ORGANIC VARIATION AND VOICE QUALITY

JANET MACKENZIE BECK

Thesis submitted for the degree of Ph.D.
University of Edinburgh
1988

EX BIBL. UNIV. EDINBURGEN.

ACKNOWLEDGEMENTS

I have many people to thank that this thesis has been completed, and cannot hope to include here a full list of all the colleagues and friends who have helped in one way or another. My thanks go to everyone who has been involved.

Special thanks must go to John Laver, for his thought and his time, and for more patience than should ever be asked of a supervisor. Steve Hiller also deserves special thanks, for both practical and moral support. I am indebted to Sheila Wirz, who first prompted me to start voice quality research, and to Edmund Rooney, who worked on the later stages of the acoustics project.

I owe much to our consultants and collaborators on the two MRC projects. Dr. W.I. Fraser, Mr. T. Harris, Mrs. S. Collins, Mrs. M. Mackintosh, Mrs. R. Nieuwenhuis and Dr. A. Maran all allowed access to both patients and medical records. I am grateful to Stewart Smith, Jeff Dodds, Norman Dryden and Irene Macleod, for technical assistance, and to Anne Anderson for advice on statistics.

Nicola Robinson gave me time to complete the writing by caring for my children, as did my parents and other members of the family.

Above all, I want to thank my husband Al, for the tedious task of proof reading, and for generally bearing the brunt of it all.

DECLARATION

This thesis was composed by myself, and represents an original and substantial contribution to the work of a research group investigating aspects of voice quality. I was employed as a Research Associate on two projects funded by the Medical Research Council ('Vocal Profiles of Speech Disorders, MRC Grant No. G978/1192 and 'Acoustic Analysis of Voice Features', MRC Grant No. 98207136N), and was responsible for the collection and interpretation of clinical data, as well as being closely involved in theoretical developments.

.....

Janet M. Mackenzie Beck

March 1988

TABLE OF CONTENTS

ACKNOWLEDGEMENTS
DECLARATION
TABLE OF CONTENTS
LIST OF FIGURES
ABSTRACT

INTRODUCTION 1

PART ONE: BUILDING THE VOCAL APPARATUS
1.1 STRUCTURAL COMPONENTS
1.1.1 Topographical orientation 6
1.1.2 Basic building blocks (cells and tissues) 8
1.1.3 Mechanical characteristics of tissues with special reference to the vocal fold 33
1.2 PRINCIPLES OF GROWTH AND CHANGE
1.2.1 Growth mechanisms during development 47
1.2.2 Growth mechanisms in maintenance and repair 73
1.2.3 Degenerative change and neoplastic growth 81
1.2.4 Growth and change of the vocal apparatus 92
1.2.5 The consequences of growth and change for voice quality 118
1.3 CONCLUSION TO PART ONE 126

PART TWO: ORGANIC AND PHONETIC RELATIONSHIPS IN SPEECH
2.1 VOICE QUALITY ANALYSIS TECHNIQUES
2.1.1 Introduction 128
2.1.2 Perceptual analysis of voice quality: the Vocal Profile Analysis Scheme 131
2.1.3 Acoustic analysis of voice quality 190
2.2 PERCEPTUAL ANALYSIS OF NORMAL VOICE QUALITY 196
2.3 VOICE QUALITY IN DOWN'S SYNDROME
2.3.1 Introduction 202
2.3.2 Organic characteristics of Down's Syndrome 206

2.3.3 Experimental investigation 232
2.3.4 Discussion and conclusions 238
2.4 ACOUSTIC CHARACTERISTICS OF NORMAL PHONATION 243
2.5 ACOUSTIC ANALYSIS IN LARYNGEAL PATHOLOGY
2.5.1 Introduction 248
2.5.2 Predicted consequences of vocal fold pathology 250
2.5.2 Organic vocal fold pathologies 262
2.5.4 Experimental investigation 276
2.5.5 Discussion and conclusions 296

PART THREE: CONCLUSIONS 298

BIBLIOGRAPHY 303

APPENDIX 1: "Vocal Profile Analysis Scheme: A User's Manual"

APPENDIX 2: "A perceptual protocol for the analysis of vocal profiles" (Reprint of Laver et al. 1981)

APPENDIX 3: "Structural pathologies of the vocal folds and phonation" (Reprint of Mackenzie et al. 1983)

APPENDIX 4: "Automatic analysis of waveform perturbations in connected speech" (Reprint of Hiller et al. 1983)

APPENDIX 6: "An acoustic screening system for the detection of laryngeal pathology" (Reprint of Laver et al. 1986)

APPENDIX 7: "Voice quality as an expressive system in mother-to-infant communication (Reprint of Marwick et al. 1984)

APPENDIX 8: "The use of two voice analysis techniques in clinic (Reprint of Nieuwenhuis and Mackenzie 1986)

LIST OF FIGURES

1.1.1/1 Anatomical planes

1.1.1/2 Schematic representation of the vocal apparatus

1.1.1/3 Lateral view of the skull

1.1.1/4 External surface of the base of the skull, to show the position of the superior constrictor muscle

1.1.1/5 Medial view of the mandible

1.1.1/6 Sagittal section through the skull

1.1.1/7 Schematic diagram showing suspension of the hyoid bone

1.1.1/8 Lateral view showing the constrictor muscles of the pharynx

1.1.1/9 The cartilages and ligaments of the larynx

1.1.1/10 Coronal section of the larynx

1.1.1/11 Median section of the larynx

1.1.2/1 Schematic representation of a generalized animal cell

1.1.2/2 Schematic representation of the principal types of epithelium lining the vocal apparatus

1.1.2/3 Exocrine and endocrine glands

1.1.2/4 Unicellular exocrine glands

1.1.2/5 Multicellular exocrine glands

1.1.2/6 Schematic representation of types of connective tissue proper

1.1.2/7 Schematic representation of types of cartilage

1/1/2/8 Schematic representation of dense and spongy bone

1.1.2/9 Longitudinal section of a tooth

1.1.2/10 Schematic diagram of skeletal muscle

1.1.2/11 Diagram of skeletal muscle contraction

1.1.2/12 Variation in muscle architecture

1.1.3/1 Schematic view of the vocal folds, seen from above

1.1.3/2 Diagrammatic representation of the tissue layers of the vocal folds

1.1.3/3 Schematic representation of the ligamental portion of the vocal fold, seen in cross section

1.1.3/4 Graphic representation of tissue thickness variation along the glottal edge of the ligamental portion of the vocal fold

1.1.3/5 Schematic diagram of the vocal fold in horizontal section to show the maculae flavae

1.1.3/6 A summary of the mechanical properties of vocal fold tissues

1.1.3/7 Diagram of Titze's 16-mass model of vocal fold vibration

1.2.1/1 Standard height growth curves for British children

1.2.1/2 Standard weight growth curves for British children

1.2.1/3 Height growth velocity curves for British children

1.2.1/4 Weight growth velocity curves for British children

1.2.1/5 Graphic representation of changes in bodily proportions during development

1.2.1/6 Growth curves of reproductive tissue, brain and head and lymphoid tissue, compared with the general growth curve

1.2.1/7 Interstitial and appositional growth

1.2.1/8 Schematic diagram of mitosis

1.2.1/9 Schematic diagram of the developmental origin of exocrine and endocrine glands

1.2.1/10 Schematic diagram of hyaline cartilage development

1.2.1/11 Schematic diagram of intramembraneous ossification

1.2.1/12 Schematic diagram of long bone growth

1.2.1/13 Schematic diagram of tooth development

1.2.2/1 Schematic diagram of bone fracture repair

1.2.3/1 Schematic representation of neoplastic growth patterns

1.2.4/1 Changing proportions of the skull from birth to maturity

1.2.4/2 Sagittal view of the skull at birth, showing fontanelles

1.2.4/3 Schematic diagram of cranial growth

1.2.4/4 Sex differences in cranial width measurements

1.2.4/5 Landmarks of the facial skeleton

1.2.4/6 Graphic summary of palatal growth

1.2.4/7 Schematic diagram of mandibular growth

1.2.4/8 Changing angle of the mandible

1.2.4/9 Incisor relationships in infancy and adulthood

1.2.4/10 Angle classes of malocclusion

1.2.4/11 Spinal curvature changes from birth to old age

1.2.4/12 Lung growth curves

1.2.4/13 Thoracic growth curves

1.2.4/14 Rib elevation and thoracic volume in infancy and adulthood

1.2.4/15 Sex differences in thyroid cartilage contour

1.2.4/16 Graphic representation of vocal fold tissue thickness in subjects aged 20-29 years

1.2.4/17 Graphic representation of vocal fold tissue thickness in subjects aged 50-59 years

1.2.4/18 Typical ages of tooth eruption

1.2.4/19 *Newborn and adult vocal tracts*

1.2.5/1 A graphic summary of reported average speaking

F0 at different ages
A. Females
B. Males

2.1.2/1 Radiographic diagram of the vocal tract in a neutral setting

2.1.2/2 Vocal Profile Analysis Protocol

2.1.2/3 Diagram of changes in vocal tract configuration and vowel distribution in neutral and a fronted and raised tongue body setting

2.1.2/4 A summary of laryngeal tension parameters in different phonation settings

2.1.2/5 Larynx configurations in different phonation settings

2.1.2/6 Summary of key segments for vocal quality settings

2.1.2/7 Criteria for assessing judge agreement

2.1.2/8 Table showing levels of inter- and intra-judge agreement for control voices (MRC project staff)

2.1.2/9 Table showing levels of inter- and intra-judge agreement for PD voices (MRC project staff)

2.1.2/10 Histograms showing distribution of trainee judge agreement levels

2.1.2/11 Table showing trainee judge agreement levels

2.1.2/12 Example of summated Vocal Profile Analysis results; a group of hearing impaired speakers

2.1.3/1 Flow chart of perturbation analysis system

2.1.3/2 Diagrammatic representation of raw and smoothed F0 curves

2.1.3/3 Acoustic Profile form

2.2/1 Subject group characteristics

2.2/2 Summated Vocal Profile Analysis results for all normal subjects

2.2/3 Table of mean scalar degrees and standard deviations for all subjects

2.2/4 Summated Vocal Profile Analysis results for female subjects

2.2/5 Summated Vocal Profile Analysis results for male subjects

2.2/6 Table of mean scalar degrees, standard deviations and statistical significance of sex differences for male and female subjects

2.3/1 Graph of reported speaking F0 in DS and normal children

2.3/2 Normal human chromosome complement, and DS variants

2.3/3 Normal and DS height growth curves

2.3/4 Tracings of lateral skull X-rays for normal and DS adults

2.3/5 Summary table of normal and DS cranial measurements

2.3/6 Summary table of normal and DS palatal measurements

2.3/7 Diagram of normal palatal contour and "steeple" palate

2.3/8 Summary of reported organic characteristics of DS and predicted voice quality settings

2.3/9 Summated protocol for control group

2.3/10 Summated protocol for DS group

2.3/11 Table of mean scalar degrees and standard deviations for DS and control groups

2.3/12 Graphic representation of significant differences between DS and control groups

2.3/13 Comparison of predicted voice quality findings and VPAS results for DS group

2.3/14 Pairwise calculations of Vocal Profile differences for DS/DS, control/control and DS/control subjects

2.4/1 Subject group information

2.4/2 Table of acoustic results

2.4/3 Table of acoustic results for three age bands, and statistical significance of age-related differences

2.5/1 Schematic representation of tissue layer disruption

2.5/2 Classification system for vocal fold pathology

2.5/3 Structural vocal fold pathologies arranged according to the classification system outlined in Figure 2.5/2

2.5/4 A summary of mechanical characteristics of vocal fold pathologies *and predicted acoustic consequences*

2.5/5 Laryngeal disorders diagnosed in pathological subject group

2.5/6 Subject group information

2.5/7 Percentage of subjects with acoustic values deviating from control group means by more than 2SD

2.5/8 Scattergram of F0 vs. S-DPF: male pathological subjects

2.5/9 Discrimination success of selected bivariate plots

2.5/10 Results of linear discriminant analysis

2.5/11 Comparison of three statistical procedures

2.5/12 Table of acoustic values for different pathologies

2.5/13 Average acoustic profiles in different pathologies: males

2.5/14 Average acoustic profiles in different pathologies: females

2.5/15 Acoustic profiles of speakers with unusually regular phonation

2.5/16 Graphic representation of the relationship between laryngeal tension and perturbation

2.5/17 Acoustic profile for a patient with Reinke's oedema

2.5/18 Acoustic profile for a patient with a sessile

vocal polyp

2.5/19 Acoustic profile for a patient with keratinization and hyperplasia

2.5/20 Acoustic profile for a patient with squamous carcinoma

2.5/21 Longitudinal study of a patient with squamous carcinoma

2.5/22 Longitudinal study of a patient with vocal nodules

ABSTRACT

This thesis examines the contribution of organic factors to voice quality. As background, the first part of the thesis examines the structure and properties of the tissues which make up the vocal apparatus, and discusses the growth patterns of these tissues in normal development, and in response to trauma and disease. The normal changes in the vocal apparatus which occur during the human life cycle are summarised.

The second part of the thesis focusses on some experimental investigations into the relationship between specific types of organic variation and voice quality. One problem in this field has been the lack of objective means of voice quality analysis, so that a subsidiary aim of the thesis has been the development and testing of appropriate voice assessment procedures. Two techniques for voice quality evaluation were used in this study; a perceptually based scheme and one using acoustic measurements. The development and use of these procedures are discussed, and examples of their application are discussed. Two main types of organic disorder are used to illustrate the links between measurable voice quality features and organic factors. The first of these, Down's Syndrome, involves a global disruption of growth and development, which results in a well documented set of physical anomalies. Voice quality findings for Down's Syndrome speakers seem to be clearly related to their organic features. The second class of disorder involves structural changes in the vocal folds, such as laryngeal cancer. Acoustic analysis of phonation in the presence of these disorders is discussed, with a view to the future development of acoustic voice analysis as a means of detecting vocal fold pathology.

ORGANIC VARIATION AND VOICE QUALITY

INTRODUCTION

The title of this thesis immediately raises two questions which are basic to the whole work. Firstly, what, within the framework of this thesis, is meant by "voice quality"? Secondly, what is meant by organic variation? Before going any further, it is therefore necessary to lay out some working definitions which will answer these questions.

Voice quality is given a broader definition than is often used. To quote Laver (1980:1), it is "the characteristic auditory colouring of an individual speaker's voice". This involves much more than just the activities of the larynx. The habitual posture of a speaker's lips or tongue may be just as important as her habitual phonation in making her voice characteristic and immediately recognizable to any acquaintances. The whole of the vocal apparatus is therefore seen as contributing to voice quality.

The word "organic" will be used to describe any factors which are to do with anatomical structure, and with the constraints imposed by that structure and its mechanical properties on the capability for physiological action. Organic variation, therefore, refers to any anatomical feature which may differ from one individual to another. Such variation ranges from relatively minor differences between individuals, such as details of dentition, to gross distortions of the vocal tract such as may occur in cleft palate or laryngeal cancer. Since this study concerns the relationship between such organic variations and voice quality, the primary area of interest is obviously the vocal apparatus.

In speech, an important distinction can be drawn between underlying organic factors and phonetic factors. This distinction follows work by Laver (1980:9). An individual is endowed with a characteristic organic make-up, and, short of surgery or some other medical intervention, there is little that can be done to change the situation. The potential range of speech output will always be constrained by an individual's organic state. Some features of speech may be clearly identifiable as having a specific organic basis, and as such they will be outside the speaker's control. Phonetic features of speech, on the other hand, are those features which are under the speaker's control. They are due to voluntary adjustments, i.e. volitional, learned actions (which are not necessarily under conscious control), of the vocal apparatus musculature.

MOTIVATION

The title is a very wide one, and it is important to quell any expectations that a thesis of this sort will be able to do more than bring some initial structure to what is a very little explored field. Until recently, phonetics has understandably been mainly concerned with the development of a general phonetic theory which allows a unified system of analysis, applicable to the linguistically relevant speech output of the whole human race. This has required phoneticians to assume that all speakers possess more or less similar vocal apparatuses, and that differing speech sounds are mainly due to various adjustments of vocal tract posture.

It is, of course, clear that the detail of an individual's speech patterns, both at the segmental level and at the longer term 'voice quality' level, will be influenced by his or her vocal anatomy. A simple example may be used to illustrate this, comparing two speakers

who differ in details of alveolar and palatal contour. Speaker A has a rather broad alveolar ridge with a low, flattened palatal arch which does not afford much volume in the front of the oral cavity. Speaker B has a narrow alveolar ridge and a very high, arched palate, which gives her a much larger oral volume. Assuming that the two speakers have similar jaw and tongue relationships, they are likely to display all sorts of minor differences in speech output which are directly attributable to their vocal anatomies. For example, the trajectory of tongue movement needed to move from an alveolar segment to one involving some degree of palatal approximation, as in the English word deed, will be different in two such speakers, and this is likely to be reflected in the acoustic detail of the transitions. In terms of voice quality, unless some compensatory adjustments are made, speaker A, who has a smaller palatal volume, will tend to have a greater degree of constriction in the front of the oral cavity throughout speech.

The motivation for beginning an examination of the relationship between organic factors and voice quality may stem from several sources. Foremost of these, as far as this thesis is concerned, is the importance of this relationship at the interface between medicine, speech therapy and phonetics. Whilst even the healthy population shows a considerable amount of organic variation of the vocal apparatus, the amount of variation seen in populations attending speech therapy or Ear, Nose and Throat clinics is greatly increased. The assessment of speech output by speech therapists relies heavily on general phonetic techniques, but the paucity of phonetic research into the relationship between non-standard vocal tract anatomy and phonetic output inevitably leads to problems in applying phonetic assessment techniques to the clinical population. Some studies have begun to address this problem, but most have taken a segmentally-

oriented approach (e.g. Vieregge 1981). Since organic factors are relatively invariable, and will tend to exert their influence throughout speech, and not just on individual segments, it seems logical to look at their effect on voice quality as a long-term ingredient of speech. The segmental effects will, of course, contribute to voice quality assessment in a way which will be explained more fully in Section 2.1.2.

Further motivation for examining the effects of normal organic variation on voice quality may stem from the increasing interest in acoustic systems for speaker recognition and speech analysis. An understanding of the ways in which interspeaker and transient intraspeaker differences in organic state may affect speech output may be highly relevant in this area, but a full discussion of the implications of organic variation for speech technology is beyond the scope of this thesis.

THESIS AIMS

The overall aim of this thesis is to develop a soundly-based account of the relationship between growth of the vocal apparatus and voice quality. To do this, it will be necessary first to organise and integrate relevant aspects of the medical literature on the materials which make up the vocal tract, their arrangement, and the ways in which they grow and develop throughout life. It is hoped that Part One of the thesis will make this information available in a way which is helpful to phoneticians and speech therapists. There is no shortage of detailed decriptions of vocal apparatus structure (e.g Kaplan 1960, Hardcastle 1976, Laver 1980, Dickson & Maue-Dickson 1982), but it is often difficult to extract information about the role of specific tissue types, or the growth relationships between different parts of the vocal tract. Since such information is basic to a proper

understanding of the kind of organic variation which may influence speech output, there is a strong argument for attempting to collect such information together and to present it in a digestible form. Inevitably there may be areas where this section will fall short of providing all the relevant information, because rapid developments in medical research make it difficult for any account to be fully up to date. It is hoped, however, that this section will at least offer a useful background to the specific types of organic variation considered in Part Two of the thesis.

The second part of the thesis aims to examine the relationship between specific types of organic variation and voice quality experimentally. One problem in this field has been the lack of objective means of voice quality analysis. A subsidiary aim of this work has therefore been the development of appropriate voice assessment procedures. Two techniques for voice quality assessment will be discussed here. The first of these is a perceptually based system (the Vocal Profile Analysis Scheme), and the second is an acoustic system for the analysis of phonatory characteristics. The theoretical bases and the practical procedures for using these analysis systems will be presented. Finally, the experimental application of perceptual and acoustic analyses to groups of speakers with organic abnormalities affecting the vocal apparatus, and to normal control groups, will be described. Links will then be drawn between voice quality findings and underlying organic factors.

PART ONE: BUILDING THE VOCAL APPARATUS

1.1. STRUCTURAL COMPONENTS

The aim of this section is to describe the major structural components or building materials which make up the vocal apparatus. The cells and tissues involved each have their own characteristic structures, biological capabilities and mechanical properties which suit them for their various roles within the vocal apparatus. An understanding of these properties allows a better prediction of the consequences for speech of the kinds of abnormality in tissue relationships which appear in many individuals at some stage in their lives.

1.1.1 TOPOGRAPHICAL ORIENTATION

Although the emphasis of this section is on the basic building materials which make up the vocal apparatus rather than on the overall structure of the apparatus, there will be some instances where it is useful to illustrate the function of a particular tissue by commenting on its geographical distribution. It may therefore be useful to provide some basic diagrams of the vocal apparatus at this stage. These will serve two purposes. Firstly, they can act as simple maps, showing the major topographical details of the vocal apparatus. Secondly, they may help to clarify some of the anatomical terminology which will be used in this thesis. Authors do unfortunately vary somewhat in their choice of anatomical labels, so it is as well to present at this stage some of the more important labels which will be used throughout the thesis.

Full accounts of the anatomy and physiology of the vocal apparatus can be found elsewhere (Kaplan 1960, Hardcastle 1976, Romanes 1978, Laver 1980, Dickson and Maue-Dickson

1982). Most of the anatomical diagrams presented in this thesis, together with some of the diagrams in section 1.1.2, were originally prepared by the author for a text book on the anatomy and physiology of speech written by Laver and Mackenzie Beck (forthcoming), which includes a very complete description of the vocal apparatus and its constituent parts.

Figure 1.1.1/1 introduces the standard anatomical terminology, used to describe the planes of the human body. The transverse plane divides the body, or some part of the body, across its longitudinal axis. The sagittal plane divides the body along its length, into right and left parts. This can be remembered because a cut along the sagittal plane would be parallel to the sagittal suture of the skull. A vertical cut at right angles to the sagittal plane, which divides the body into front and back parts, is the coronal plane. In this case, the coronal suture of the skull acts as a reminder.

Other anatomical terminology used in this thesis should be fairly self explanatory, but the following definitions may help to avoid any confusion.

Superior = above
Inferior = below
Posterior = towards the back of the body
Anterior = towards the front of the body
Lateral = away from the midsagittal plane
Medial = towards the midsagittal plane
Superficial = towards the surface of the body
Deep = away from the surface of the body

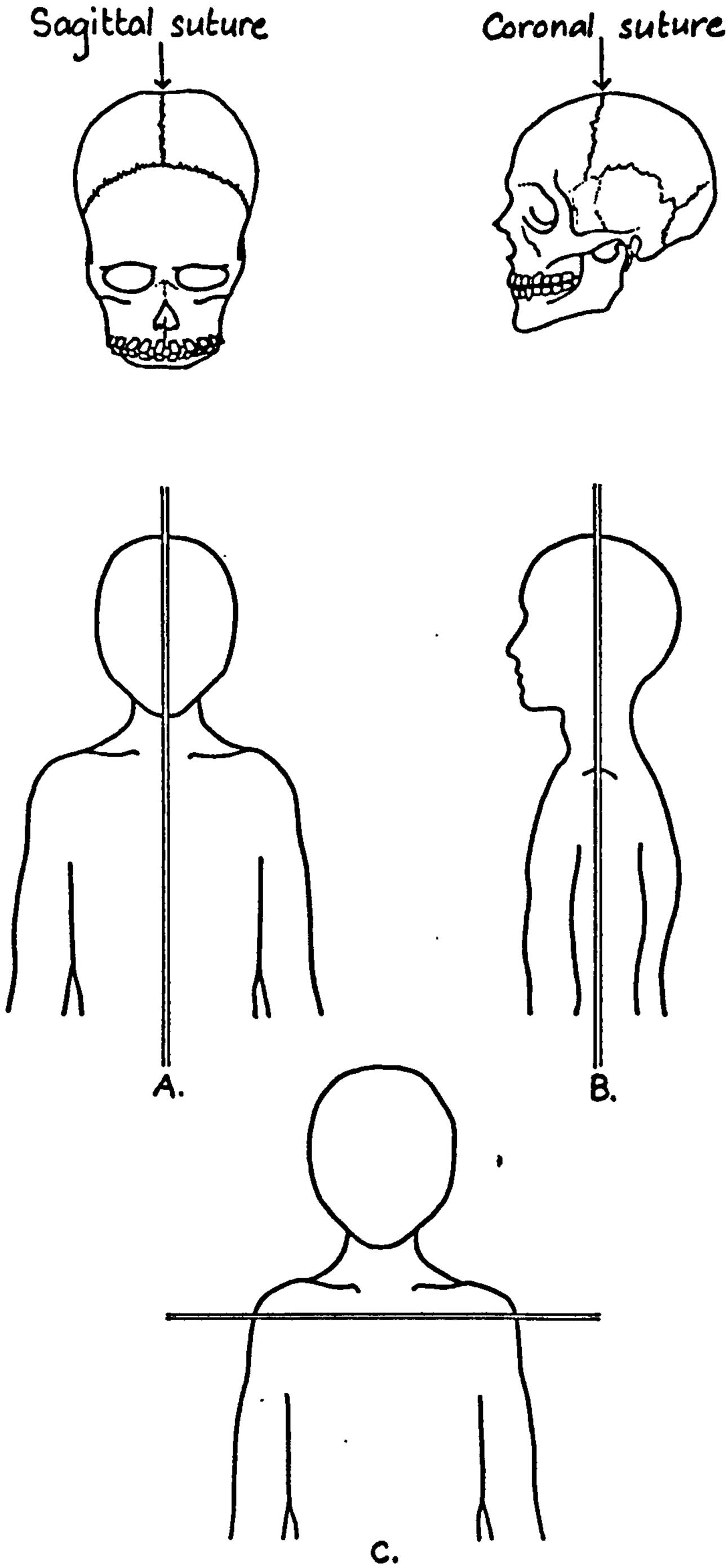
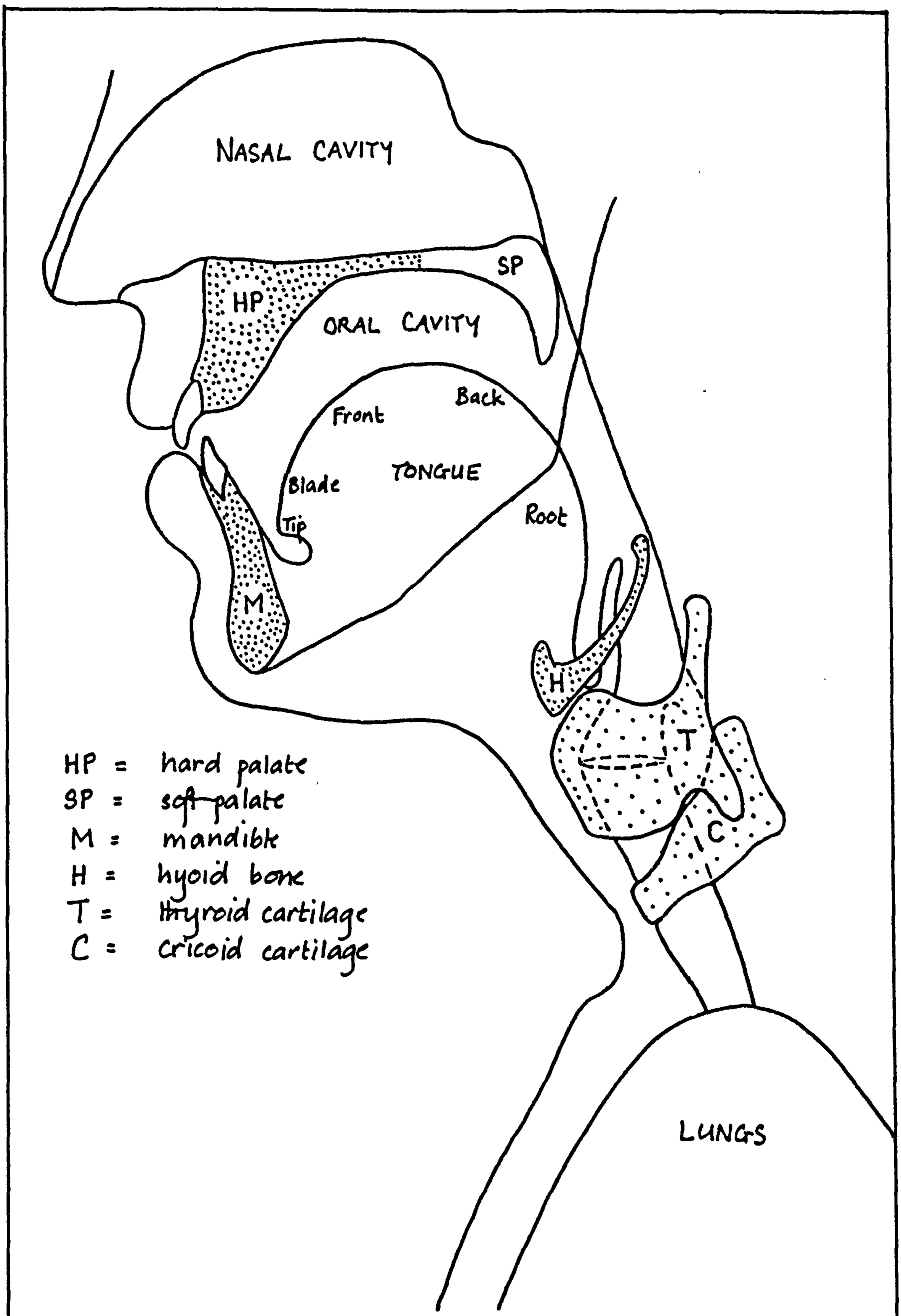


