A. (1985) Color display system for connected speech to be used for the hearing impaired. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no. 1, pp. 164-173
- [164] WATSON, C. (1992a) An acoustic investigation into higher partial enhancement in the voices of a group of opera singers. PhD thesis, Edinburgh: Queen Margaret College
- [165] WATSON, C. (1992b) Higher partial enhancement and glottal source manipulation by the trained opera singer. *Voice*, vol. 1, no. 1, pp. 1-45
- [166] WATSON, C. (1993) Personal communication to the author
- [167] WATSON, C., & MACAULAY, R. (1994) Modal instability of vocal fold vibration in patients with Parkinson's disease shown by jitter analysis. *Voice*, vol. 4, no. 1, pp. 27-36
- [168] WELCH, G.F. (1986) A developmental view of children's singing. *British Journal of Music Education*, vol. 3, no. 3, pp. 295-303
- [169] WELCH, G.F., HOWARD, D.M., & RUSH, C. (1989a) Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music*, vol. 17, pp. 146-157
- [170] WELCH, G.F., RUSH, C., & HOWARD, D.M. (1988) The SINGAD (SINGing Assessment and Development) system: first applications in the classroom. *Proceedings of the Institute of Acoustics*, vol. 10, no. 2, pp. 179-185
- [171] WELCH, G.F., RUSH, C., & HOWARD, D.M. (1991) A developmental continuum of singing ability: evidence from a study of five-year-old developing singers. *Early Child Development and Care*, vol. 69, pp. 107-119
- [172] WELCH, G.F., SERGEANT, D.C., & MACCARTAIN, F. (1989b) Xeroradiographic-electrolaryngographic analysis of male vocal registers. *Journal of Voice*, vol. 3, no. 3, pp. 244-256

- [173] WENZEL, E.M. (1992) Localization in virtual acoustic displays. *Presence*, vol. 1, pp. 80-107
- [174] WHITE, P.J., WALKER, R., & WELCH, G.F. (1994) Spectral energy measures of trained and untrained children's voices. Paper in preparation, Leverhulme Project, Roehampton Institute, London
- [175] WILLIAMS, C.E., & STEVENS, K.N. (1972) Emotions and speech: acoustical correlates. *Journal of the Acoustical Society of America*, vol. 52, no. 4, pp. 1238-1250
- [176] WINCKEL, F. (1971) How to measure the effectiveness of stage singers voices. *Folia Phoniatica*, vol. 23, pp. 228-233
- [177] WIRZ, S.L., & ANTHONY, J. (1979) The use of the voicscope in improving the speech of profoundly deaf children. *British Journal of Disorders of Communication*, vol. 14, no. 2, pp. 137-151
- [178] YANAGISAWA, E., ESTILL, J., KMUCHA, S., & LEDER, S. (1989) The contribution of aryepiglottic constriction to 'ringing' voice quality - a video-laryngoscopic study with acoustic analysis. *Journal of Voice*, vol. 3, no. 4, pp. 342-350
- [179] YOUDELMAN, K., MACEACHRON, M., BEHRMAN, M. (1988) Visual and tactile sensory aids: integration into an ongoing speech training program. *Volta Review*, vol. 90, no. 4, pp. 197-207