FIGURE 1.1.1/1: Illustration of anatomical planes.
A. Sagittal plane, B. Coronal plane
C. Transverse plane



HP = hard palate
SP = soft palate
M = mandible
H = hyoid bone
T = thyroid cartilage
C = cricoid cartilage

FIGURE 1.1.1/2: Schematic diagram of the vocal apparatus

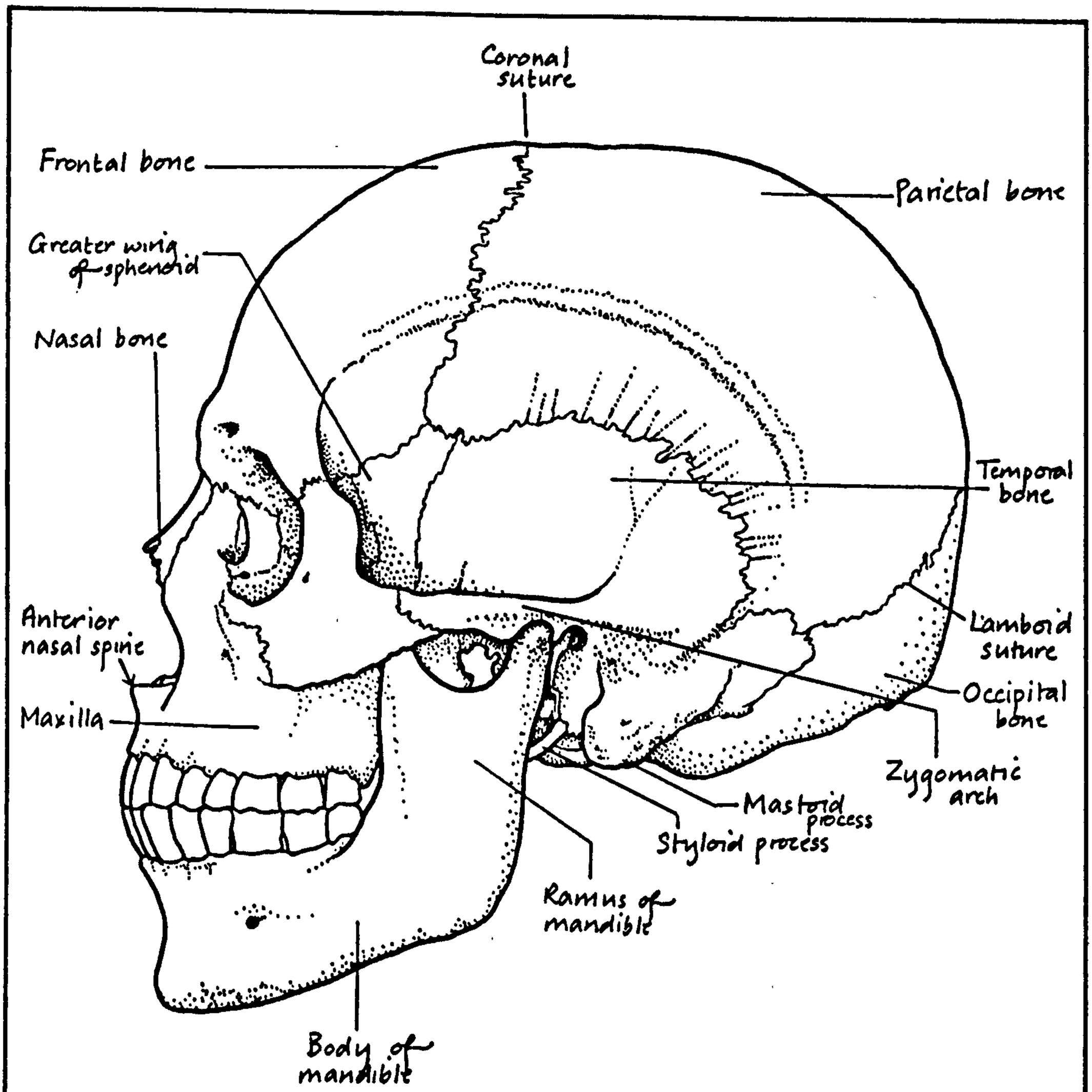


FIGURE 1.1.1/3: Lateral view of the skull (adapted from Romanes 1978:265)

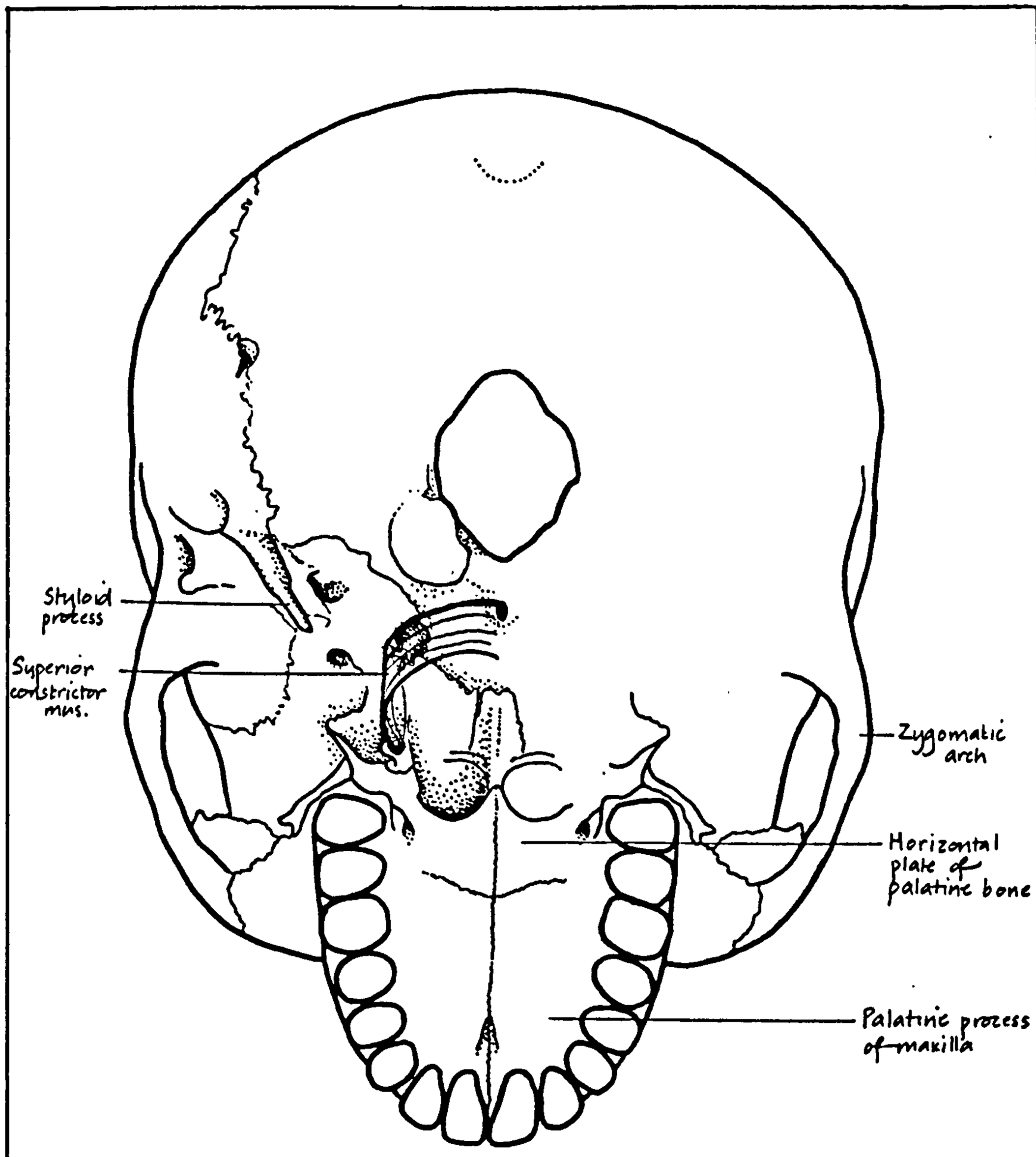


FIGURE 1.1.1/4: External surface of the base of the skull, showing the position of the superior constrictor muscle

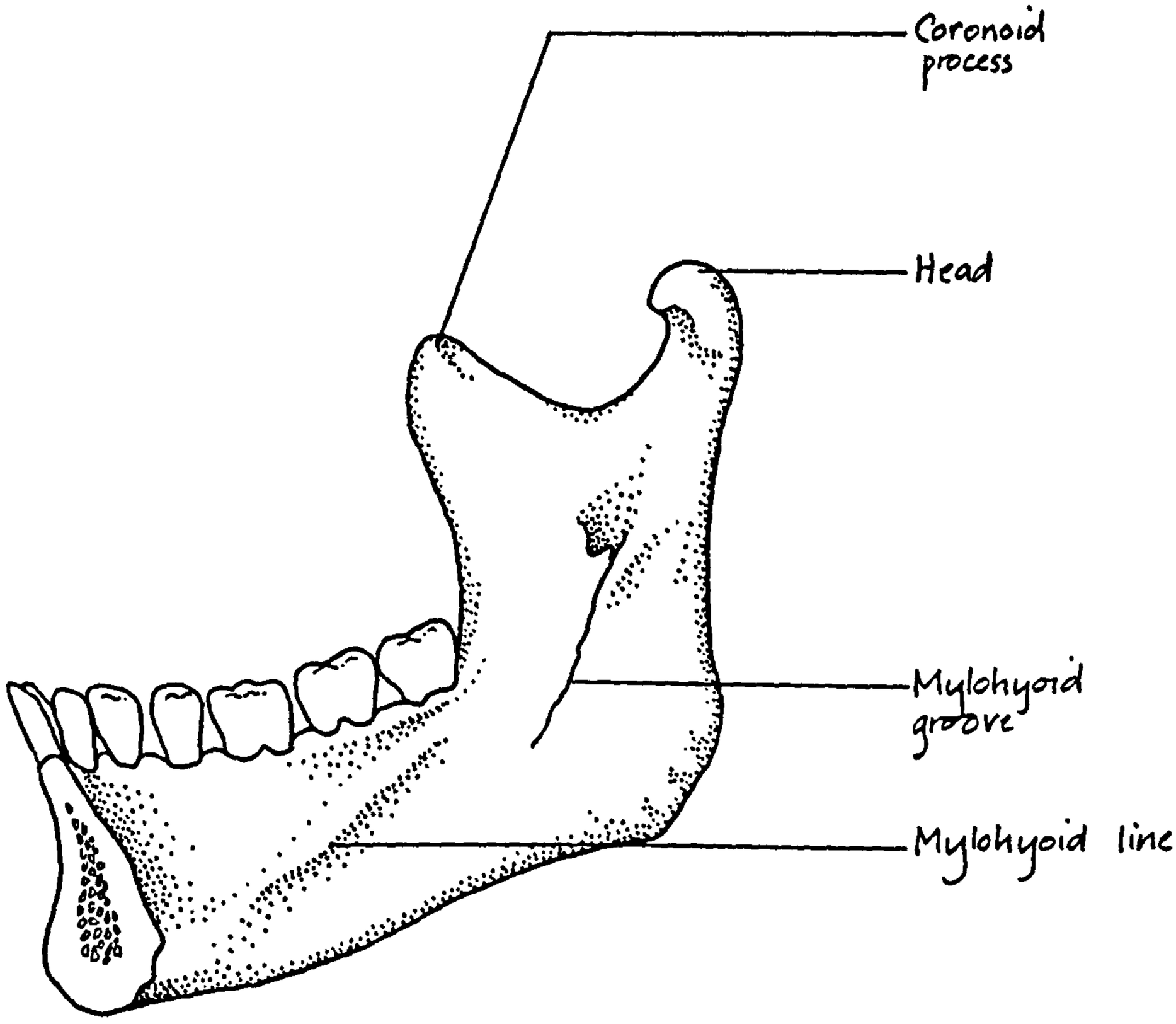


FIGURE 1.1.1/5: Medial view of the mandible (adapted from Davies and Davies 1962:327)

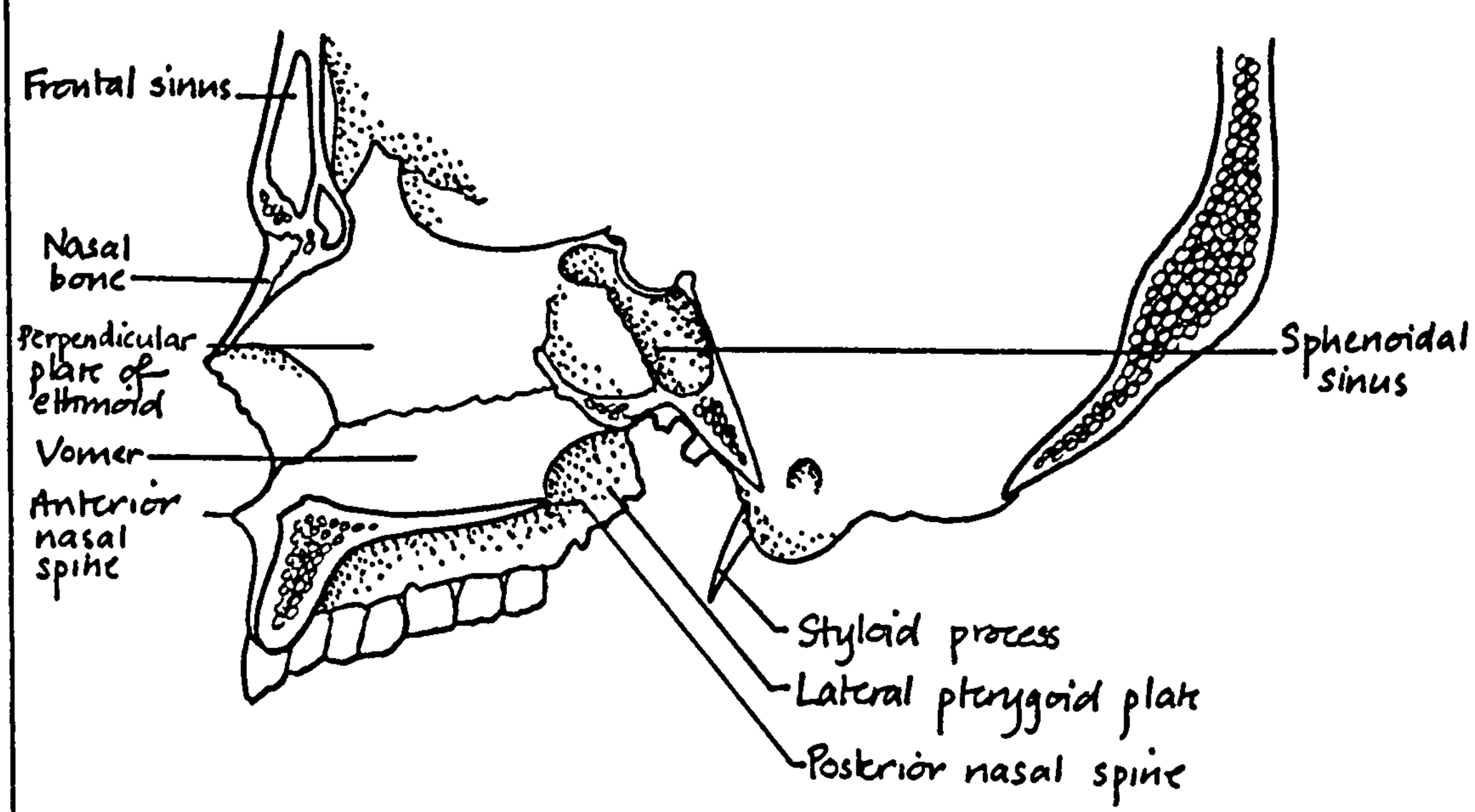


FIGURE 1.1.1/6: Sagittal section through the lower part of the skull (adapted from Davies and Davies 1962:324)

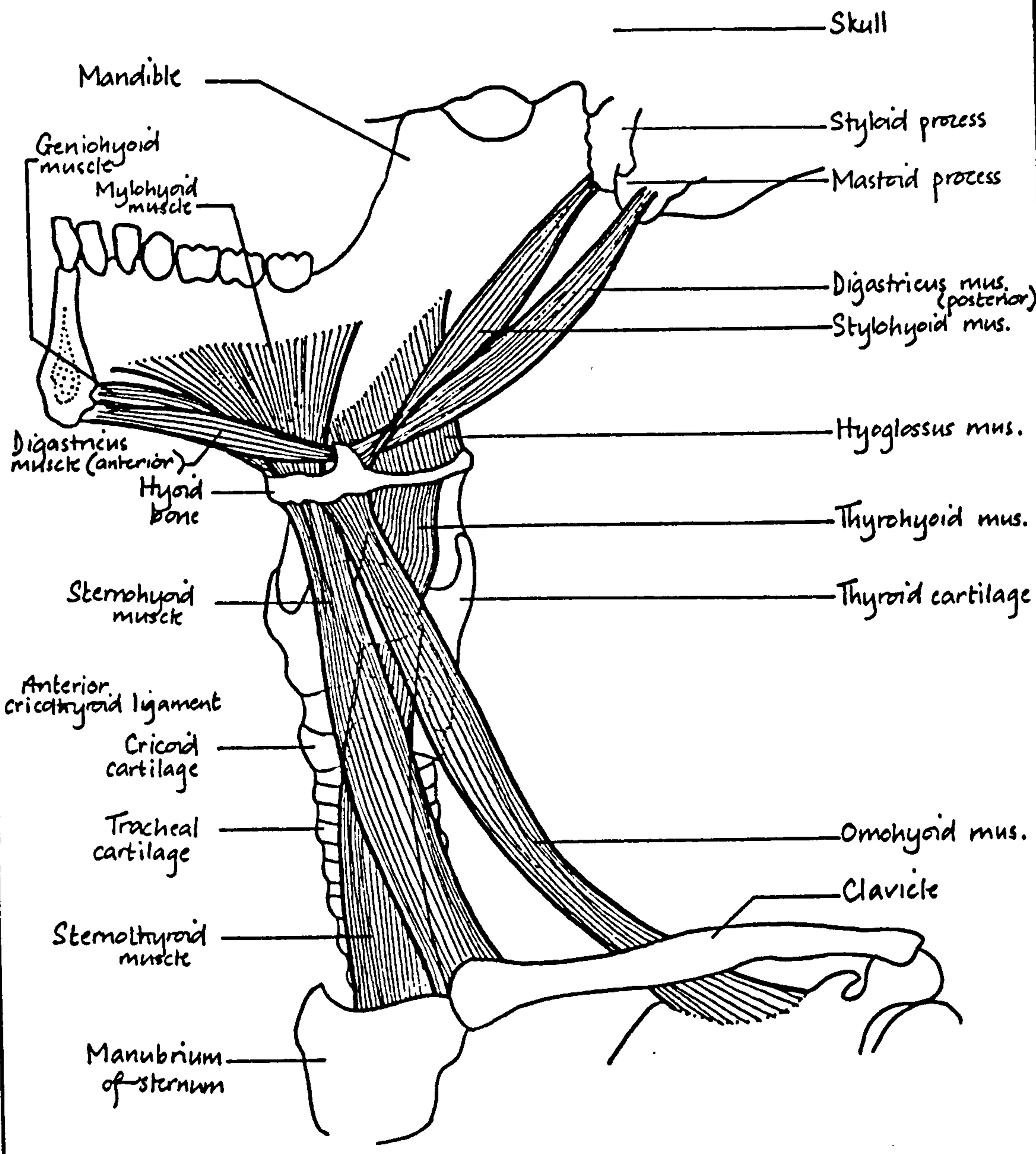


FIGURE 1.1.1/7: Schematic diagram showing suspension of the hyoid bone

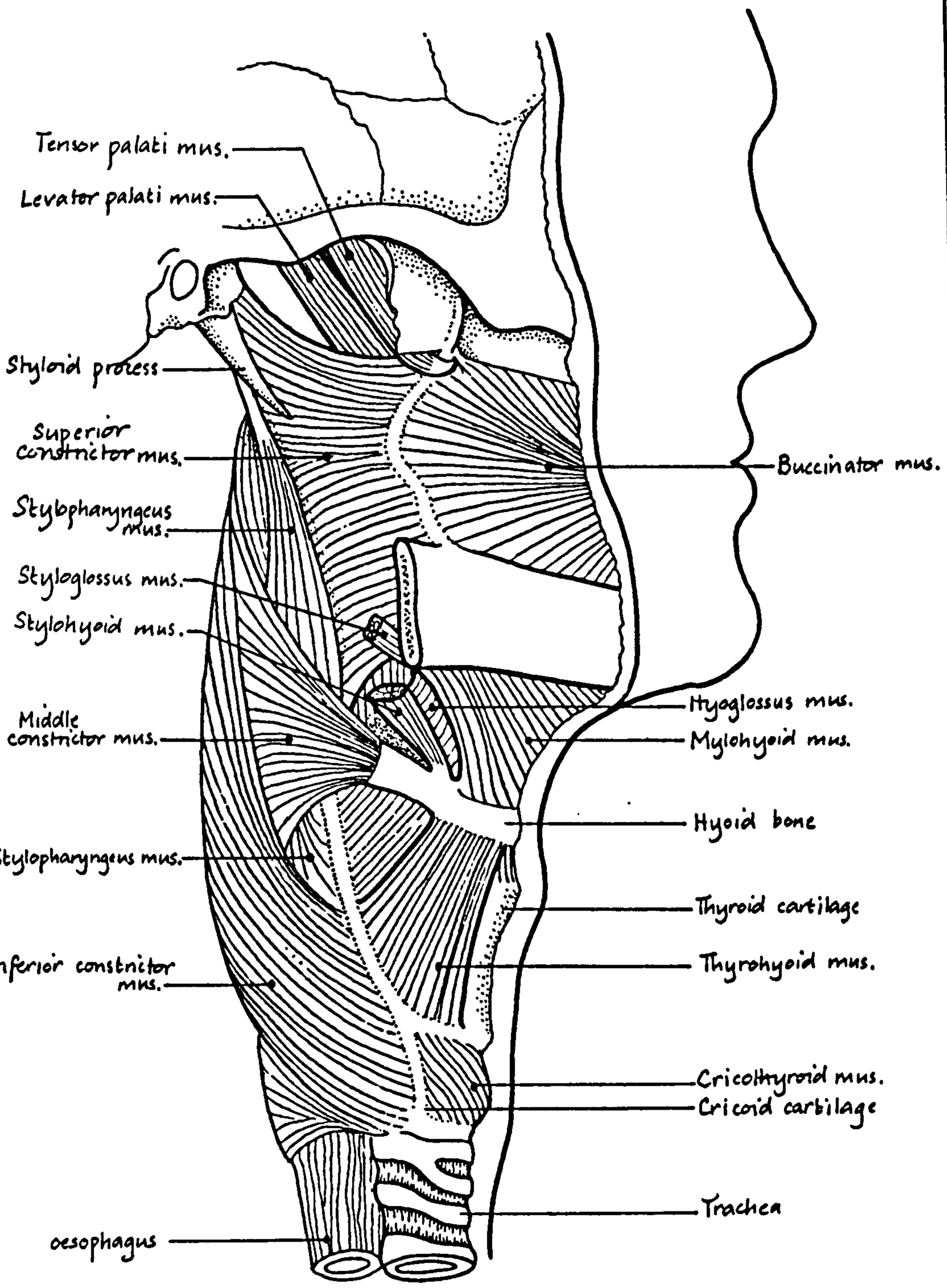


FIGURE 1.1.1/8: Lateral view of the constrictors of the pharynx and associated muscles (adapted from Romanes 1978:114)

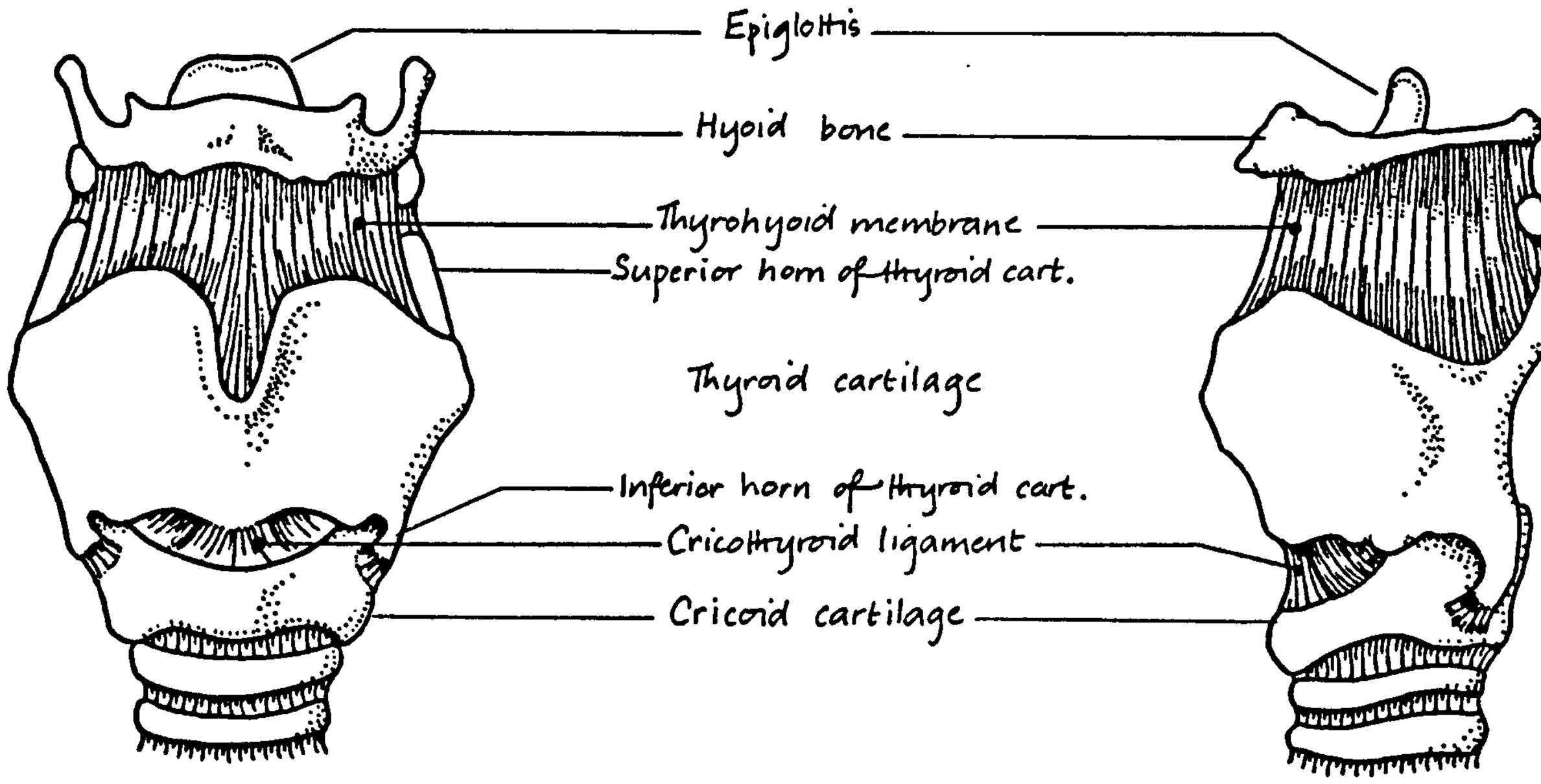


FIGURE 1.1.1/9: The cartilages and ligaments of the larynx (adapted from Romanes 1978:130,131)
A. Anterior view, B. Profile view

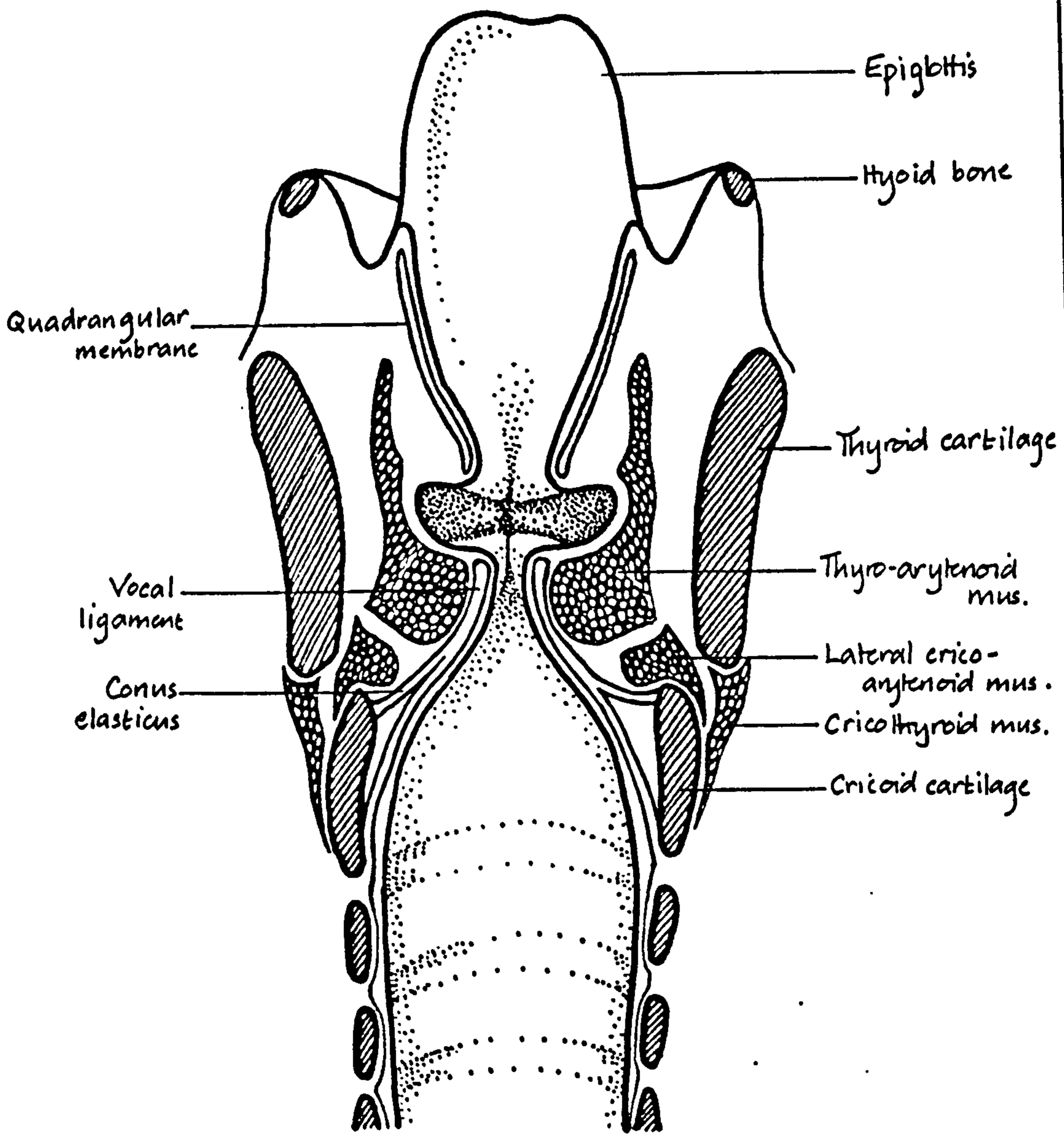


FIGURE 1.1.1/10: Coronal section of the larynx
(adapted from Romanes 1978:133)

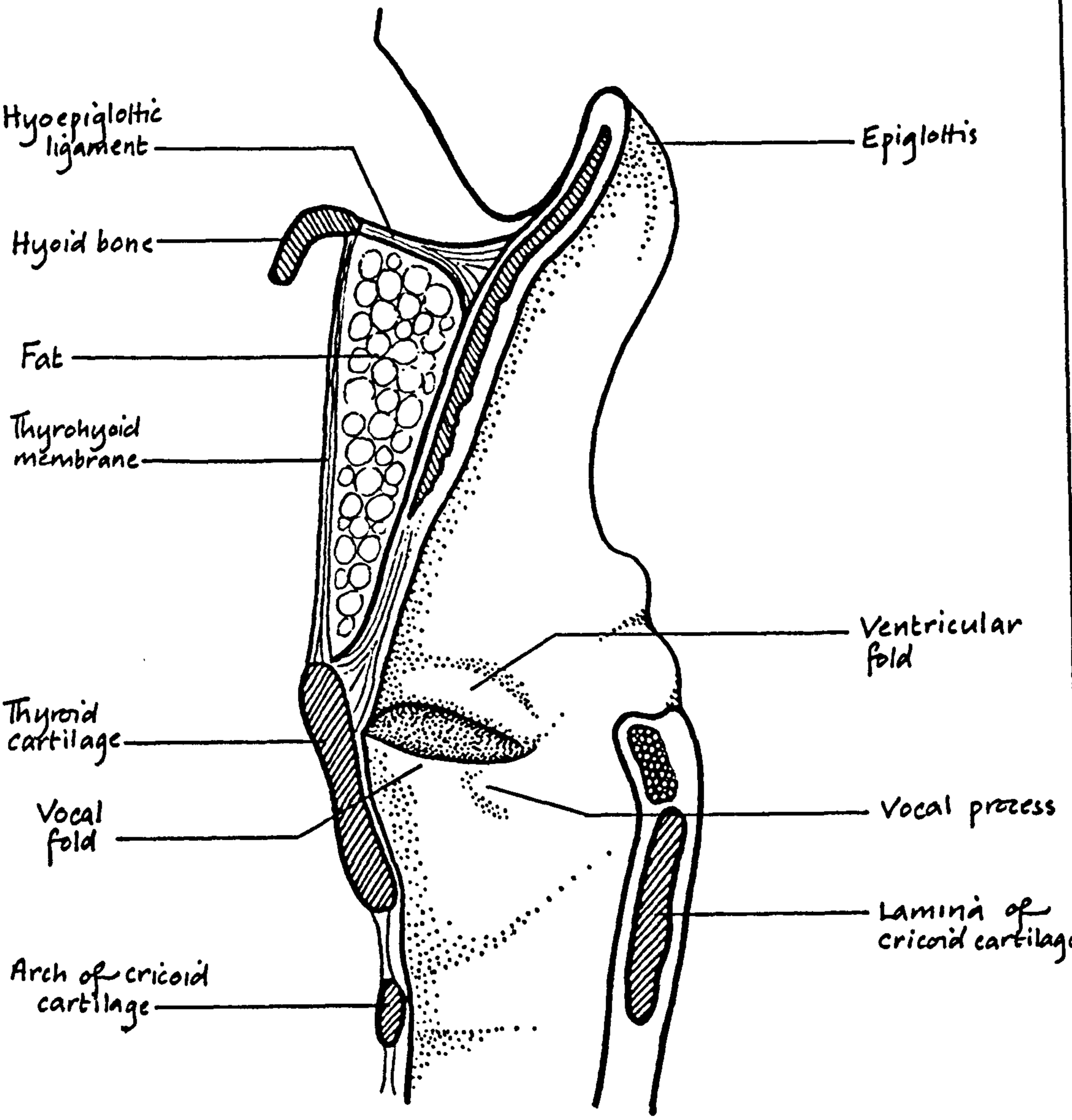


FIGURE 1.1.1/11: Median section of the larynx (adapted from Romanes 1978:132)

1.1.2 BASIC BUILDING BLOCKS

Later sections on growth, and on development of the vocal apparatus, make frequent reference to various types of cell and tissue. Some preliminary discussion of these basic building materials is therefore needed, in order to define a vocabulary for the following chapters. The aim of this section is to present some basic histology (histology is the study of tissue structure), with particular reference to those tissue types which act as major structural components in the vocal apparatus.

Much of the information presented is a synthesis of relevant material from many basic texts on human anatomy and histology. Rather than making the text too unwieldy by repeated reference to the same works, I shall begin by acknowledging my debt to those authors who have contributed a broad range of background information. Amongst these are Clegg and Clegg (1963), Freeman and Bracegirdle (1967), Bloom and Fawcett (1968), Leeson and Leeson (1976) and Junqueira and Carneiro (1980). Some of the material is also summarised in Dickson and Maue-Dickson (1982). Further reference will be made in the text to authors who have made specific contributions to any topic.

Cells

Most anatomy and histology texts start with a description of a typical animal cell, and take this as the basic structural and functional unit of the body. Since most of the properties of cells will be important in later descriptions of tissue growth, I will follow this precedent. Two things need to be remembered, however. One is that much of the bulk of the body is made up of material which is not contained in cells. This may consist of fibres of various sorts, or of various types

of ground substance, which vary widely in make up and consistency. The other is that the "typical" cell does not resemble very closely the cells which are actually found in many real tissues. At the start of embryonic development, each cell has the potential for all the essential activities of life; respiration, assimilation of nutrients, excretion of waste products, growth, manufacture and secretion of various substances, response to stimuli, and reproduction. During development some cells become specialized in one or more of these activities at the expense of others, and the cell structure may be dramatically altered as it becomes adapted to its special function.

Most cells do, none the less, share some common structural features. The cell can be visualised as a bag of fluid or semi-fluid substance (cytoplasm), within which are the structures necessary to perform the various cellular activities.

Figure 1.1.2/1 is a schematic diagram of a generalized animal cell, showing the main features which can be seen using an electron microscope. The material of the bag which encloses the whole structure is the cell membrane, but it is more than a mere container. It plays an active role in controlling the transport of substances in and out of the cell, and it acts as a kind of biochemical magnet, to capture substances which are important for a particular cell. For example, cells which need to respond to the presence of growth hormone by becoming more active have special receptors on the cell membrane which bind passing molecules of growth hormone to the cell. The membrane is also involved in binding cells together in certain types of tissue, and in the normal limitation of growth.

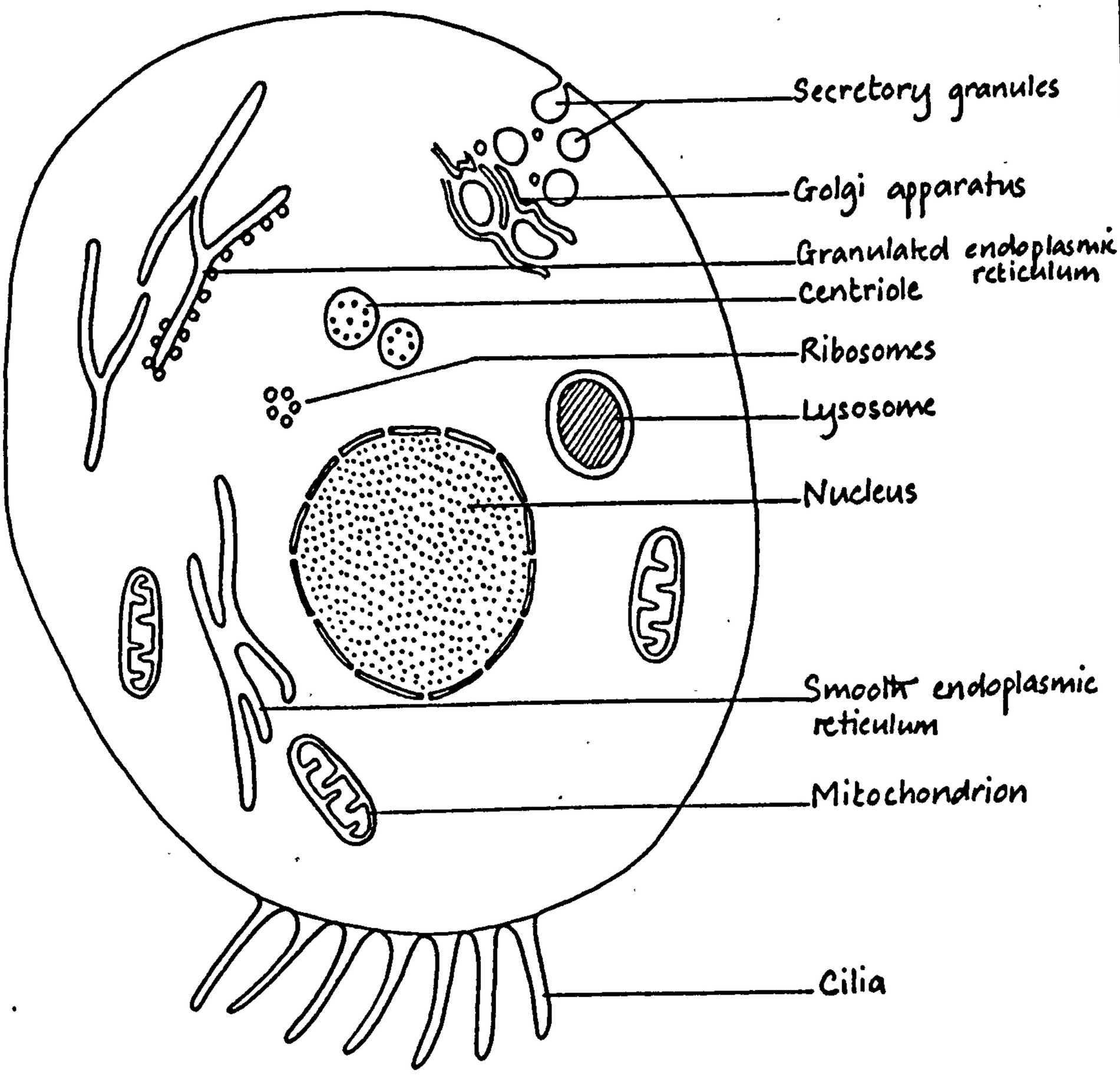


FIGURE 1.1.2/1: Schematic representation of a generalized animal cell

The Nucleus

Within the cell, the most conspicuous structure is usually the nucleus, which is enclosed within the nuclear membrane. Most cells have a single nucleus, although a few exceptional cells, such as muscle cells (see below), have several. The nucleus contains the bulk of the genetic material, in the form of chromosomes. These are so thin and elongated that they are not easily seen, except during cell division, when they become shorter and thicker (see section 1.2.1.b).

With the exception of cells involved in reproduction, all normal human cells have 23 pairs of chromosomes in each nucleus. In these are coded all the instructions necessary for proper functioning of the cell, and hence of the whole body. Each chromosome consists of a long length of DNA (deoxyribonucleic acid), which can be thought of as a long list of separate instructions. The lengths of DNA corresponding to each of these instructions are known as genes. With the exception of certain controlling genes, each gene is responsible for the formation of one protein, which consists of a string of amino acids. The information carried by a gene governs the identity and the ordering of the amino acids which make up a protein. The characteristics which we think of as being inherited via our genes, such as red hair or short stature, are simply the gross consequences of the presence or absence of particular proteins. The genetic control of individual growth and morphology will be particularly relevant in Section 2.3 on voice quality in Down's Syndrome.

The simplest description of a DNA molecule is of a spiralling ladder, where the rungs may be of 4 possible types. Each group of 3 adjacent rungs corresponds to one of the 20 possible amino acids. When a gene is activated the specified set of amino acids (usually a few hundred)

will be assembled in the correct order outside the nucleus. The information is relayed from the nucleus by messenger molecules of RNA (ribonucleic acid). They have a similar structure to DNA, and use the DNA as a template to copy the correct sequence of rungs.

The full complexity of the genetic system is still not fully understood, but it is clear that there are some sections of DNA which are concerned with controlling the activity of other genes rather than with the manufacture of proteins. It should be stressed that every cell in the body, with the exception of reproductive cells, contains identical genetic information. It is the selective activity of genes within each cell which differentiates the huge variety of cell types within the body. The other main structures within the cell are summarised below.

Mitochondria These are the centres of respiration, converting energy into a useable form, and are hence commonly called the "power-packs" of the cell.

Ribosomes These are responsible for protein manufacture, in collaboration with messenger molecules of RNA, which carry information from genes within the nucleus.

Endoplasmic reticulum This membranous stucture is the site of protein manufacture. It may appear smooth or granulated, depending on whether or not it has ribosomes associated with it.

Golgi body (or golgi apparatus) This is another membranous structure, which is most clearly developed in secretory cells, and seems to be involved in the accumulation of substances which are to be secreted.

Centrioles There are two centrioles in each cell, which lie together, and are involved in cell division.

Lysosomes These are small envelopes of digestive enzymes which break down particles within the cell.

Microtubules These thin, straight tubular structures, made of the protein tubulin, give some rigidity to the

cell and help to maintain cell shape. They are also involved in the movement of cilia and flagella (see below), and aid the transport of water and other substances within the cell.

Cilia and flagella Cilia are tiny hair-like protrusions from the cell surface, which display rapid beating movements interspersed by a slower recovery of their original position. Some types of epithelium are covered with cilia, which bend in synchrony to produce a wave-like movement, thus moving a surface film of fluid or mucus. Flagella are essentially similar, but are longer and occur singly, as in the tail of a spermatazoan.

Microfilaments Microfilaments are often grouped into parallel bundles (fibrils) and are found in most cell types. Together with microtubules they seem to act as an intercellular skeleton, giving the cell some rigidity, resilience and tensile strength. Some microfilaments are capable of contraction and are similar to the highly specialised filaments found in muscle tissues (see under muscle, below). In keratinizing epithelium microfilaments play a part in keratin formation. Keratin is the horny material which characterizes the external skin (see the section on epithelium below).

Cells may also contain inclusions such as fat droplets, pigment, and glycogen (a form of carbohydrate) granules.

The cell is not, of course, a static structure, but a dynamic system in which there is constant recycling and renewing of many constituents. In addition to constant turnover and movement at the molecular level, there are larger scale exchanges of material amongst the structures of the cell. An illustration of this is the transfer of membrane from the endoplasmic reticulum to the golgi body, and hence to the cell surface, associated with the secretion of protein (Leeson and Leeson 1976:33-36). Membrane formation is a continuous process within the

cell, and membranes may move from one site to another, changing in structure and function as they go. In secretion of protein, the protein is manufactured at the endoplasmic reticulum, sealed in an envelope of membrane (= a transfer vesicle) and passed to the golgi body. The vesicle membrane fuses with the membrane of the golgi body, and the protein is condensed and modified as it passes through the body. It is then repackaged in a membrane envelope (= a condensing vacuole) for transfer to the cell surface. The membrane of the condensing vacuole finally fuses with the outer cell membrane, releasing the contents from the cell. Many such processes may be active in any one cell, depending on its specialization, and the dynamics of cell function and of intercellular relationships are enormously complex.

Tissues

It has already been said that cells, as they develop, may become differentiated and specialized in structure and function. Cells with the same sort of speciality tend to be organised into one tissue type. A tissue is a collection of similar cells, together with varying amounts and types of intercellular substance. The functional and mechanical properties of a tissue depend, therefore, partly on the cells themselves, and partly on any intercellular substances which may be present.

Tissues are normally classed into four main types.

A. Epithelial tissue. This forms thin sheets of cells, which cover all internal and external surfaces of the body. Epithelium may also become folded in on itself, and develop into glands. Less commonly, epithelium may take on a sensory function, as in the cochlea of the ear (Dickson and Maue-Dickson 1982:16).

B. Connective tissue. This group of tissues includes the various forms of bone and cartilage, which form the skeletal framework of the body. Other types of connective tissue act as structural coordinators, binding organs, muscles and nerves to each other, and to the skeleton. Transport of many substances round the body is also the task of two specific types of connective tissue; lymph and blood.

C. Muscle tissue. Muscle tissue is a highly differentiated type of tissue, specialized for contraction.

D. Nervous tissue. This is specialized to be able to transmit electrochemical impulses.

Each of these tissue classes can be further subdivided, and the human body contains an enormous variety of tissue types. Only tissues which play significant structural roles in the vocal apparatus will be described below.

A. Epithelial tissue

Covering epithelium

Covering epithelium, i.e. the sheets of tissue which cover the surfaces of the body, has very little in the way of intercellular substance. The cells are very tightly packed together, separated only by a thin layer of intercellular cement. They lie on a non-cellular basement membrane, which may be derived from the underlying tissue.

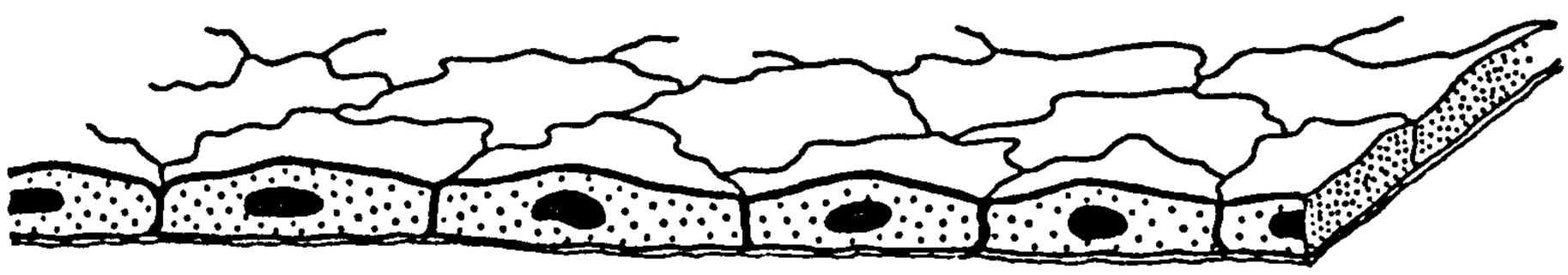
Epithelium of different sorts lines the whole of the vocal apparatus, from the lungs to the lips and nose. It is interrupted only by the teeth. The importance of this layer to speech may be out of all proportion to its small

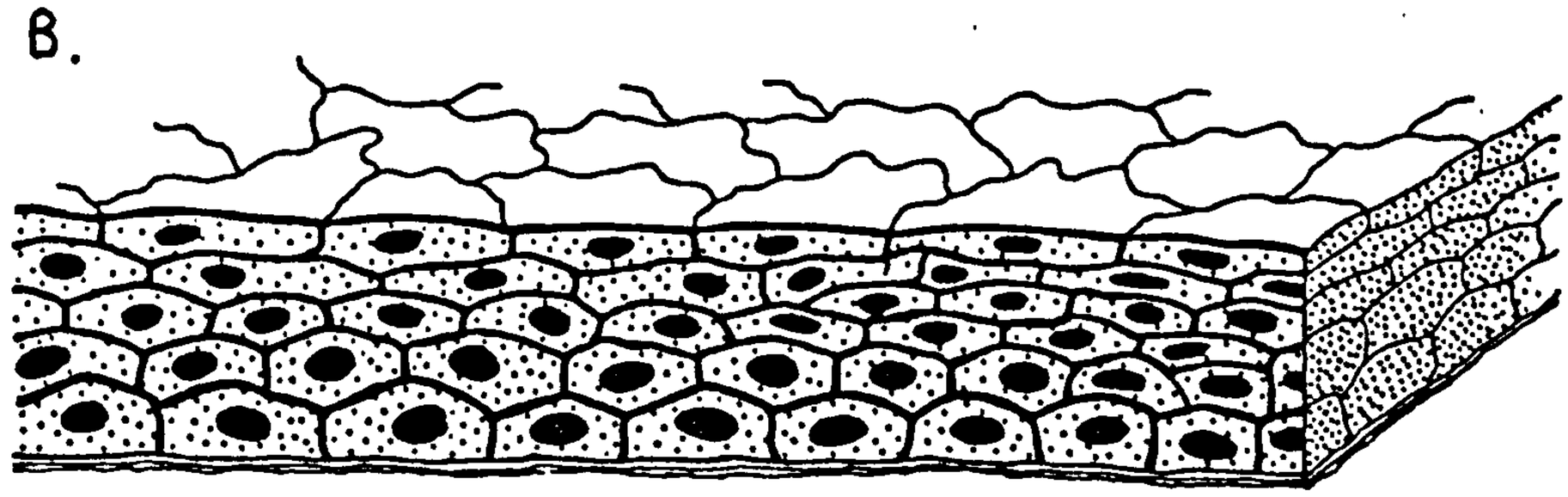
thickness. In the larynx, the state of the epithelium may have a profound influence on vocal fold vibration (see Sections 2.3 and 2.5). Throughout the rest of the vocal tract, the epithelium may well affect the extent to which acoustic energy is absorbed by the resonating cavities.

Covering epithelium is classified according to cell shape, the number of cell layers, and the nature of the free surface (see Figure 1.1.2/2). Cells may be cuboidal, columnar or squamous in shape. Cuboidal cells are approximately isodiametric in shape, columnar cells are taller than they are wide, and squamous cells are flattened, so that they take on the appearance of somewhat irregular paving stones.

Epithelium which is only one cell in thickness is known as simple epithelium, and is found in situations where the epithelium is not subjected to mechanical stress, or where absorption of nutrients or gases must take place. Simple epithelium is found, for example, in the innermost parts of the lung. Where epithelial tissue is more than one cell thick, the cells tend to be arranged in fairly orderly layers, and the tissue is described as *stratified* epithelium. *Stratified* epithelium is generally found in places where mechanical trauma is a particular problem. It is found, for example, covering the free borders of the vocal folds, where the folds make contact during adduction for phonation. Occasionally the tissue may appear to be more than one cell in thickness, but a closer examination shows that all the cells are, in fact, resting on the basement membrane. This is known as pseudostratified epithelium.

The free surface of the epithelium may be smooth and unadorned, or it may be covered with small thread-like projections called cilia (= ciliated epithelium). Cilia are able to move in a rhythmical, wave-like manner, and





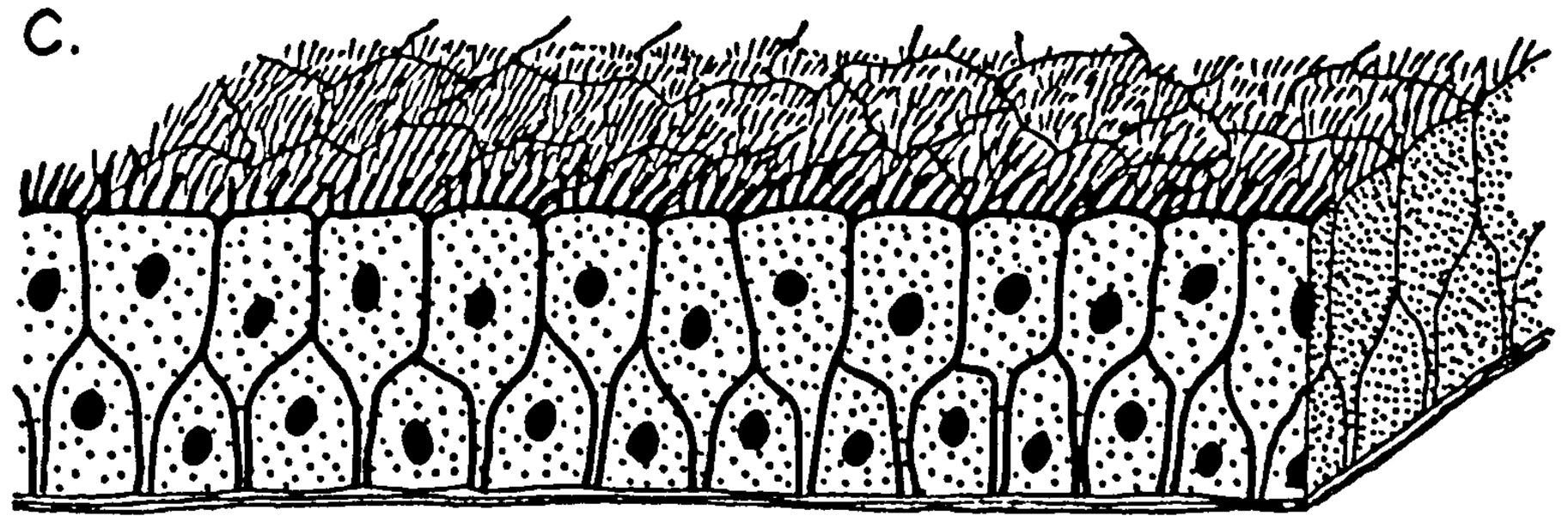


FIGURE 1.1.2/2: Schematic representaion of the principal types of epithelium found lining the vocal apparatus
A. Simple squamous epithelium
B. Stratified squamous epithelium
C. Pseudostratified ciliated columnar epithelium

they are found in areas such as the trachea, where they trap dust and secretions in a surface film of mucus, and help to move them away from the lungs. The beating of cilia on an area of ciliated epithelium is rather like the movement of long grass in a field as it is blown by the wind.

In the external skin, which has to withstand a considerable amount of trauma, the free surface is protected by a layer of horny protein, keratin. This is produced from the superficial cell layers, which lose their nuclei and are largely converted into keratin. Normally, keratin is found only at the outer limits of the vocal tract, at the lips and nares, but abnormal deposition of keratin may occur elsewhere within the vocal apparatus in some pathological conditions (see section 2.5).

Only a few of the various types of covering epithelium which are found in the body need to be discussed in relation to the vocal apparatus. These are illustrated in Figure 1.1.2/2.

i) Keratinized stratified squamous epithelium.

ii) Non-keratinizing stratified squamous epithelium.
This is found in the oral cavity, the oro-pharynx, the laryngo-pharynx, and on the free borders of the vocal folds.

iii) Pseudostratified ciliated columnar epithelium and

iv) Ciliated columnar epithelium.
One or other of these two types is found in most of the sections of the vocal apparatus which are primarily to do with respiratory function, and which do not also constitute part of the digestive tract; i.e. the nasal

cavity and most of the respiratory pathway between the epiglottis and the bronchioles of the lung.

v) Cuboidal epithelium.
This is found in a small transitional area where the ciliated columnar epithelium of the naso-pharynx changes to the stratified squamous epithelium of the oro-pharynx.

vi) Simple squamous epithelium.
This is found lining the inner airways of the lung.

Glandular epithelium

The last major class of epithelial tissue is concerned with the secretion of substances such as hormones, enzymes and mucin. Some epithelial cells are highly specialized for manufacture and secretion of these substances, and may be grouped together to form secretory organs or glands. Glands vary in complexity, but are of two main types (see Figure 1.1.2/3): exocrine glands release their secretions at an epithelial surface, whilst endocrine glands release their products into the blood or lymph system.

The simplest type of exocrine gland is a single cell, the mucous or goblet cell, which is found amongst the columnar epithelium of many mucous membranes (see Figure 1.1.2/4). It secretes mucin, which dissolves in water to form mucus. Mucus, and other substances, may also be produced by more complex multicellular glands. Some examples of multicellular exocrine glands are shown in Figure 1.1.2/5. They show considerable variation in form, but all remain connected to the surface epithelium (from which they developed, see Section 1.2.1) by ducts, through which secretions are released.

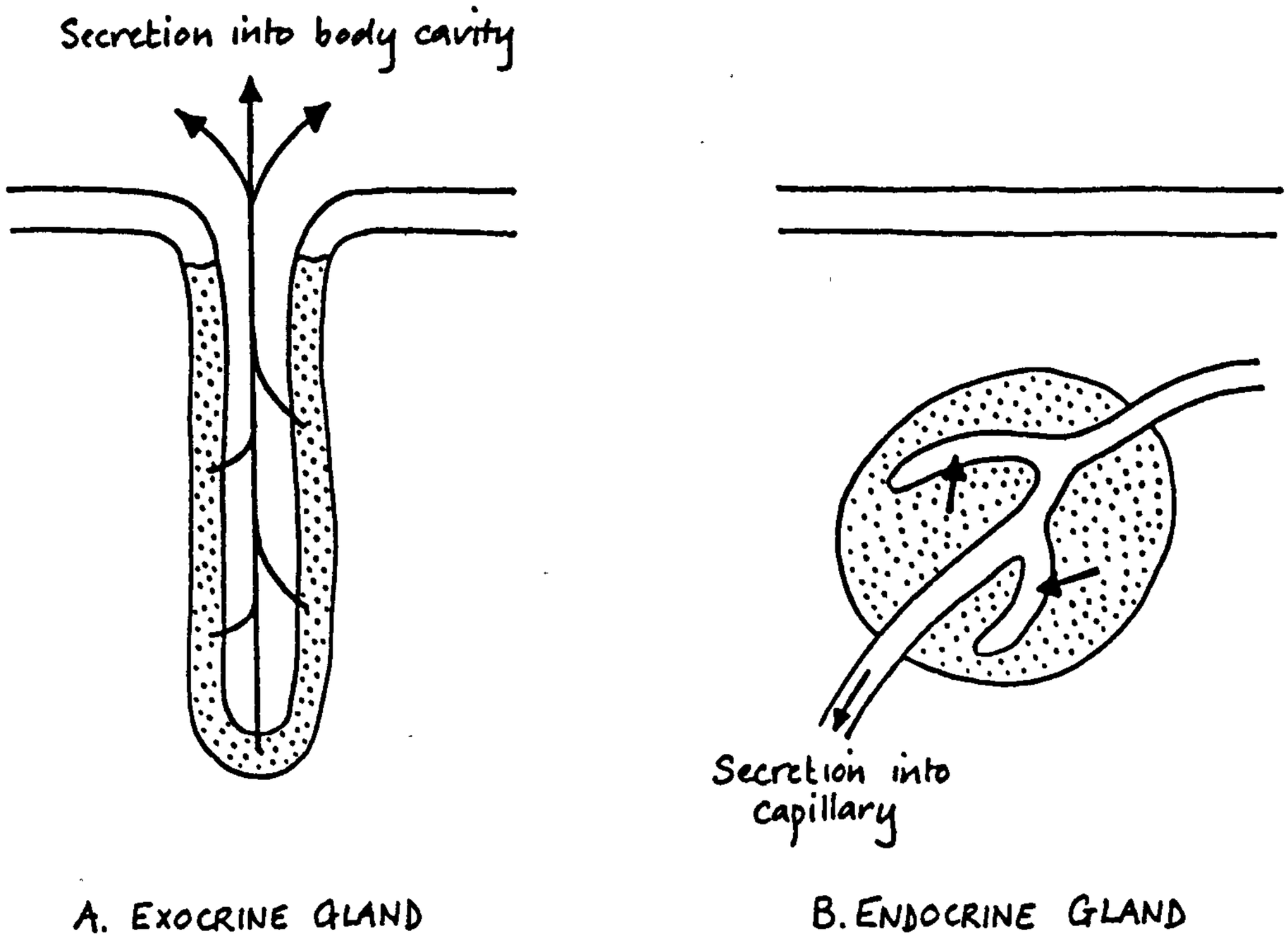


FIGURE 1.1.2/3: Exocrine and endocrine glands (adapted from Freeman and Bracegirdle 1966:6)

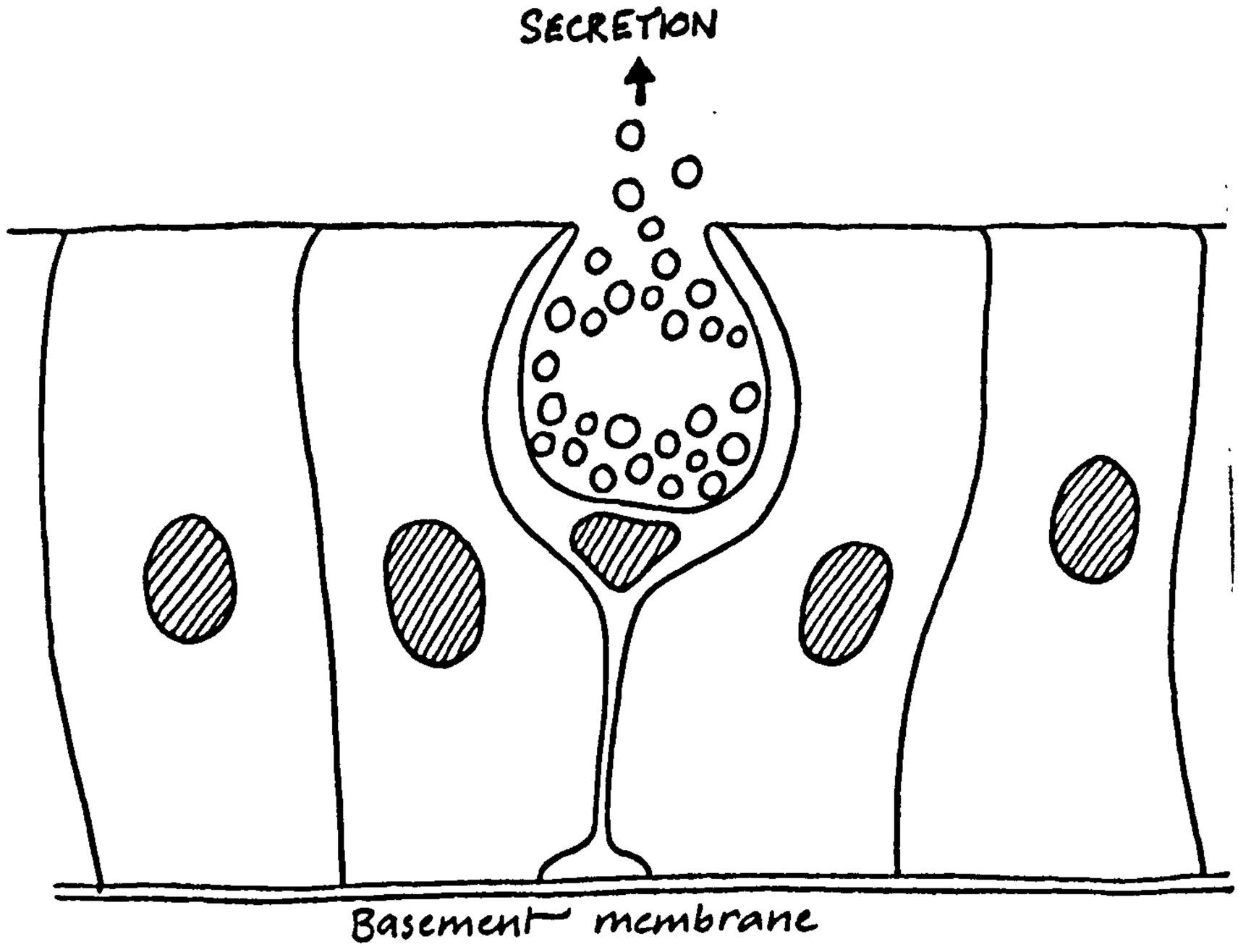


FIGURE 1.1.2/4: Unicellular exocrine gland (goblet cell - adapted from Freeman and Bracegirdle 1966:69)

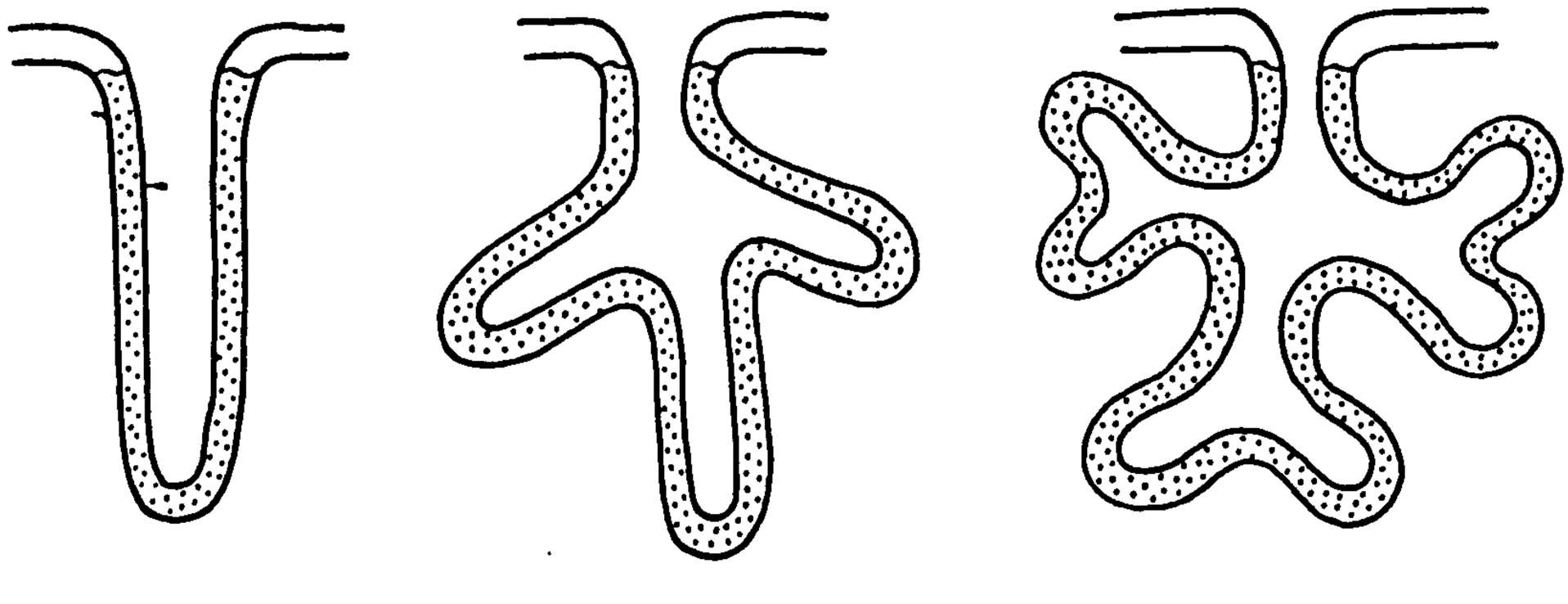


FIGURE 1.1.2/5: Multicellular exocrine glands

Endocrine glands, in contrast, have no ducts, but are closely associated with blood or lymphatic vessels into which their products are passed. Substances produced by endocrine glands are known as hormones, and because of their transport by the blood stream or the lymph system, they may exert an influence on parts of the body far removed from the glands which produce them.

Exocrine glands are of great importance within the vocal apparatus, since mucus is essential as a lubricant to keep the membrane lining in good condition. Healthy mucous membranes are of particular relevance to speech in the laryngeal area, since even minor changes in the mucous covering of the vocal folds may reduce the mechanical efficiency of the larynx with quite obvious phonetic consequences (see Chapter 2.5). Mucus also has an important cleansing function, helping to trap and remove particles of dust from the upper airways. If the ducts of exocrine glands become blocked, cysts of trapped secretory material may develop. These can produce very dramatic phonetic effects if they protrude into the glottis or constrict some other part of the vocal tract.

Endocrine glands have less direct effects on the vocal apparatus, but some forms of hormone imbalance, such as hypothyroidism, may have deleterious effects on the mucous membrane, and others may influence the overall development of the vocal organs.

B. Connective tissue

Connective tissue is very diverse in its manifestations, but all types are characterized by the presence of a considerable amount of intercellular substance (matrix). Connective tissue cells are of various types, some of which are responsible for producing the different sorts of matrix. It is the nature of the matrix which is of

primary importance in determining the mechanical and physiological properties of any given connective tissue.

Some types of connective tissue, such as blood and lymph, need not concern us here. Although both these fluids are of vital importance in servicing the tissues which make up the vocal apparatus, they do not contribute much to the bulk of these structures. The emphasis here will be on the class of tissues known as "connective tissue proper", and on cartilage and bone. The dentine and enamel of the teeth, which are partially derived from connective tissue will, for convenience, be dealt with as part of a separate description of teeth at the end of this section.

Connective tissue proper

This class of tissue consists of cells together with a matrix of fibres embedded within an amorphous ground substance. The appearance and behaviour of the tissue varies according to the relative proportions and arrangement of these constituents.

Cell types

Fibroblasts: These are responsible for the manufacture of fibres and of amorphous ground substance. They are described as fixed cells, but may actually be capable of some movement near healing or inflamed tissues.

Fibrocytes: Mature and relatively inactive fibroblasts are often called fibrocytes.

Macrophages: These are mobile cells, which are most abundant in loose areolar tissue (see below). They are able to ingest and destroy dead cells, bacteria and foreign bodies, and thus help to defend and clean the tissue.

Fat cells: Each of these cells stores a large droplet

of fat. They may be found singly or in clumps. Tissue which contains large accumulations of fat cells is known as adipose tissue.

Mast cells: These cells tend to congregate around small blood vessels, and seem to be involved in the production of heparin, an anti-coagulant, (i.e. it prevents blood from clotting) and histamine, which increases the permeability of blood vessels.

Other cells which may be found in connective tissue proper include white blood cells (leucocytes) and pigment cells.

Fibres

Three types of fibre may be present in connective tissue.

Collagen fibres: These are relatively coarse, transparent fibres, consisting of the protein collagen. They may be arranged in bundles, and when fresh they are soft, highly flexible, relatively inelastic and possessed of very high tensile strength (see section 1.1.3). Collagen fibres are present in almost all connective tissue.

Reticular fibres: These are very fine, branching fibres, forming networks around small blood vessels, muscle fibres and nerves, and within the lungs. They are abundant in the connective tissue immediately adjacent to epithelial sheets, and form part of the basement membrane of the epithelium. They are thought to have a similar molecular composition to collagen, and may be an immature form of collagen fibres.

Elastic fibres: These are fine threads or ribbons which

have a yellowish colour in bulk. They are made up largely of the protein elastin, and have, as the name suggests, the capacity of stretching easily, and of returning easily to their original length when tension is released.

Amorphous ground substance

The ground substance of connective tissues may be a viscous solution or a gel, and it has several functions.

i) It acts as a medium for the diffusion of nutrients and waste products between cells and capillaries.

ii) It may provide some support for the tissue.

iii) It may act as a selective barrier to electrically charged molecules and ions.

iv) It helps to localise invasion by bacteria.

v) It may act as a lubricant, diminishing friction in dense aggregations of collagen fibres (Bloom and Fawcett 1968:138).

The major categories of connective tissue proper which are of relevance to the vocal apparatus are summarised below. This is a fairly crude classification, and connective tissue with characteristics which are intermediate between two categories may be found.

Loose connective tissue (areolar tissue)

This is a loosely arranged tissue, which contains all the cell and fibre types described above in a fluid ground substance. Fibroblasts and macrophages are the commonest cell types. Collagenous fibres are arranged in a haphazard fashion, and elastic fibres form a loose, continuously branching network. Reticular fibres are scarce, except in the areas adjacent to other tissue types or structures. Loose connective tissue is found throughout the body as a packing or binding material,

connecting other tissues and organs, and affording a considerable degree of flexibility between structures. Figure 1.1.2/6a shows a schematic diagram of this tissue.

Dense connective tissue

In dense connective tissue the fibres are closely packed, with a corresponding reduction in the relative proportions of cells and ground substance. These tissues can be further subdivided according to the arrangement of the fibres.

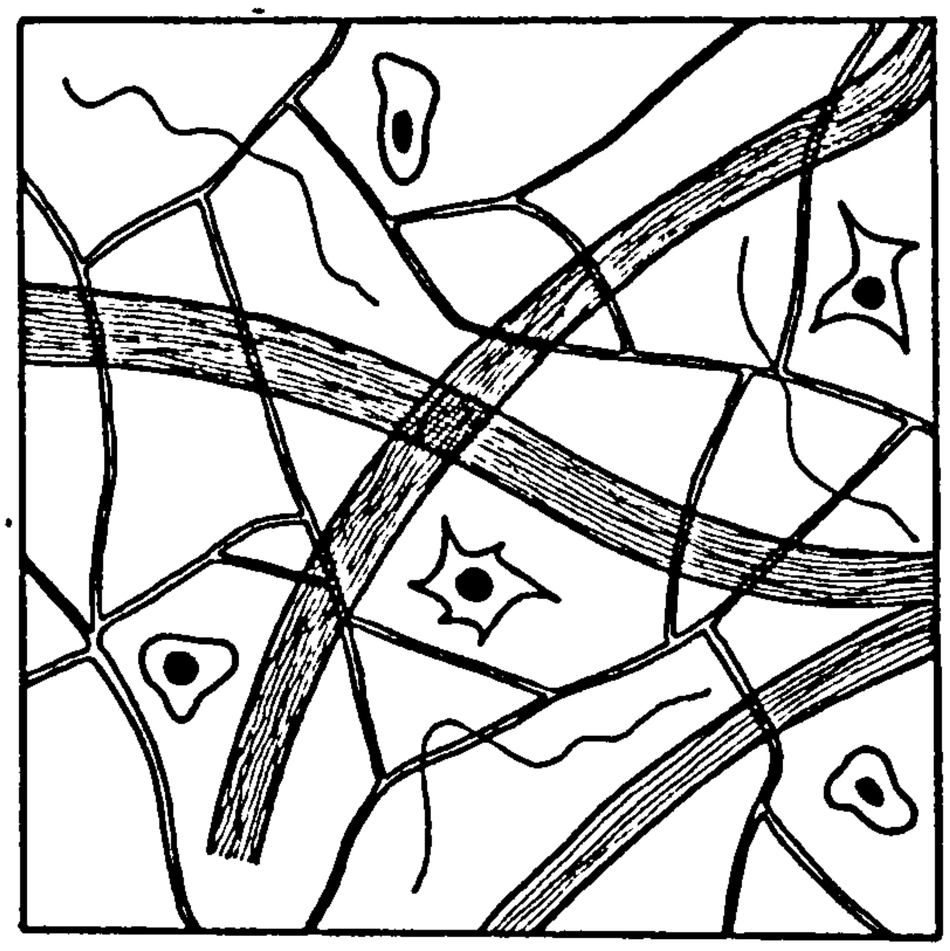
i) Irregularly arranged (see Figure 1.1.2/6b)

In areas which are subjected to tension in all directions, fibres are laid down haphazardly so that the resistance to tension is equal in all directions. This type of tissue occurs in sheets, and collagen fibres predominate, although some elastic and reticular fibres are also present. Irregularly arranged dense connective tissue is found in the dermis of the skin, and forms the fibrous sheaths of bones and cartilage. It also encapsulates some organs and lymph nodes.

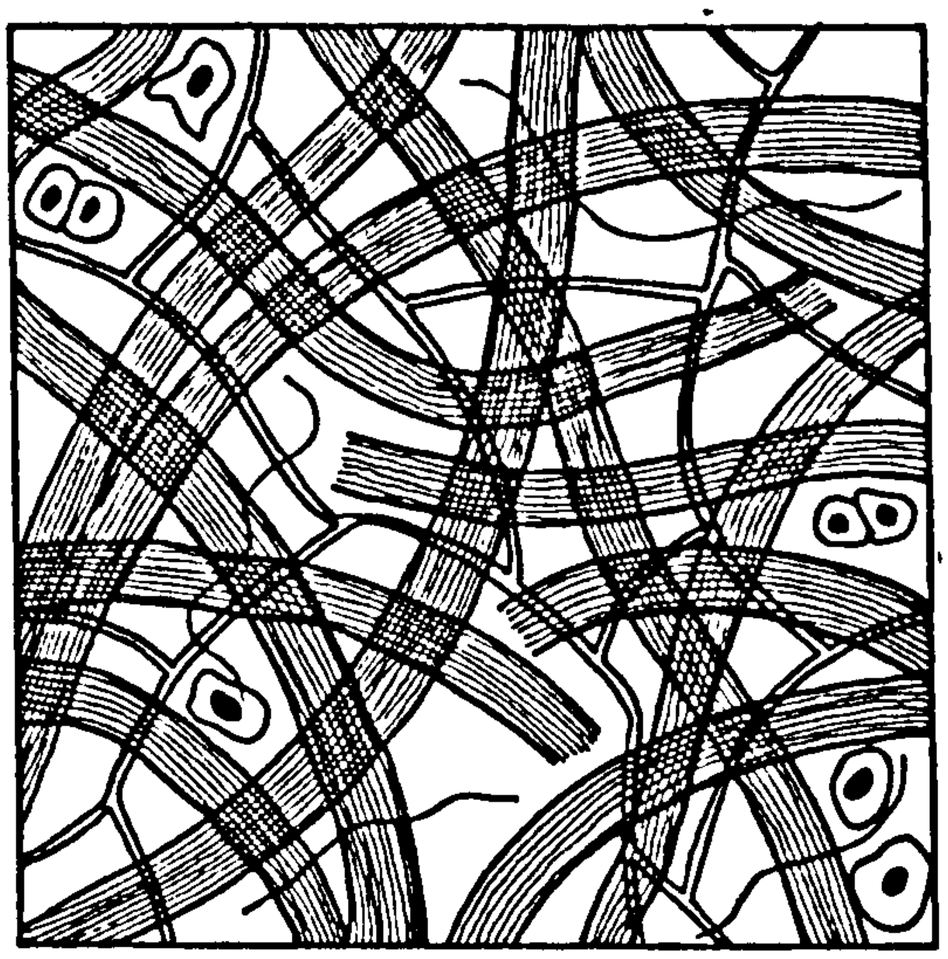
ii) Regularly arranged (see Figure 1.1.2/6 c and d)

Where tissues have to withstand tensions in one direction only, fibres are arranged in an orderly parallel fashion. In tendons, for example, where great tensile strength is demanded, collagenous fibres are present in high density and the only cells present in significant numbers are fibroblasts. Such tissues may be termed collagenous tissues.

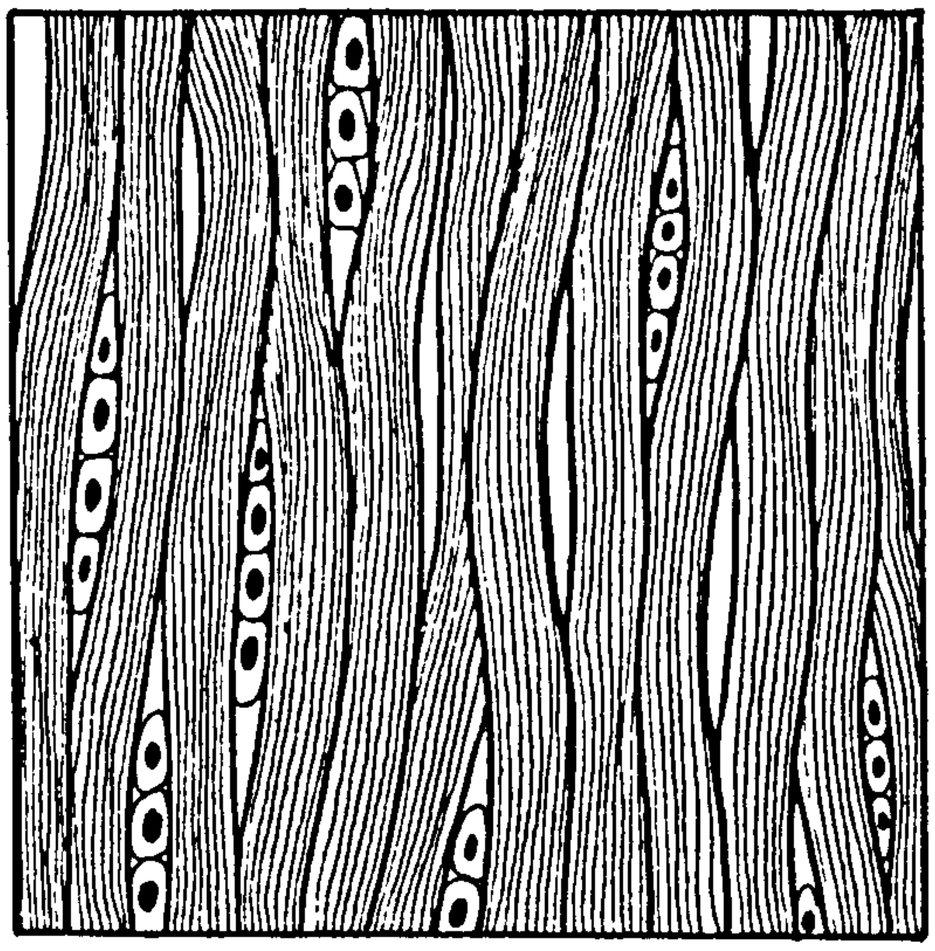
In a few areas, where the mechanical requirement is for elasticity rather than tensile strength, the bulk of the tissue may consist of elastic fibres. A notable example of this is seen in the vocal ligament. In elastic tissue, fibroblasts are rather more prominent than in collagenous tissue, and the fibres branch repeatedly and fuse with



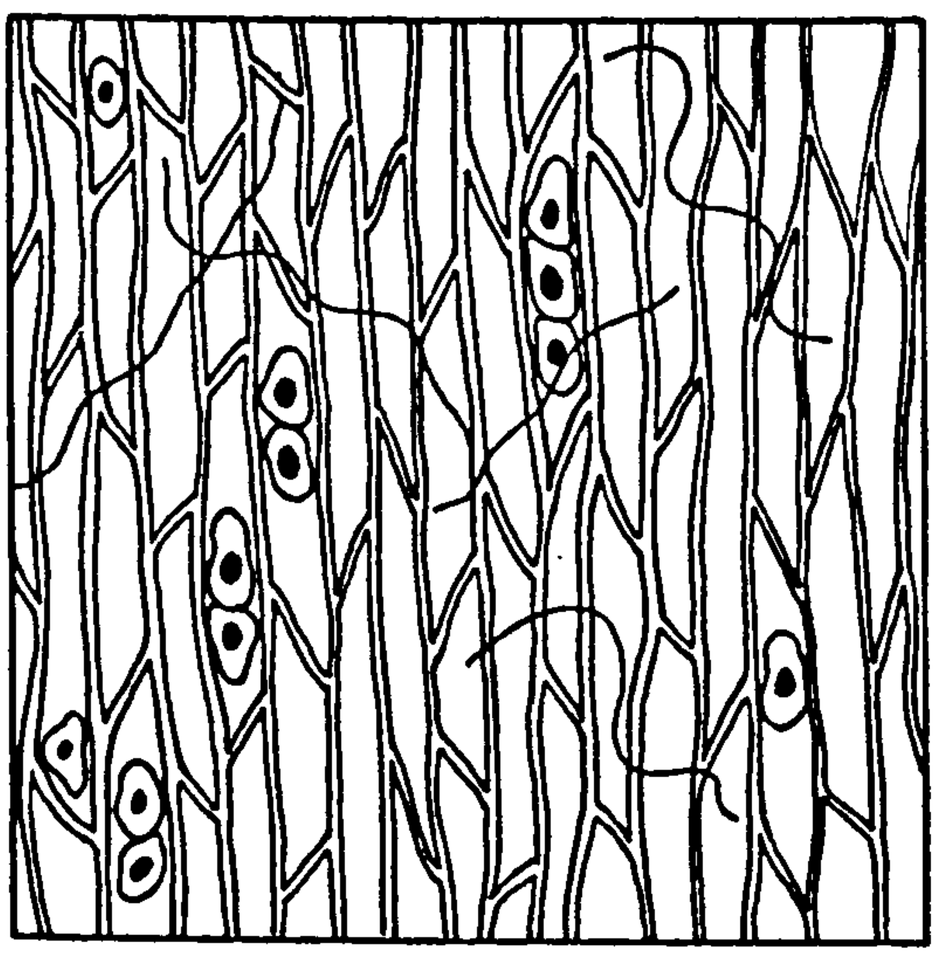
A.



B.



C.



D.

= Collagen fibre

= Reticular fibre

= Elastic fibre

= Cells

FIGURE 1.1.2/6: Schematic representation of types of connective tissue proper.
A. Loose connective tissue
B. Irregular dense connective tissue
C. Regular collagenous tissue
D. Regular elastic tissue.

one another. The tissue also contains a fine network of reticular fibres.

Cartilage

Cartilage is a rigid, but fairly flexible type of connective tissue, which forms part of the skeletal framework of the body. It may be classified into three types, depending on the types of fibres embedded in the matrix.

i) Hyaline cartilage. This contains fine collagen fibres.
ii) Elastic cartilage. This contains a predominance of elastic fibres, together with some collagen fibres.
iii) Fibrocartilage. This contains densely packed, coarse collagen fibres.

Only the first two types of cartilage are of particular relevance to the architecture of the vocal apparatus, since fibrocartilage is limited to sites which are subjected to considerable pressure, such as in intervertebral discs. Hyaline cartilage forms the nasal cartilage, the thyroid and cricoid cartilages, the bulk of the arytenoid cartilages, and the tracheal rings. Elastic cartilage is found in the epiglottis and at the tips of the arytenoid cartilages, and forms the cuneiform and corniculate cartilages (Leeson and Leeson 1976:398).

Two types of cell are present in all cartilage. Chondroblasts are concerned with manufacture of the cartilage matrix. As they mature and become isolated within the matrix, they change in appearance and are called chondrocytes. It may be better to think of chondroblasts and chondrocytes as being different developmental stages of one cell type rather than as separate cell types.

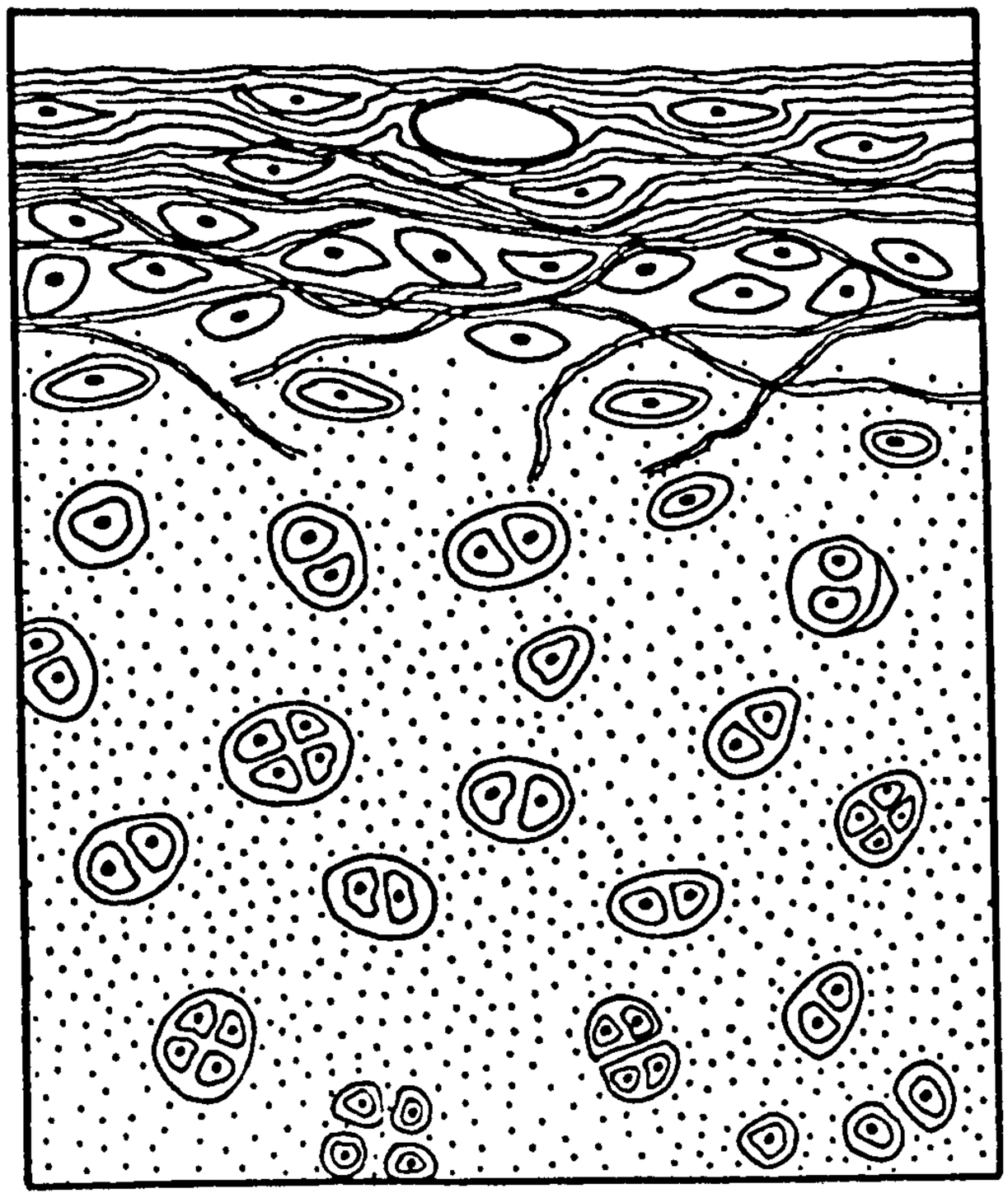
Hyaline cartilage

Hyaline cartilage earns its name by its transparent, glassy appearance, which is due to the fact that the fine collagen fibres have a similar refractive index to the surrounding intercellular substance. Junqueira and Carneiro (1980:122) state that 40% of the dry weight of hyaline cartilage is made up of these fine collagen fibres, embedded in amorphous ground substance. The collagen is mostly in the form of fibrils, which are finer than those found in most connective tissue. The rigidity of the cartilage results from chemical linkage between these fibrils and large molecules found in the ground substance.

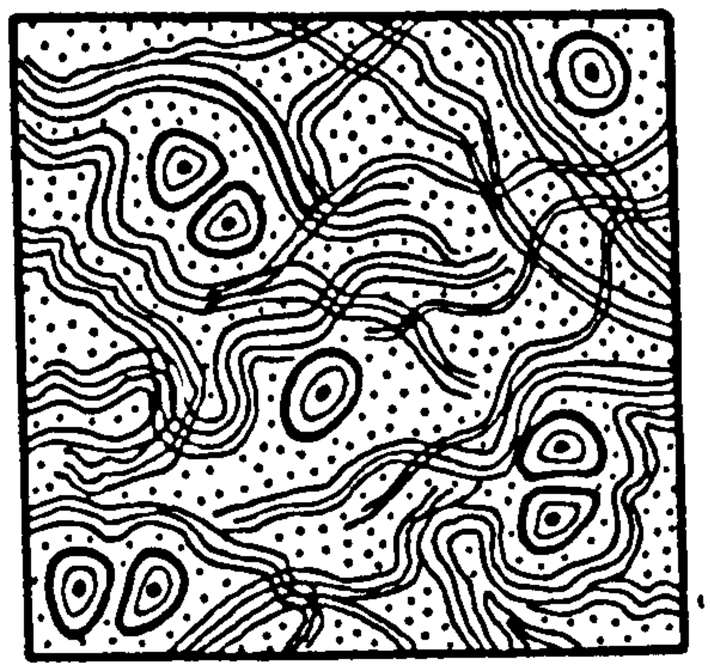
The cartilage is usually enclosed within a layer of tough, dense connective tissue, the perichondrium, which is rich in collagen, and contains cells which resemble fibroblasts. These cells are more numerous at the junction with the cartilage, and they may be precursors of chondroblasts, rather than true fibroblasts. Within the cartilage proper the chondrocytes are enclosed within the cartilage matrix. The area immediately surrounding each chondrocyte is rich in ground substance chemicals, but has little collagen content, and is called the capsule. Chondrocytes are elliptical near the surface of the cartilage, with the long axis lying parallel to the surface. Towards the centre of the cartilage the cells become rounder in shape, and may form groups derived from the division of single cells (see Figure 1.1.2/7). Cartilage derives nutrients and oxygen from blood vessels within the perichondrium, and its maximum thickness is therefore limited by the ability of nutrients and gases to diffuse through the matrix.

Elastic cartilage

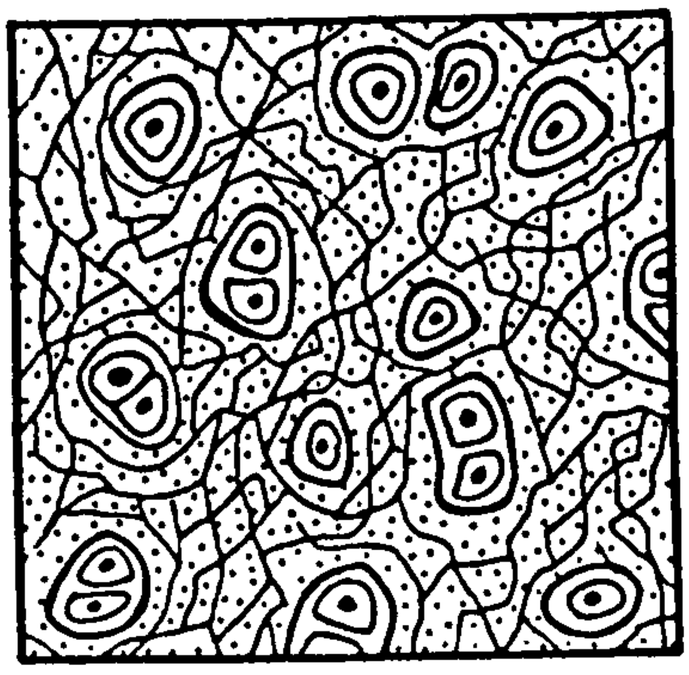
Elastic cartilage is found in sites where support needs to be associated with a high degree of flexibility. Its



A.



B.



C.

= cartilage matrix

= collagen fibres

= cells

= elastic fibres

FIGURE 1.1.2/7: Schematic representation of cartilage (adapted from Freeman and Bracegirdle 1966:25)
A. Hyaline cartilage
B. Fibrocartilage
C. Elastic cartilage

structure is broadly similar to that of hyaline cartilage, except that a preponderance of elastic fibres gives it a yellowish colour. The matrix contains a few collagen fibres, together with branching networks of elastic fibres, which are usually larger and more densely packed in the interior of the tissue.

Bone

Bone consists of specialized cells embedded within a bony matrix. The maintenance of bone as a living tissue depends on these cells receiving adequate oxygen and nutrients. The matrix is not a good medium for diffusion, so it accomodates a network of blood vessels to provide the cells' requirements.

Cell types

There are three cell types which are characteristic of bone: osteoblasts, osteocytes and osteoclasts. It seems likely that these cell types are capable of transformation from one type to another when necessary.

Osteoblasts: these are responsible for the manufacture of the bony matrix, and possibly also for its calcification (see section 1.2.1.b).

Osteocytes: during bone development osteoblasts become imprisoned within the bone matrix and become less active manufacturers of protein. They are then known as osteocytes.

Osteoclasts: these have several nuclei, and are thought to arise by fusion of other bone cell types. They seem to be associated with the resorption of bone (see section 1.2.1.b).

Bone matrix

Bone matrix is composed of organic material, water and inorganic matter (bone mineral). The bone mineral content increases throughout development, reaching a maximum of

about 65% of dry weight (Bloom and Fawcett 1968:229, Leeson and Leeson 1976:144). Bone mineral has a crystalline structure, and is composed largely of calcium phosphate. The organic material consists mostly of collagen fibres within an amorphous ground substance.

Bone architecture

Bone architecture varies according to the relative needs for strength and lightness in any particular bone. Two broad classes of architectural design can be distinguished: spongy bone and dense bone.

Spongy bone is made up of a system of bony struts (trabeculae), between which there are large cavities containing blood vessels and cells, which make up the bone marrow (see Figure 1.1.2/8). This type of bone has the virtue of lightness, but is relatively weak.

Dense bone, which is stronger, consists of dense deposits of bone matrix, laid down in concentric circles. At the centre of these circles are canals, which carry blood and lymph vessels. Each canal, together with the surrounding layers of bone matrix and the osteocytes which it supports, is called a Haversian system (see Figure 1.1.2/8). The osteocytes lie within lacunae (small spaces within the matrix), and communicate with the central canal via a network of fine canaliculi.

Periosteum

Most bones are enclosed within a fibrous sheath of collagenous connective tissue containing some elastic fibres and a network of blood vessels. This is the periosteum, which is closely attached to the bone by collagenous fibres which penetrate the bone.

In bones containing a cavity of bone marrow (a centre for blood cell production), the cavity is lined with a

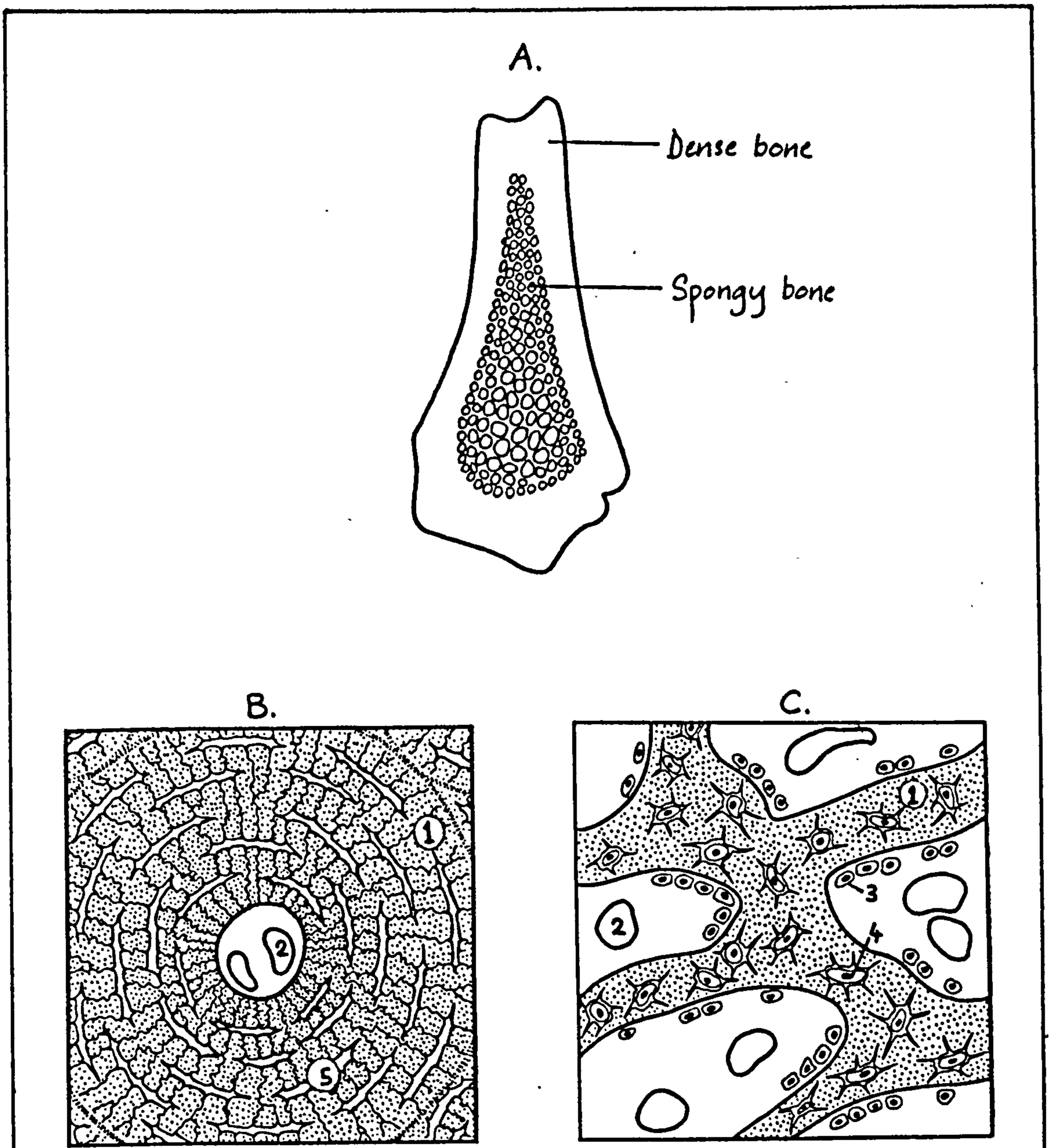


FIGURE 1.1.2/8: Schematic representation of dense and spongy bone (adapted from Freeman and Bracegirdle 1966:30,31)
A. Section through the mandible to show the distribution of dense and spongy bone
B. Dense bone
C. Spongy bone

similar but thinner layer of connective tissue, the endosteum.

Lymphoid (lymphatic) tissue

Lymphoid tissue forms part of the body's defence system, helping to limit infection by filtering lymph (lymph is a circulatory fluid) and destroying dead cells and organisms. Lymphatic tissue is made up of reticular tissue, within which are free cells, most of which are lymphocytes.

Reticular tissue, which is found chiefly in lymphoid tissue, bone marrow and the liver, is characterized by a structure of interconnected reticular fibres associated with primitive reticular cells. These are star-shaped cells, which seem to be linked to other cells by long cytoplasmic protrusions. Some behave very like fibroblasts, whilst others are phagocytic, i.e. they are capable of engulfing and destroying debris and foreign organisms. Reticular cells are thought to retain the potential to develop into a variety of other cell types, and may give rise to free macrophages (cells specialized for phagocytosis), to precursors of erythrocytes and leucocytes (types of blood cell), and to other cell types. Lymphocytes are involved in antibody production, so that the presence of free lymphocytes in lymphoid tissue is associated with the defensive role of the tissue.

Diffuse lymphoid tissue

In some areas of the body. lymphoid tissue is not clearly separated from the surrounding tissue, and is known as diffuse lymphoid tissue. This type shows no special organization, and is found most commonly within the lamina propria of mucous membranes.

Lymph nodules

Lymphoid tissue frequently occurs as denser, spherical aggregates of tissue, which are surrounded by a border of small lymphocytes. Lymph nodules may occur singly, or they may be organised into specific lymphoid organs such as the tonsils, or lymph nodes. Lymph nodes are aggregates of lymphoid tissue within a capsule of connective tissue. Lymph nodules vary in number and in position, arising in response to infection, and then disappearing. Infection may result in the temporary formation of lymph nodules within diffuse lymphoid tissue.

The greatest aggregated mass of lymphoid tissue within the vocal apparatus is formed by the tonsils. The three groups of tonsils (palatine, lingual and pharyngeal tonsils) share a similar structure, being made up of depressions in the covering epithelium surrounded by collections of lymph nodules.

Teeth

Teeth are highly specialized structures, derived from connective tissue and epithelium. Although they are complex structures made up of several tissue types, and therefore fit uncomfortably in a section on basic building materials, it is convenient to describe them here because they contain a unique set of tissue types. Figure 1.1.2/9 is a longitudinal section through a tooth, showing the layers of tissue in question.

The outer layer of enamel is the only part of the tooth which arises from epithelial tissue, and is the hardest substance in the body, protecting the biting and chewing surfaces of the teeth. When fully developed, the mineral content is as high as 96%, mostly in the form of calcium salts.

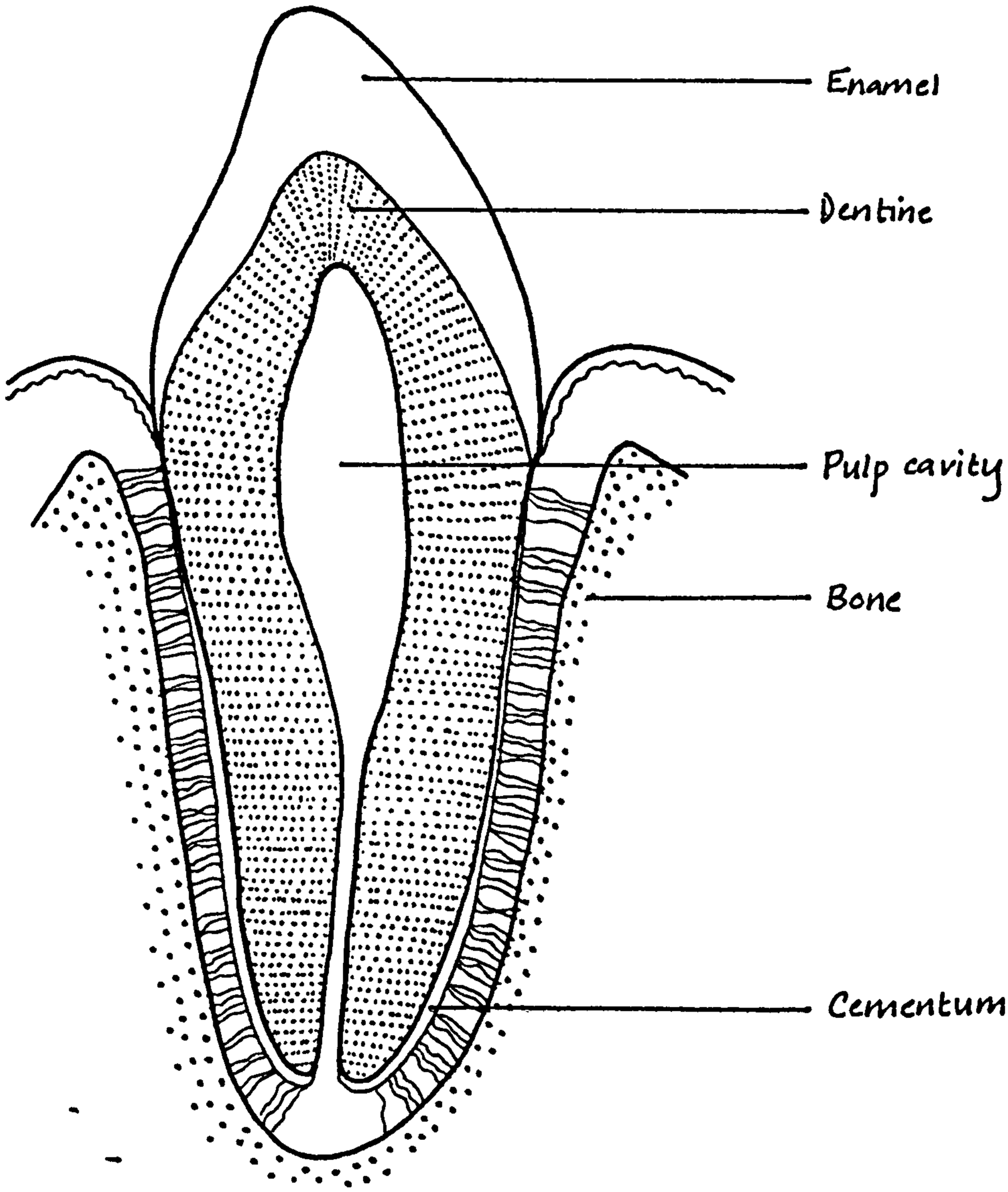


FIGURE 1.1.2/9: Longitudinal section of an incisor tooth

Dentine, which forms the bulk of the tooth, is similar to dense bone in chemical composition, with 28% organic material, and 72% mineral content. The cells of this tissue, the odontoblasts, are concentrated around the pulp cavity in fully developed teeth.

Cementum, which covers the dentine of the root, is also similar in structure to bone. Coarse bundles of collagen fibres penetrate the surrounding membrane (the periodontal membrane), thus helping to anchor the tooth in its socket. The cells of the cementum (cementocytes) are found only in the thicker, lower area of cementum, and lie in lacunae.

The pulp cavity contains connective tissue with a loose arrangement of collagen fibres in an amorphous ground substance, together with lymphocytes, macrophages and cells peculiar to pulp. The pulp contains a network of blood vessels and nerves.

The periodontal membrane also forms the periosteum for the alveolar bone, within which the teeth are embedded. It is rich in collagenous fibres, which run from the alveolar bone to the cementum, and is more rigid than most periosteum because it lacks elastic fibres.

C. Muscle tissue

Muscle tissue is a major structural component of the vocal apparatus, forming much of the bulk of the vocal tract walls. It is unique amongst the other tissue types described here in that it is capable of changing its length and shape, as well as its mechanical and acoustic properties, through the process of contraction. Many muscles can reduce their resting length by up to 50%, although a more efficient working range is usually within a 10% length change (Dickson and Maue-Dickson 1982:38).

There may be as much as a tenfold difference in elasticity between resting and contracted muscle (Hill 1970, cited by Hirano et al. 1982), and resting muscle will tend to absorb more acoustic energy (Laver 1980:143).

Muscle activity is responsible, through connective tissue attachments to bones and cartilages, for all moment to moment movements of the vocal organs. These range from subtle changes, such as the adjustments of the larynx involved in pitch control, to gross movements of the tongue and jaw.

There are three types of muscle, all of which are made up of parallel bundles of elongated muscle cells or fibres, which are specialized for contraction.

i) Cardiac muscle. This is found in the walls of the heart, and contracts rhythmically and continuously throughout life.

ii) Smooth muscle. This is found in places where slow, steady contractions are needed, which are not under voluntary, or conscious control. The walls of the intestines and of some blood vessels contain smooth muscle, as does the iris of the eye.

iii) Skeletal (or striated) muscle. This is capable of strong, rapid contractions, and may be under voluntary control. This is the muscle type which controls the movement of skeletal structures relative to one another, and since it is the only muscle type which adds significant bulk to the vocal organs, it will be the only muscle type to be described here.

Skeletal muscle

Like all types of muscle tissue, skeletal muscle is made up of elongated muscle fibres bound within a framework of connective tissue. This connective tissue is modified in

places to form tendons, which anchor muscles to the skeleton. The composition of the connective tissue sheath varies according to the function of the muscle. The muscles of the tongue, for example, which need to be highly mobile and elastic, are encased in connective tissue which is rich in elastic fibres (Clegg and Clegg 1963:98). Greater densities of collagen fibres are found in the connective tissue of muscles requiring greater tensile strength, such as the major muscles concerned with locomotion.

Figure 1.1.2/10 shows a schematic diagram of a typical skeletal muscle. Muscle cells, or fibres, are unusual in that each one contains many nuclei. Each fibre is enclosed within a specialized membrane, the sarcolemma, which seems to form strong mechanical links between the contractile elements within the fibre and the surrounding connective tissue, thus helping to transmit the contractile force to the skeleton. The contractile elements themselves, the myofibrils, are stacked longitudinally within the muscle fibre, and give the banded appearance which earns this muscle its alternative name, striated muscle. The detailed structure of myofibrils is an amazing example of bio-engineering. Protein filaments interlock in such a way that they can slide over one another to reduce overall length. This is shown schematically in figure 1.1.2/11. The filaments are linked by a system of molecular cross-bridges, and the sliding movement is achieved by these cross bridges breaking, swivelling, and rejoining at a position a little further along the neighbouring filament. This process, which has been likened to an "animated cogwheel" (Leeson and Leeson 1976:195), is a chemical response to stimulation by nerve endings at the muscle fibre membrane (= motor end plate). Relaxation involves a reversal of the sliding mechanism.

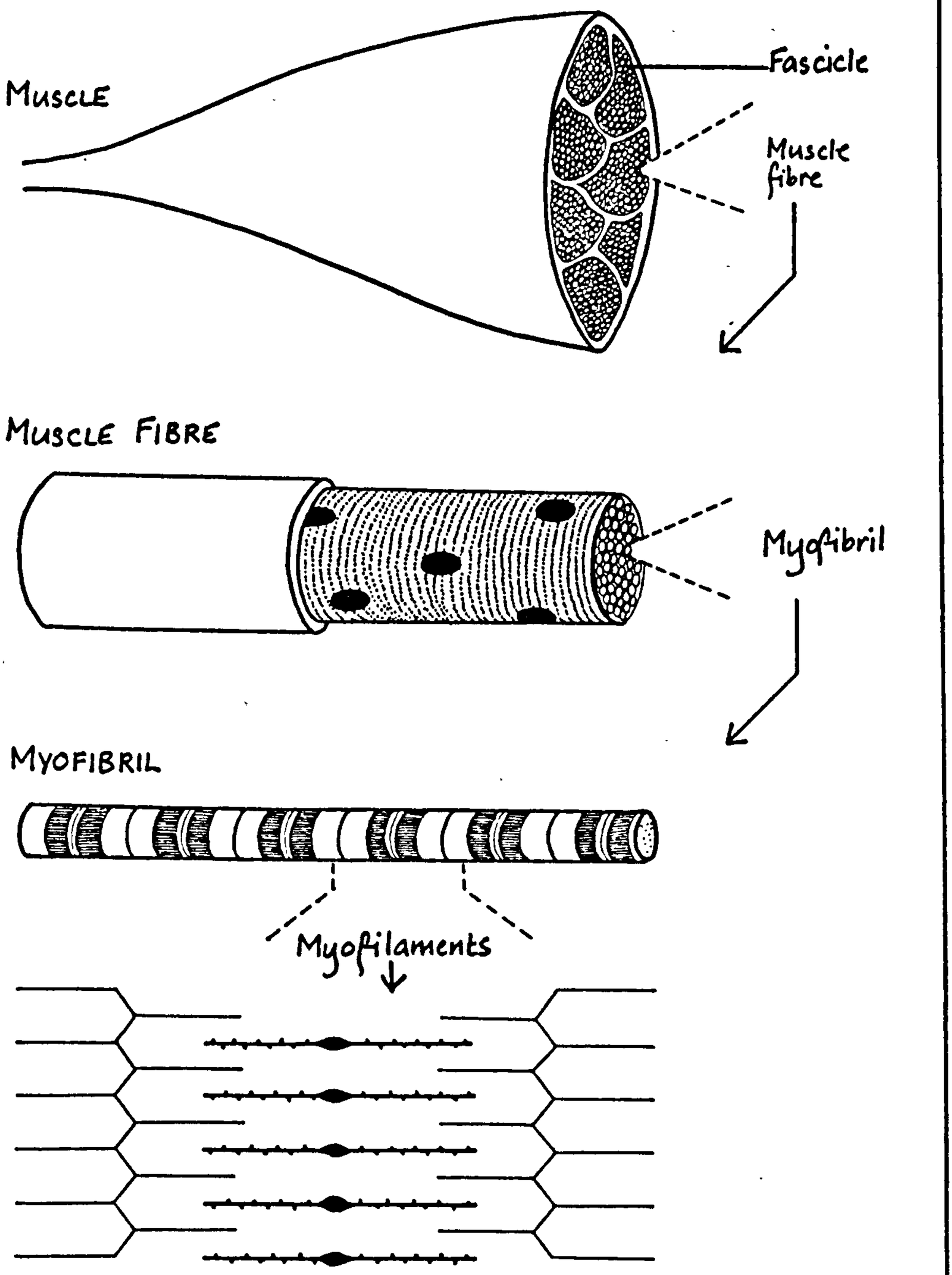


FIGURE 1.1.2/10: Schematic diagram of skeletal muscle, exploded view (adapted from Dickson and Maue-Dickson 1982:32)

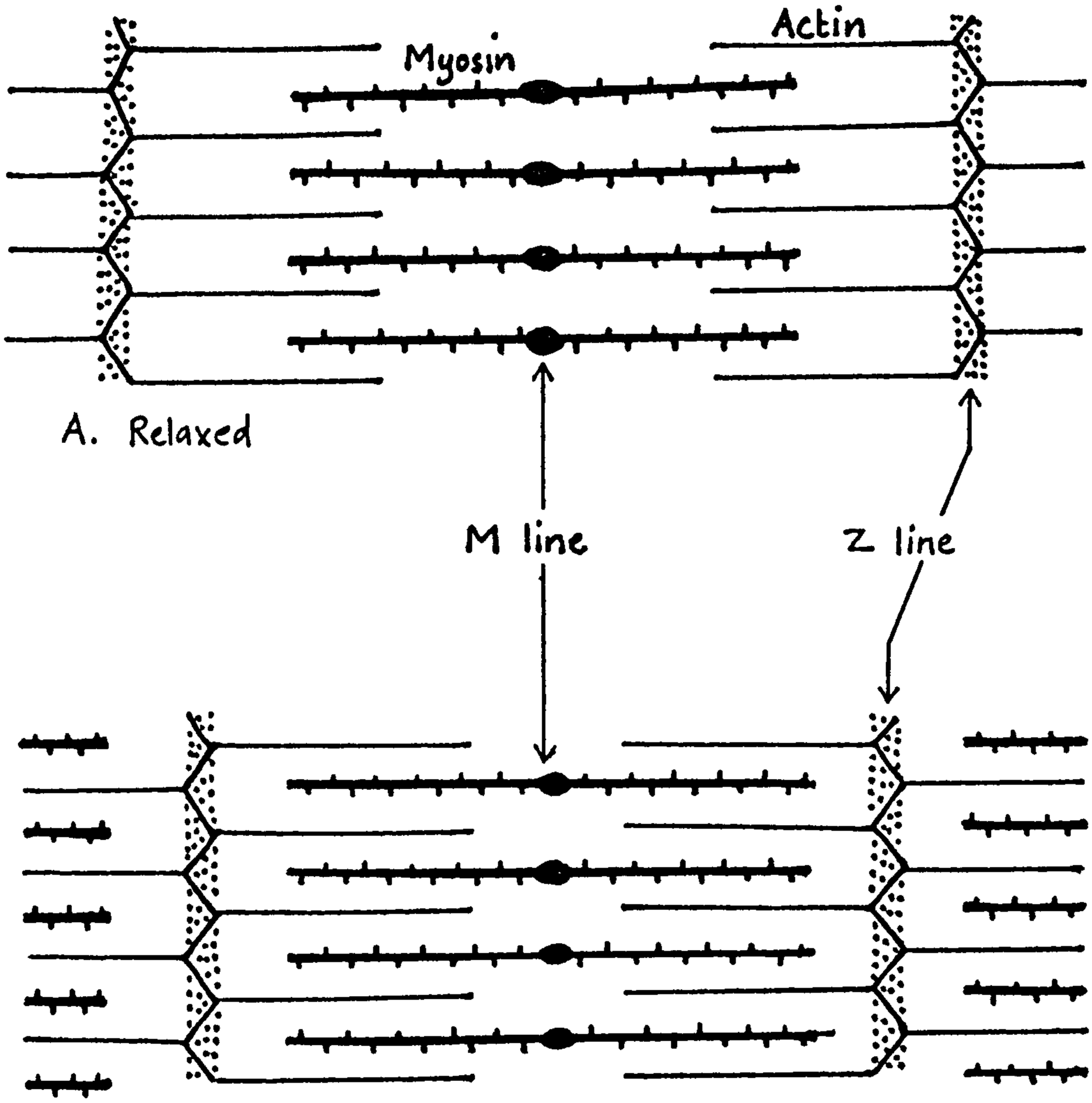


FIGURE 1.1.2/11 Schematic diagram of skeletal muscle contraction (adapted from Dickson and Maue-Dickson 1982:34)

Muscle architecture

Skeletal muscles vary greatly in size and shape, ranging from a few millimetres to tens of centimetres in length. The shape depends on several factors. The required area of attachment is one important factor governing overall shape, whilst the arrangement of fibres within the muscle will depend on the relative needs for power and range of movement. Greater power requires a greater number of muscle fibres, whilst greater range of movement demands greater length of the fascicles (a fascicle = a bundle of muscle fibres). Some examples of common muscle forms are shown in Figure 1.1.2/12.

D. Nervous Tissue

Nervous tissue is obviously crucial in controlling muscular and physiological adjustments of the vocal tract, and it also has an indirect influence on vocal tract development, since bone modelling and muscle development are in part a response to muscle activity. It does not, however, contribute significantly to the mass of the vocal organs. Since the focus of this thesis is on the overall morphological form of the vocal apparatus it does not, therefore, seem useful to include a description of the complexities of nervous tissue. Descriptions of nerve histology may be found in all the texts mentioned at the beginning of this section, and the specific role of nervous tissue in speech and language is amply covered in such texts as Espir and Rose (1976) and Dickson and Maue-Dickson (1982).

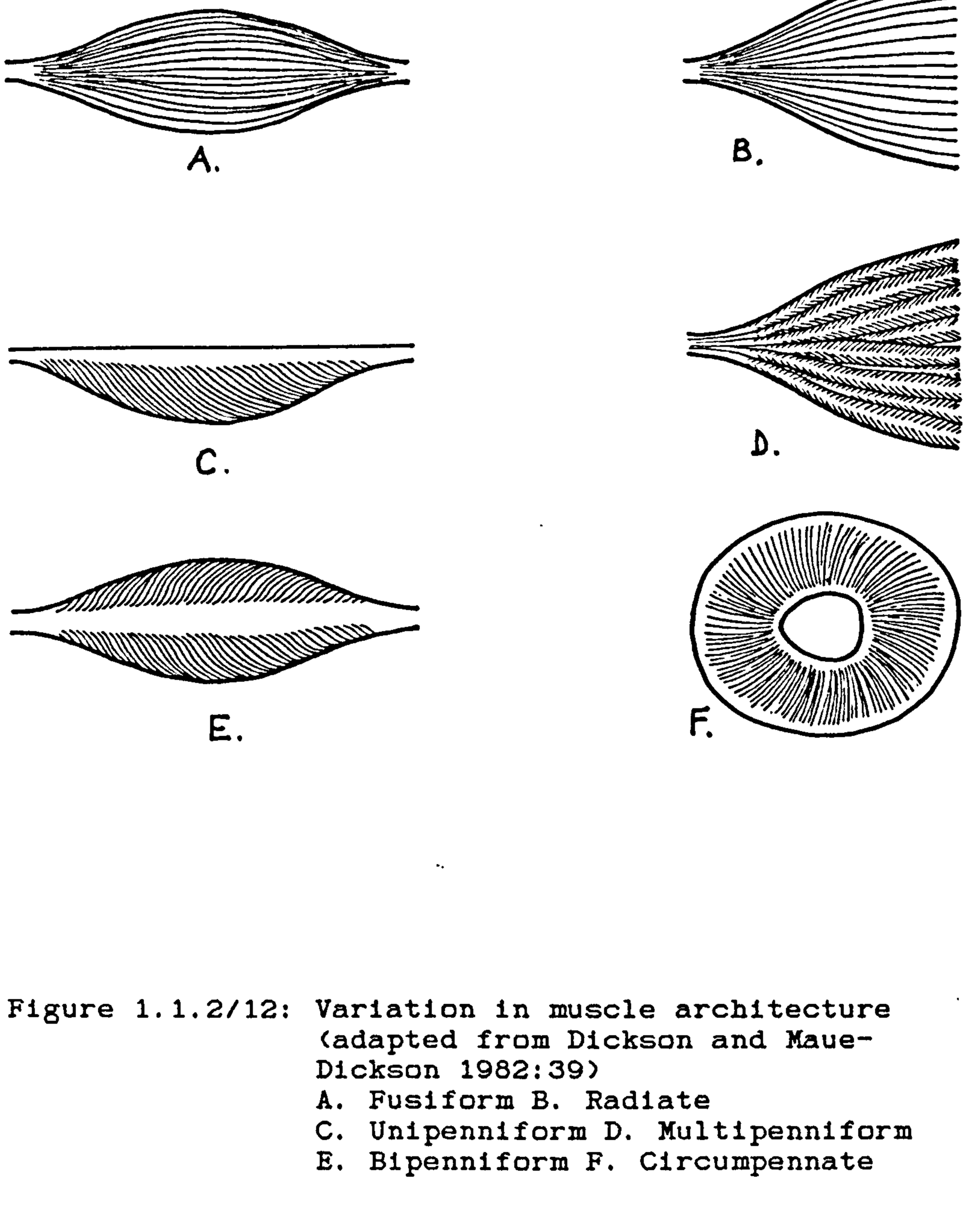


Figure 1.1.2/12: Variation in muscle architecture (adapted from Dickson and Maue-Dickson 1982:39)
A. Fusiform B. Radiate
C. Unipenniform D. Multipenniform
E. Bipenniform F. Circumpennate

1.1.3 MECHANICAL CHARACTERISTICS OF TISSUES, WITH SPECIAL REFERENCE TO THE VOCAL FOLD

It should already be clear from the previous section that the mechanical characteristics of any given tissue will depend partly on the type and arrangement of cells, and partly on the consistency and arrangement of any extracellular material which is present. The proper functioning of the vocal apparatus is very much dependent upon the mechanical characteristics of its constituent tissues, and this is especially obvious where the vocal fold is concerned. Since a later chapter (2.5) will describe some of the acoustic consequences which result from disturbances in vocal fold structure, and try to relate these to mechanical factors, it may be helpful to offer a brief summary of some of the mechanical measurements which may be applied to tissues. This will then be illustrated by reference to the detailed tissue layer structure of the vocal fold.

Basic concepts in tissue biomechanics

There are three simple measurable concepts which form the basis of all engineering and mechanical descriptions of the world (see, for example, Kenedi 1980:1). These are force, length and time. A change of force will result in a change in length (which may be evident as motion, deformation or simple stretching), and the relationship between these changes will generally be time-dependent. Force and length are examples of vector quantities, since their full description requires specification of both magnitude and direction. Time is a scalar quantity, being defined in terms of magnitude only. All the measurements which are applied to tissue biomechanics can be derived from these three quantities. Some examples of useful mechanical terms are summarised below.

Stress: this is the force applied per unit of original area of tissue.

Strain: this is defined as the change in some geometric characteristic (e.g. length, angle) per unit of original size.

Tensile loading: this is the application of a force in such a way that it tends to produce an increase in length.

Tensile stiffness and elasticity. Generally stiffness can be defined in terms of the amount of stress required to produce a given deformation. Tensile stiffness can therefore be defined as the amount of stress needed to produce a given increase in length. In other words it is a measure of how difficult it is to stretch a material. Elasticity is thus inversely related to tensile stiffness.

Tensile strength: this is the extent to which a material can withstand tensile loading without breaking or distorting irreversibly.

Isotropy: Some materials and tissues will show the same mechanical resistance to stress regardless of the direction of loading. These are known as isotropic materials. Other materials tend to show different responses to stress, depending on the direction of loading, and are known as anisotropic materials. The degree of anisotropy of a tissue is often related to the arrangement of cells and fibres within the tissue. Tissues where the major structural elements show a clear directionality, like muscle or regular fibrous connective tissue where the fibres lie parallel to one another, will tend to be anisotropic.

Compressibility: this is the ease with which a material can be compressed, i.e. reduced in length as a result of application of a force.

Problems in applying mechanical theory to biological tissues

There are several reasons why the results of standard mechanical measurement techniques may need to be treated with caution when applied to biological tissues. Some of these relate to the special mechanical properties observed in living tissue. Biological tissues show some mechanical properties which complicate standard procedures (Kenedi 1980). For example, many tissues show stress relaxation. This means that if a tissue is stretched by the application of a tensile load, the stress required to maintain that extension decreases with time. Related to this is the phenomenon known as creep. If a constant stress is applied to human tissue, it will continue to deform with time. The time course of mechanical testing of biological tissues may therefore influence the results obtained. A further problem is that some standard measurements assume that there is a linear relationship between stress and strain. Very few materials adhere exactly to such a relationship, but for most engineering relationships the assumption of a linear relationship allows a reasonable approximation to reality. The relationship between stress and strain in biological materials is so non-linear that these assumptions are not valid, and so values for such measures as Young's modulus (a measure of tensile stiffness) commonly applied to tissues need to be treated with an element of scepticism.

A further complicating factor when dealing with living tissue is that its response to mechanical stress may be physiological as well as mechanical. For example, if a

tissue is exposed to mechanical stress, there may be changes in the blood supply which can ultimately lead to necrosis (Kenedi 1980:71). The mechanical properties of a tissue will also be affected by its physiological state, and the problems of maintaining tissue in a living condition, without dehydration, during testing procedures are quite complex.

Normal vocal fold structure

The vocal fold is a structure whose efficient functioning demands a very precise mechanical balance between its tissues. The mechanical characteristics of the vocal fold are of particular interest within the general scope of this thesis because quite slight organic changes may have very dramatic effects on the mode of vocal fold vibration. This description will supply a background for the discussion of the mechanical, and hence acoustic, consequences of various types of vocal fold pathology which can be found in Section 2.5.

The anatomy of the cartilages, muscles and other tissues which make up the larynx has been extensively described elsewhere (Kaplan 1960, Saunders 1964, Hardcastle 1976, Romanes 1978, Laver 1980, Dickson and Maue-Dickson 1982). This description will concentrate only on the tissues of the vocal folds themselves, and on the cartilages with which they are intimately associated. The area of focus will thus be the region bordered anteriorly and laterally by the thyroid cartilage, and extending as far back as the posterior edges of the arytenoid cartilages. In the vertical dimension, the region includes only the true vocal folds, and so the inferior border can be drawn at the level of the upper edge of the cricoid cartilage.

A convenient distinction can be made between the anterior two thirds of each vocal fold, which is bordered at the

glottal edge by the vocal ligament, and the posterior one third, where the inner edge of the arytenoid cartilage, from the vocal process to the inner "heel" of the cartilage, forms the glottal border. We can then refer to the "ligamental" part of the vocal fold and the "cartilaginous" part. This follows the convention initiated by Morris (1953) and followed by Laver (1980) of distinguishing between the intermembraneous or ligamental glottis and the cartilaginous glottis.

A schematic plan of the vocal fold region is shown in Figure 1.1.3/1. The following account offers a brief description of the arrangement of tissues within the vocal folds, together with some comment on the mechanical properties of each tissue type.

A. The ligamental area of the vocal fold

The ligamental area of the vocal fold is the one most freely involved in vibration during phonation, and it has therefore attracted the most attention from researchers concerned with vocal fold mechanics. Hirano and his associates have recently built up a considerable body of information about the histological structure of the vocal fold, and their work necessarily forms a base for the account below (Hirano 1981, Hirano et al. 1981, Hirano et al. 1982). Background sources also include standard texts on anatomy and histology (Davies and Davies 1962, Freeman and Bracegirdle 1967, Leeson and Leeson 1976, Romanes 1978).

Tissue types

The vocal fold is a layered structure, which in the ligamental area consists of the vocalis muscle and a covering of mucous membrane. The importance of these two layers in determining the fine detail of phonation has long been accepted (Smith 1961, Perello 1962, Baer 1973),

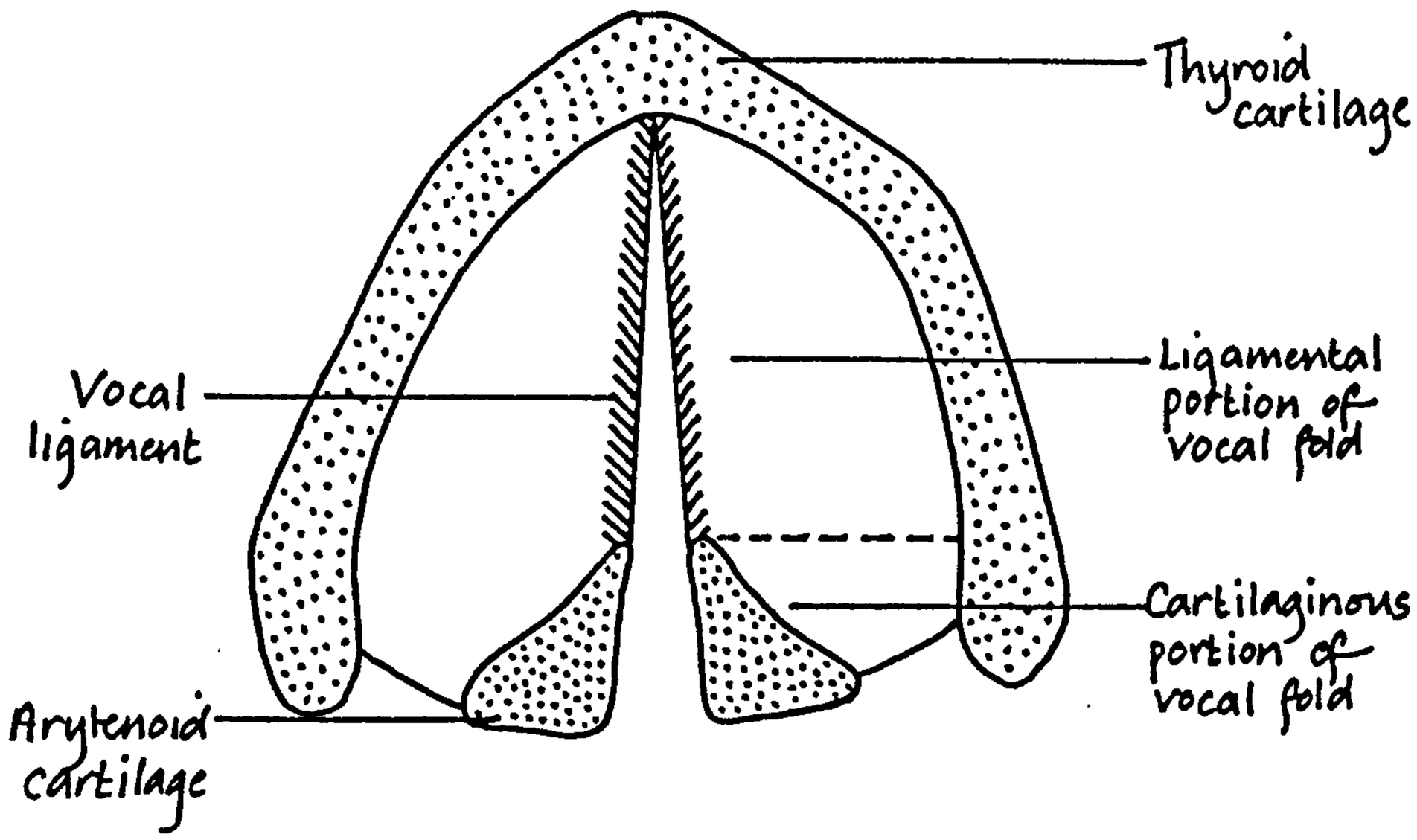


FIGURE 1.1.3/1: A schematic view of the vocal folds, seen from above.

but Hirano's work highlights further tissue type distinctions within the mucous membrane. This is divisible into four layers: an outer layer of epithelium, and three layers of underlying connective tissue. These inner three layers together make up the lamina propria (see Figures 1.1.3/2 and 1.1.3/3). The vocal ligament is formed by an uneven distribution of the layers of the lamina propria, and will be described later in this section.

Epithelium

The epithelial covering of the free border of the vocal fold is of the type known as non-keratinizing stratified squamous epithelium. This was described in the previous section, but if a reminder is needed, these three descriptive labels relate simply to the detailed structure. It is non-keratinizing because it does not produce keratin. The term "stratified" describes the arrangement of cells, which are here arranged in orderly layers, with the deepest layer resting on the basement membrane. The basement membrane is a zone where substances similar to those in the ground substance of the lamina propria are highly condensed, to form a thin sheet dividing the two tissue types. The term "squamous" refers to the shape of the cells, which are commonly likened to paving stones. In surface view they are usually polygonal, but cross sections show flattening, which is most pronounced in the surface layers.

The number of cell layers in the epithelium probably varies considerably, but in a large post-mortem study of 942 adult male larynges, Auerbach et al. (1970) found most samples of vocal fold epithelium to be between 5 and 10 cells thick. Hirano et al. (1982:278) report that there is no systematic relationship between epithelial thickness and age.

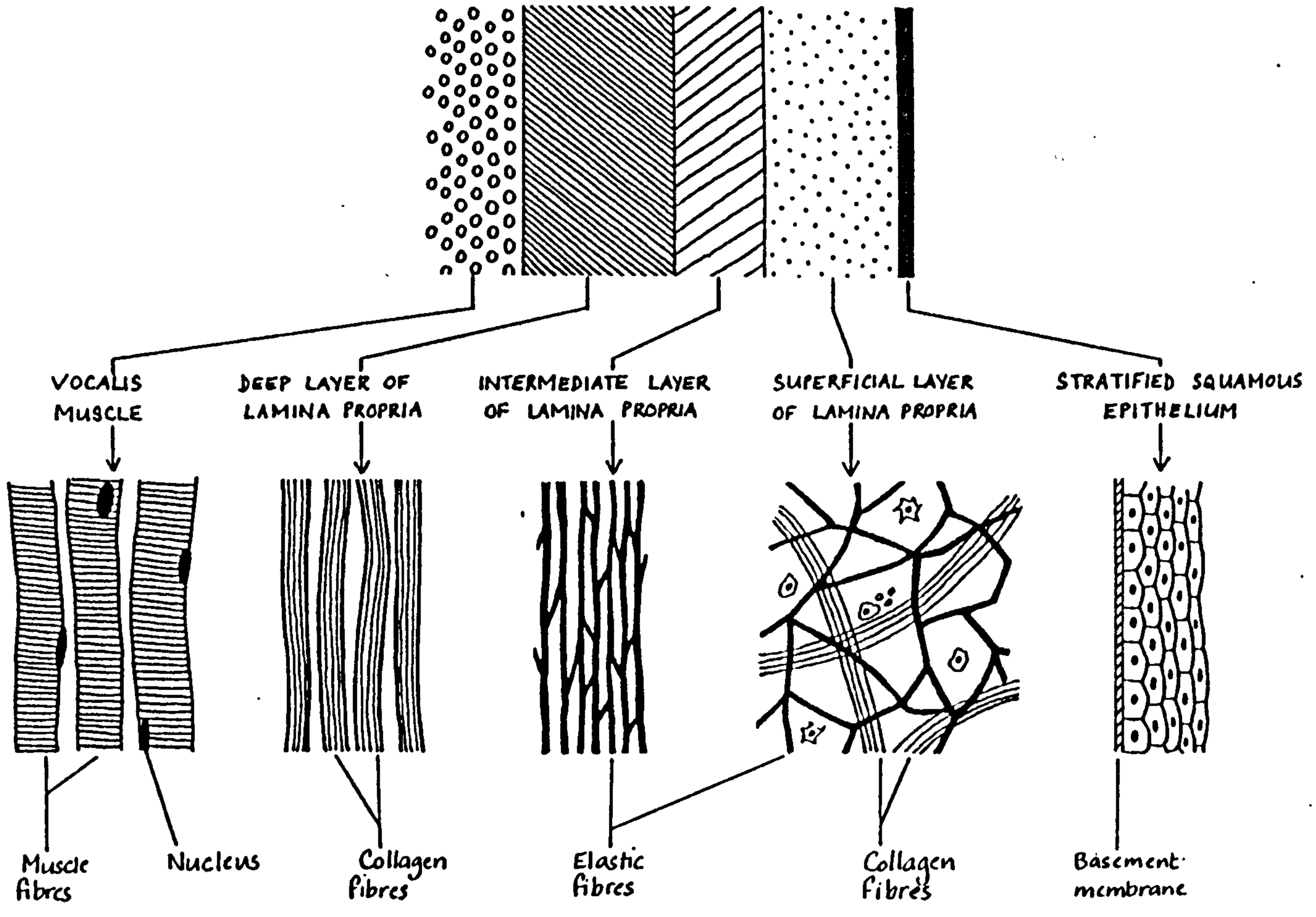
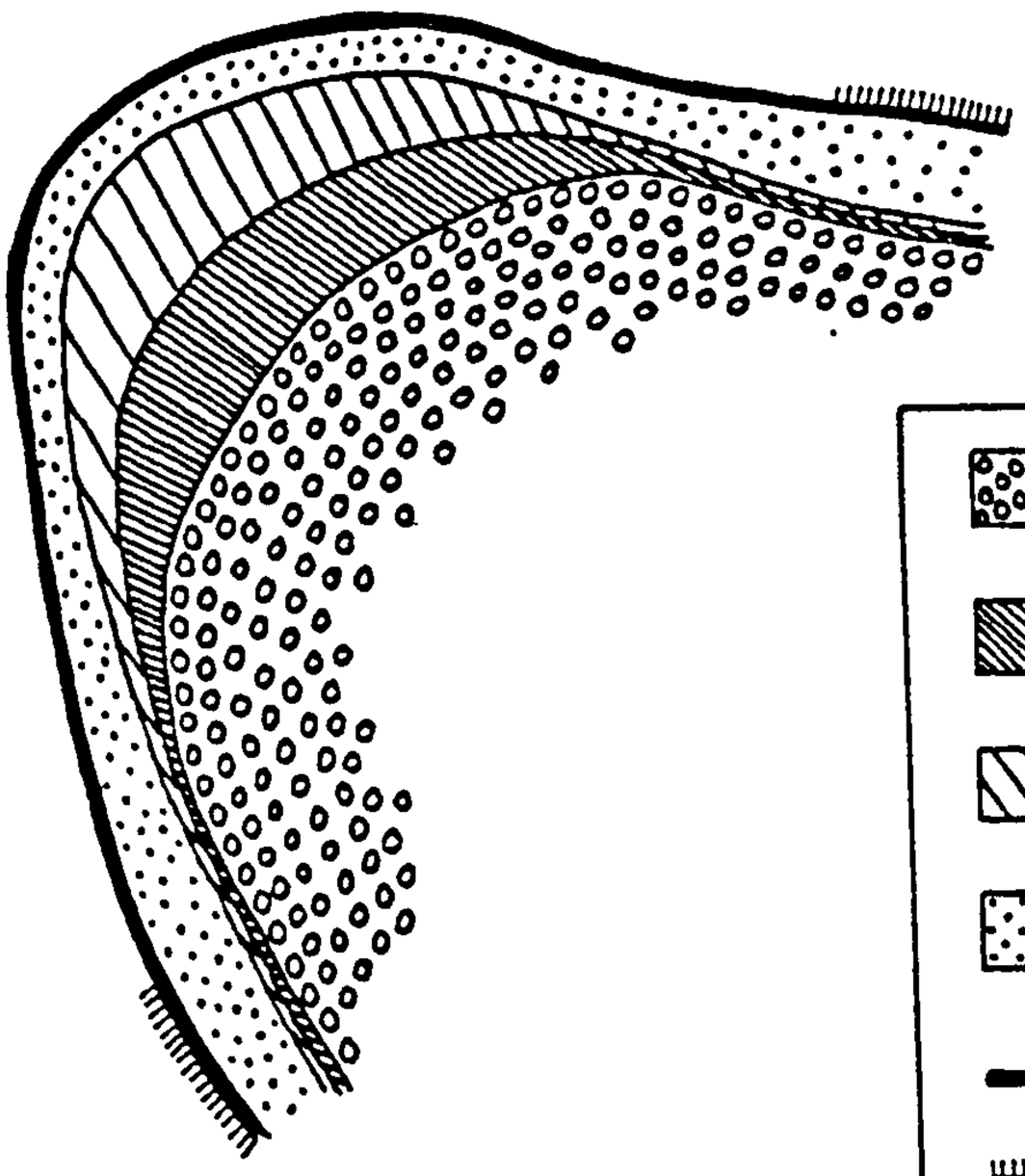


FIGURE 1.1.3/2: A diagrammatic representation of the tissue layers of the vocal folds.



= Vocalis

= Deep layer of lamina propria

= Intermediate layer of lamina propria

= Superficial layer of lamina propria

= Squamous epithelium

= Ciliated columnar epithelium

(Adapted from Hirano 1981 and Hirano et al 1982)

FIGURE 1.1.3/3: A schematic representation of the ligamental portion of the vocal fold, seen in cross section.

On the upper and lower surfaces of the vocal fold there is a transition to ciliated columnar epithelium (see Figure 1.1.3/3).

The epithelium of canine vocal folds, which appears histologically to be very similar to that in humans, has been tested mechanically by Hirano and his colleagues (Hirano et al. 1982), and it seems to be a relatively stiff, non-elastic tissue. In other words, compared with the underlying lamina propria, it requires greater stress to stretch it by a given amount. It is assumed, because the cells do not show any directionality in their arrangement, that the tissue will be isotropic. That is, it will be equally easy (or difficult) to stretch it in longitudinal or transverse directions.

Lamina propria

The lamina propria consists of three layers of connective tissue, which differ in fibre content and arrangement, and hence in mechanical properties.

a) Superficial layer of the lamina propria

The layer of the lamina propria lying immediately beneath the epithelium consists of areolar tissue. Cells are embedded in a soft, semi-fluid matrix, which contains a loose network of haphazardly arranged elastic and collagen fibres. Hirano (1981:5) likens this tissue to soft gelatin, and it is probably the most pliable of the vocal fold tissues. Titze (1973), in his mathematical model of vocal fold vibration, assumes that it behaves like a fluid. Unfortunately the experiments by Hirano et al. (1982) on canine lamina propria do not allow extrapolation to human tissue, because the lamina propria of the dog does not exhibit a comparable three-layered structure.

An alternative name for this layer, Reinke's space, signals that this is a potential site for loss of the normally tight attachment of the mucous membrane to the vocalis muscle.

b) Intermediate layer of the lamina propria

The next layer of connective tissue has a much higher fibre content. These are mostly elastic fibres, arranged in an orderly fashion so that they run parallel to the free border of the vocal fold (i.e. anterior to posterior). These fibres are primarily responsible for the mechanical properties of the tissue as a whole. Hirano's analogy between elastic fibres and rubber bands (1981:5) highlights, as does their name, their marked elastic properties. Freeman and Bracegirdle (1967:20) describe them as having considerable elasticity and little tensile strength. Fields and Dunn (1973) report that they are three times easier to stretch than collagen fibres. In other words, three times less stress is needed to produce an equivalent length increase. The parallel arrangement of the fibres is assumed to cause considerable anisotropy. The tissue is also assumed to be incompressible (Titze 1973).

c) The deep layer of the lamina propria

This layer, which lies next to the vocalis muscle, is similar to the intermediate layer in being rich in fibres which lie parallel to the edge of the vocal fold. In this layer, however, the fibres are mostly formed from collagen, so that the mechanical properties of the tissue are rather different. Hirano's analogy here is with cotton thread, emphasising a high degree of flexiblity allied with relatively low elasticity (Freeman and Bracegirdle 1967, Fields and Dunn 1973). Like the intermediate layer, the deep layer is assumed to be anisotropic and incompressible.

The vocalis muscle

The body of the vocal fold is composed of part of the thyro-arytenoid muscle, the vocalis, which is made up of ordinary skeletal muscle. In spite of controversial suggestions by Goerttler (1950) that the vocalis muscle fibres (cells) run at an angle to the edge of the vocal fold, it is now generally accepted that they in fact run parallel to the edge of the fold.

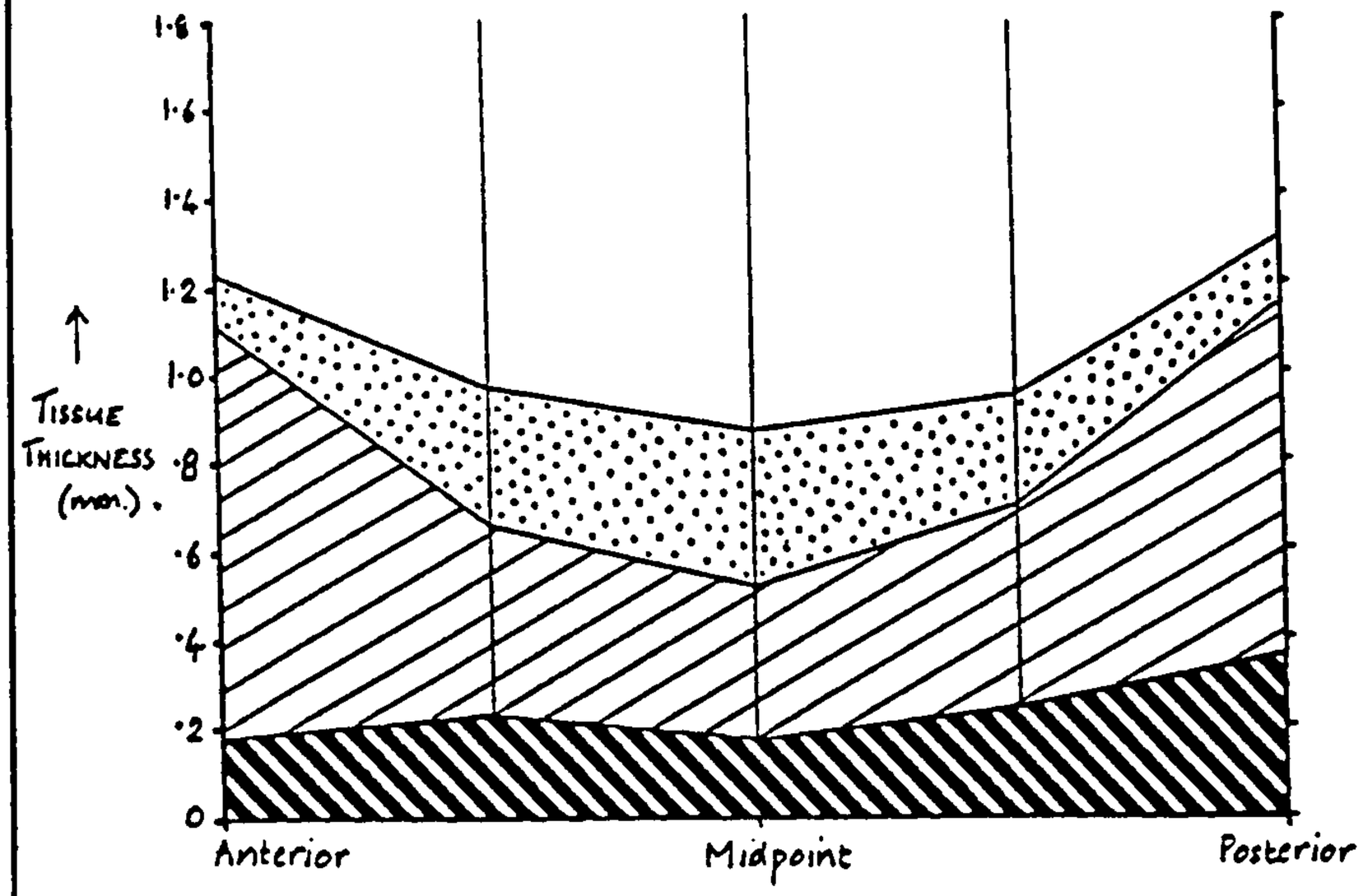
The mechanical properties of muscle vary dramatically, depending on its state of contraction. Hill (1970), cited by Hirano et al. (1982), suggests as much as a tenfold difference in elasticity between resting and contracted muscle. Resting muscle from the canine vocal fold is easier to stretch than either the lamina propria or the epithelium, but like them it is assumed to be incompressible. Anisotropy is also expected, because of the parallel arrangement of the muscle cells.

The vocal ligament

Figure 1.1.3/3, which represents a cross section of the vocal fold at the midpoint of the ligamental area, shows the uneven distribution of these tissue layers. Over the upper and lower surfaces of the vocal fold the intermediate and deep layers of the lamina propria are very thin, but at the glottal edge they become greatly thickened, and constitute the part known as the vocal ligament.

The relative thicknesses of the layers of the lamina propria vary also along the length of the vocal ligament. The superficial layer is thinner at the ends than in the middle, whilst the intermediate, elastic layer is thicker at the ends (Hirano 1981:7, Hirano et al. 1982:276). Figure 1.1.3/4 represents the author's calculations for

1. Females



ii. Males

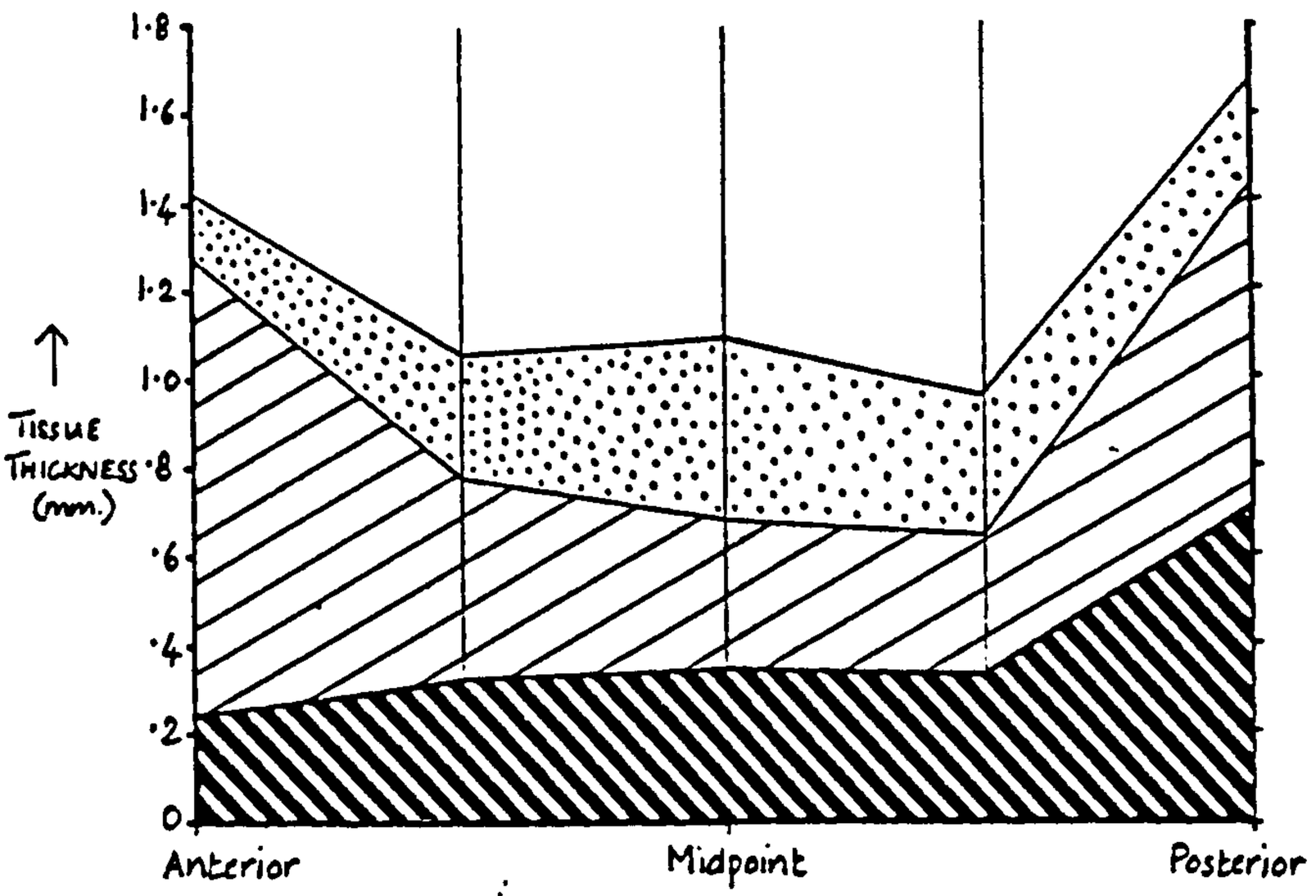


FIGURE 1.1.3/4: A graphic representation of tissue thickness variation along the glottal edge of the ligamental portion of the vocal fold (using data for subjects aged 20-29 years given in Hirano et al. 1982:274).

longitudinal variations in tissue thickness based on data presented by Hirano et al. (1982:275) for five females and five males. This is a rather small sample, but the figures can probably be accepted as being illustrative of general tendencies.

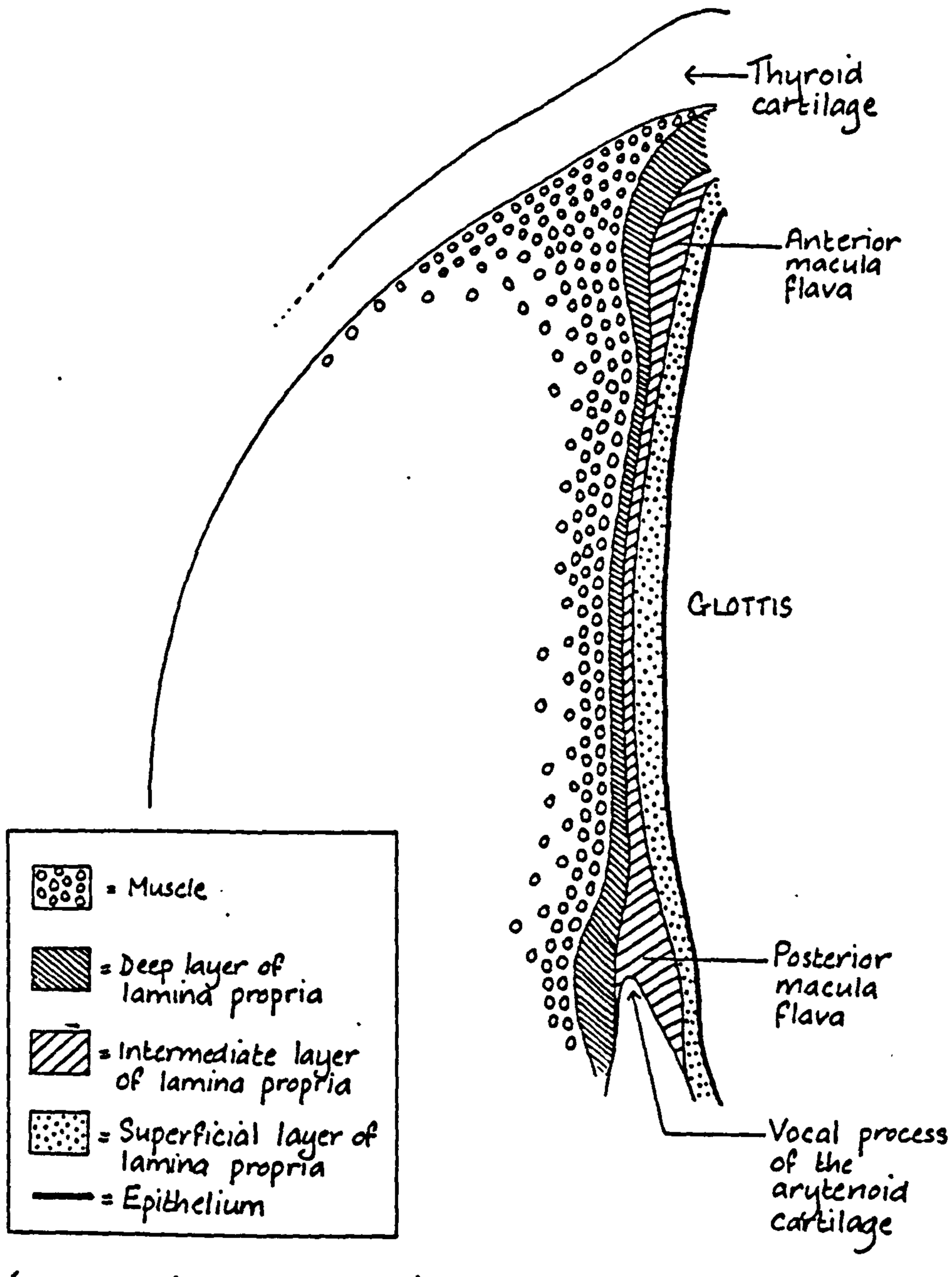
Figure 1.1.3/5 shows that the intermediate layer of the lamina propria is greatly thickened in a small area at each end of the vocal ligament. These thickened areas, which are known as the anterior and posterior maculae flavae, act as cushions of elastic material, and probably afford some protection against impact during vocal fold vibration. The reduced depth of collagen and elastic fibres at the centre of the ligamental portion of the fold increases pliability in this area.

A summary of the mechanical properties of vocal fold tissue and their consequences for vocal fold vibration

Given the preceding description of vocal fold structure, it is now possible to summarise the mechanical properties of each tissue type, and to consider how they might interact during vibration. Figure 1.1.3/6 summarises the tissue properties which have already been discussed.

Independence of tissue layers

The picture so far is of a structure with clearly defined layers, separated from each other by well marked boundaries, but this is something of an oversimplification. The extent to which tissue layers are actually differentiated and kept separate from each other has important implications of two kinds. Firstly, it is relevant to the mechanical independence of each layer, and secondly, it is relevant to the ease of spread of infection and pathological change from one layer to



(Adapted from Hirano 1981)

FIGURE 1.1.3/5: A diagram of the vocal fold in horizontal section, to show the maculae flavae (adapted from Hirano 1981:6).

TISSUE LAYER		ANISOTROPY		TENSILE STIFFNESS	
		Canine	Human	Canine	Human
EPITHELIUM		-	-	high*	high
LAMINA PROPRIA	SUPERFICIAL	+*	-	moderate*	(fluid)
	INTERMEDIATE		+		low
	DEEP		+		high
VOCALIS MUSCLE		+*	+	LOW *(relaxed) → HIGH	LOW (relaxed) → HIGH

*Indicates an entry based on experimental evidence from canine tissue. Remaining entries are based on information about histological structure of the tissues, or on reports of tissue behaviour during vocal fold vibration.

FIGURE 1.1.3/6: A summary of the mechanical properties of vocal fold tissue.

another. The importance of tissue boundaries in limiting disease will be mentioned again in Chapter 2.5.

It is a reasonable assumption that two tissue layers are more likely to behave independently of one another during vocal fold vibration if they fulfil two basic criteria:

i. they should exhibit clearly different mechanical properties, and
ii. there should be a rapid transition of mechanical properties at the border between the two tissues.

It can be seen from the mechanical properties of the tissues as they have already been described that each of the five tissue layers differs from its neighbours in at least one mechanical parameter. The question of transitions and interconnections between the tissue layers now needs to be addressed.

a) Epithelium/lamina propria

The basement membrane of the epithelium forms a well defined boundary between the tightly packed cells of the epithelium and the gelatinous superficial layer of the lamina propria, so that both the suggested criteria for mechanical independence are fulfilled. The epithelium is, however, very thin, and the connective tissue layer behaves as a fluid. Titze (1973) suggests that the two layers do, therefore, act in concert, with the epithelium mimicking the effect of a high surface tension.

b) Superficial/intermediate layers of lamina propria

Hirano et al (1982:274) report that there is generally a clearly marked and rapid transition between these two layers. There is a very dramatic difference in mechanical properties between the fluid or semi-fluid superficial areolar tissue of the superficial layer and the much denser, anisotropic tissue of the intermediate layer. A

fairly high degree of mechanical independence may therefore be expected.

c) Intermediate/deep layers of lamina propria

In the same study, Hirano et al. found that the border between elastic and collagen tissue is not so well defined. There is a gradual transition, with an intervening area where elastic and collagen fibres occur in equal numbers. In spite of their different mechanical properties these two layers are not, therefore, likely to act truly independently.

d) Deep layer of the lamina propria/vocalis muscle

Skeletal muscles are typically contained within connective tissue sheaths, and although there may be some continuity between the collagen tissue of the lamina propria and the enclosing sheath, the muscle is clearly delimited and separated from the lamina propria. The degree of disparity in mechanical properties of collagen and muscle tissue depends on the contractile state of the muscle. The mechanical characteristics of the collagen tissue are relatively invariable, but the tensile stiffness of the muscle may show wide fluctuations. It is probable that under at least some conditions of muscular contraction these two layers are sufficiently different to act with a degree of independence.

Many researchers have noted that a travelling wave can be observed on the surface of the vocal fold (Farnsworth 1940, Smith 1956, cited in Laver 1980:98, Berg et al. 1960, Perello 1962, Hiroto 1966, Baer 1973, Titze and Strong 1975, Broad 1977). This ripple-like mucosal wave can be taken as an indication that at least the outer two layers of the vocal fold (the fluid-like superficial layer of the lamina propria and the epithelium) are acting relatively independently of the deeper tissues.

It may be useful to examine some approaches to mathematical modelling of vocal fold vibration in the light of the above comments on tissue biomechanics. Workers in this field have been conscious for some time of the need to consider at least two semi-independent masses when modelling cross sectional movement of the vocal fold (Ishizaka and Flanagan 1972, Titze 1973, 1974). Ishizaka and Flanagan (1972:1235) comment that "a two-mass approximation can account for most of the relevant glottal detail, including phase differences of upper and lower edges". Titze's model further subdivides the mass of each vocal fold longitudinally into 8 individual sections (see Figure 1.1.3/7). One of the suggested virtues of this sixteen-mass model is that it allows simulation of longitudinal variations in mass and stiffness, and so can simulate some of the effects of vocal fold pathologies. The shortcoming of both the Ishizaka and Flanagan and the Titze models is that they are not capable of modelling the differential effects of changes in the intermediate and deep layers of the lamina propria and the vocalis muscle, because all these layers are represented by a single mass. In a later paper Titze and Strong (1975) do, indeed, conclude that a more accurate model would require at least three masses in cross section. This conclusion is supported by Hirano and his associates (Hirano 1981, Hirano et al. 1981, 1982). Whilst they do not offer a comparable mathematical model, they do stress the need to discriminate between three mechanically different tissue groupings. The five tissue layers which make up the vocal fold are regrouped as follows:-

Epithelium
Superficial layer of lamina propria = COVER

Intermediate layer of lamina propria
Deep layer of lamina propria = TRANSITION LAYER

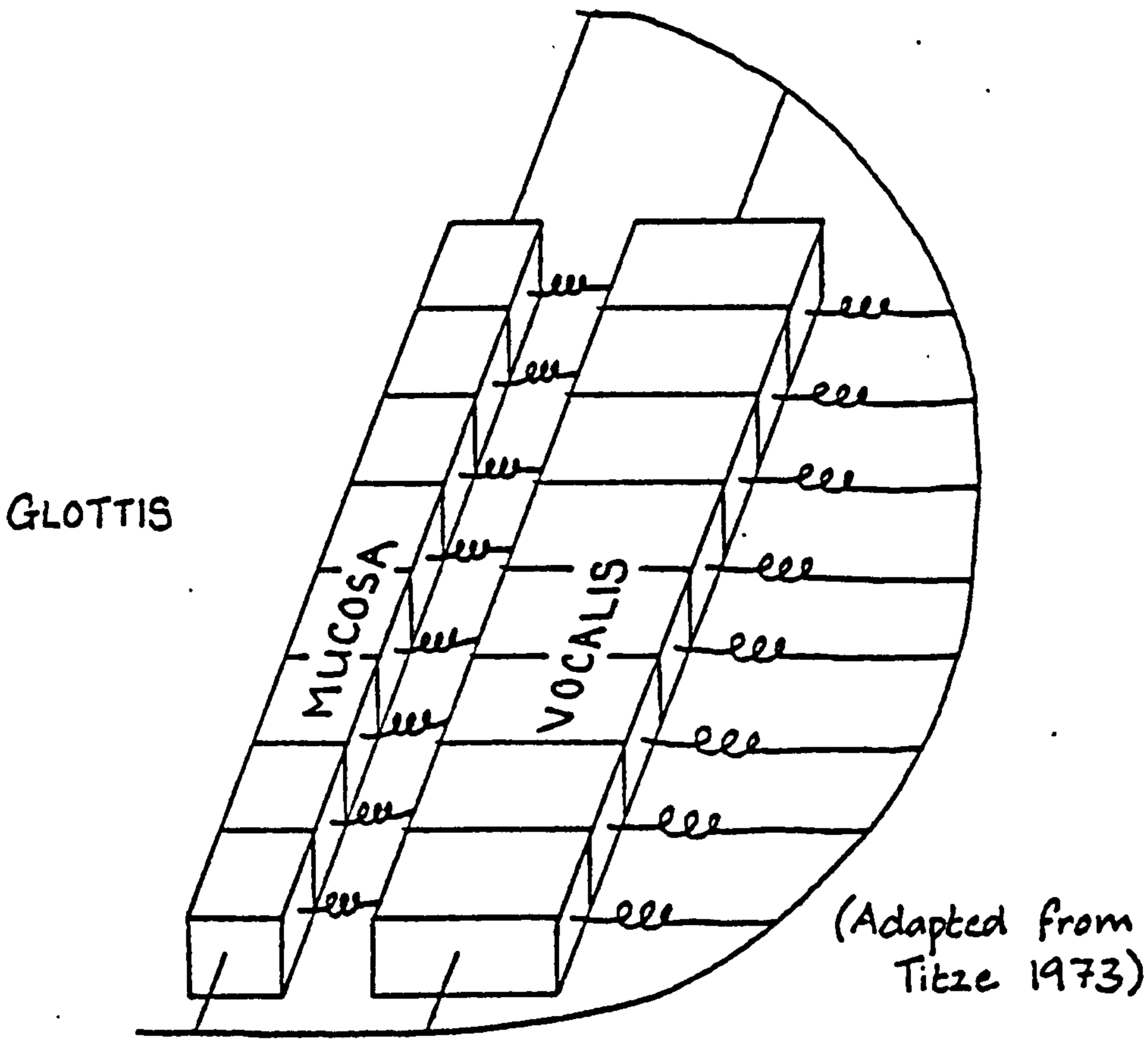


FIGURE 1.1.3/7: A diagram of Titze's (1973, 1974) sixteen-mass model of vocal fold vibration (adapted from Titze 1973).

Vocalis muscle = BODY

This grouping relates well to the expectations of mechanical independence discussed above, and the three-mass system offers a framework for classification of vocal fold pathologies which will be used in Chapter 2.5 as the basis for predictions about the acoustic consequences of organic change.

B. The cartilaginous area of the vocal fold

Much of what has been said about the ligamental area of the vocal fold applies equally to the cartilaginous area of the fold, which is also built up from a series of tissue layers. The body in this area, however, includes the arytenoid cartilage, and is therefore much more rigid. The mucous membrane covering is roughly similar to that covering the ligamental area, with one major difference: because the cartilage lends rigidity to the edge of the vocal fold, taking the place of the vocal ligament, there is no modification and thickening of the intermediate and deep layers of the lamina propria.

This area of the vocal fold is much less freely involved in vibration than the ligamental area, so that organic disruptions of the tissues may have minimal consequences for phonation. Most important in terms of acoustic output will be the inhibition of approximation by any mass protruding into the glottis.

1.2 PRINCIPLES OF GROWTH AND CHANGE

1.2.1 GROWTH MECHANISMS DURING DEVELOPMENT

The task of fully elucidating the mechanisms and control of growth during human development is immense. From an apparently simple cell, the fertilized ovum, grows an adult human, containing many millions of cells, which are organized to form a body capable of performing a multitude of complex activities. Many of these adult cells are highly differentiated, and bear little superficial resemblance either to the ovum from which they originate or to other types of differentiated cells. The means by which coordinate growth of cells and tissues is controlled and patterned to produce the adult body is still not fully understood, although the last twenty years have seen much progress in this field (Goldspink 1974, Sinclair 1978, Tanner 1978, Falkner and Tanner 1986). All that will be attempted in this chapter is a brief summary of the growth processes which are known to play major roles in determining the architecture of the human vocal tract at different stages in human development.

A. Measuring growth: growth curves

Methodologically, the simplest way to look at growth is to measure gross overall changes in size, such as height or weight. Figures 1.2.1/1 and 1.2.1/2 show standard growth curves for height and weight in British children, using data taken from Tanner and Whitehouse (1976). There is considerable individual variation, but typically a gradual deceleration of growth between birth and adulthood is interrupted by a period of accelerated growth which occurs round about the time of puberty. This is seen more clearly in growth velocity curves, which

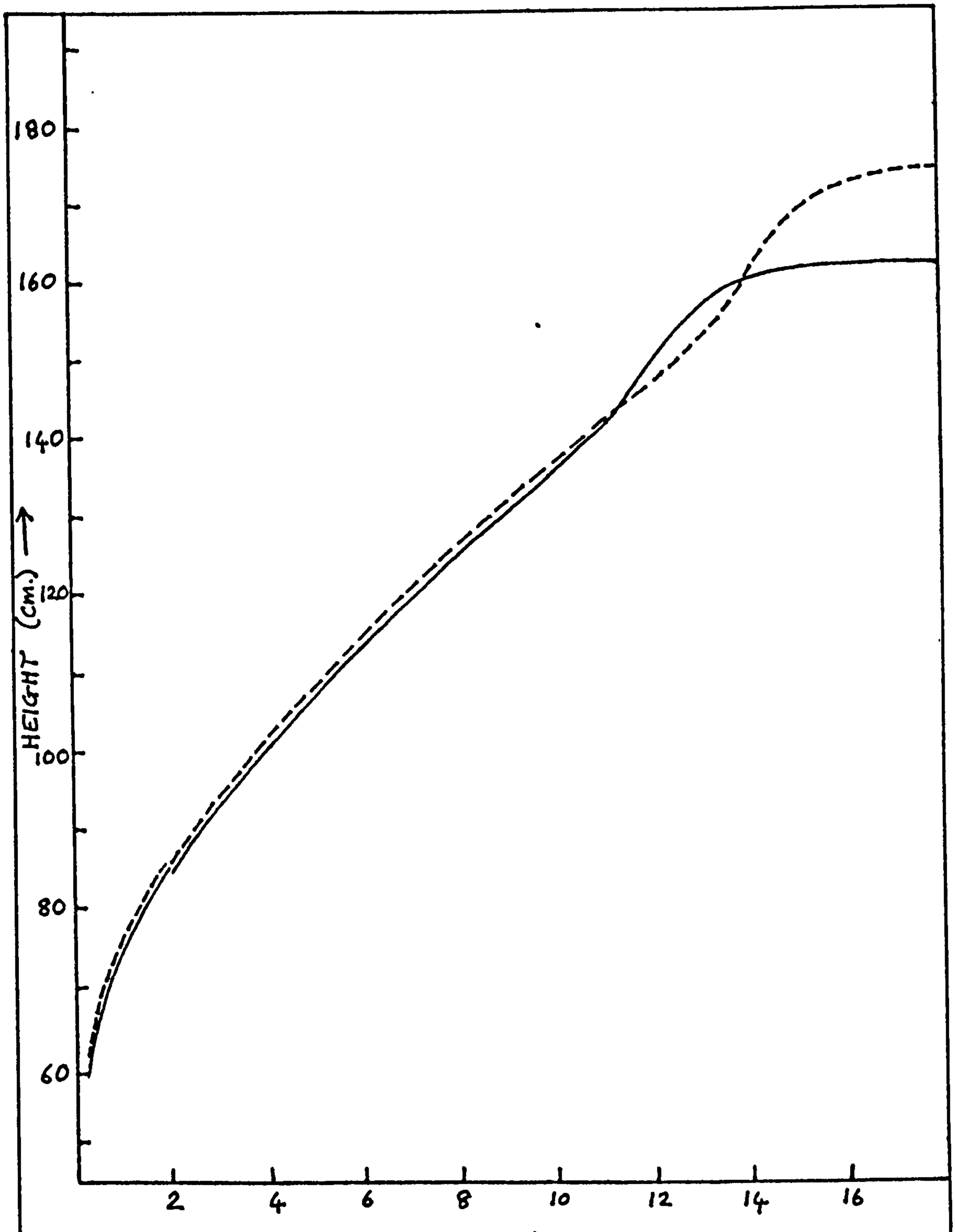


FIGURE 1.2.1/1: Standard height growth curves for British children (adapted from Tanner 1978:169,170)

——— = Girls - - - - - = Boys

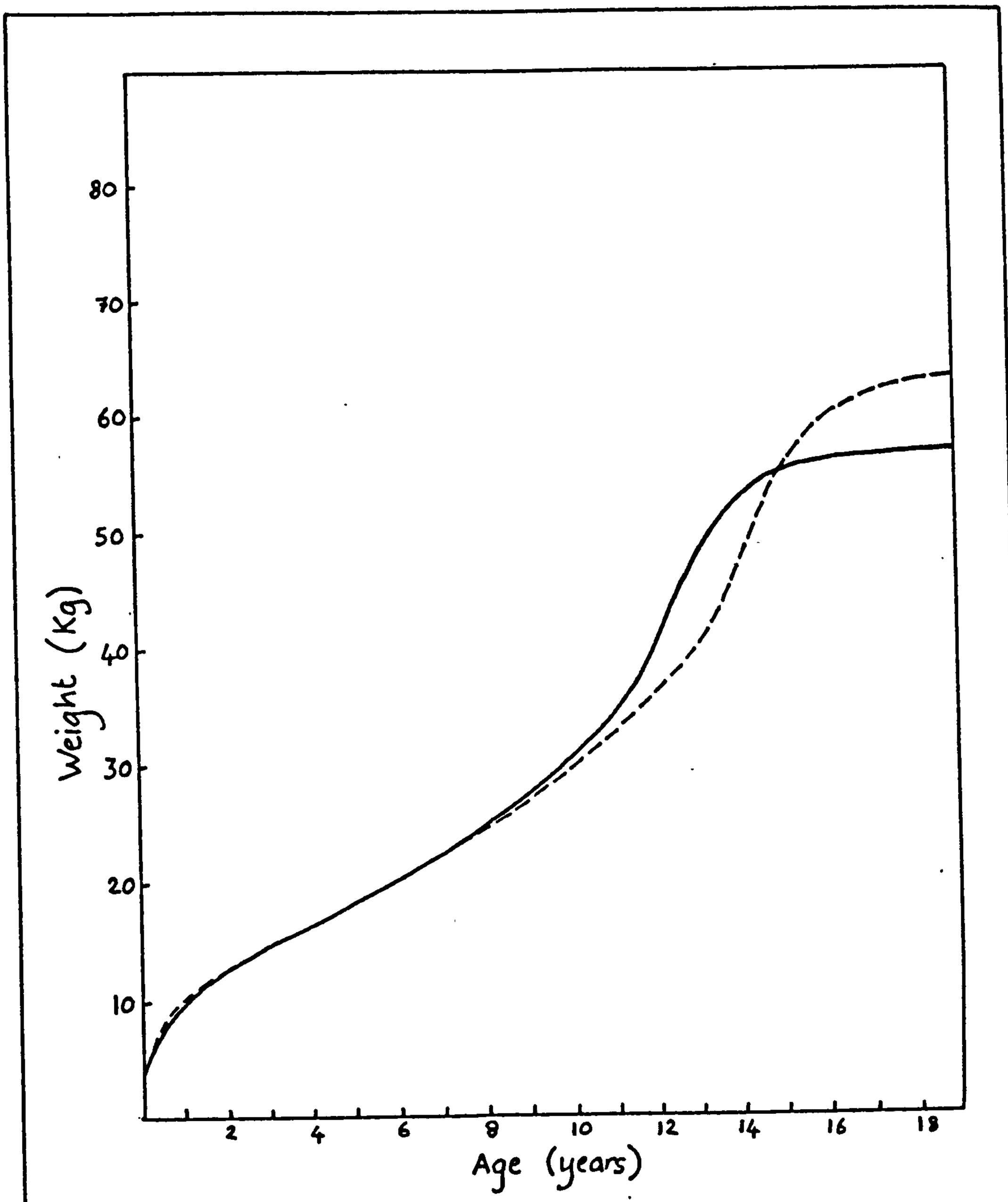


FIGURE 1.2.1/2: Standard weight growth curves for British children (adapted from Tanner 1978:180,181)

graph the same data in terms of rate of size increase (see Figures 1.2.1/3 and 1.2.1/4).

Boys are slightly longer at birth, and remain taller until adolescence. Between 11 and 14 years the female growth curve overtakes the male, because the female adolescent growth spurt occurs about two years earlier than the males'. The male growth curve then regains the lead as the female growth spurt ends and the males' begins. Weight curves follow a broadly similar pattern. The data presented here represents the results of the largest longitudinal study of growth in British children. Similar studies have been conducted elsewhere, and are discussed in some detail in Eveleth and Tanner (1976) and in Marshall (1981). Some geographical and ethnic variations are evident, but the same general trends are apparent in all studies. For example, a comparison of London children and well-off Chinese children in Hong Kong shows that the Chinese children are less tall throughout childhood, and reach adult height earlier than the London children, but that the overall shape of the size and velocity curves is broadly similar (Tanner 1978,137-8).

Whilst this type of overall growth measurement is very valuable, it may be misleading if a particular part of the body such as the vocal tract is the focus of attention. The processes of growth which are brought into play as child changes to adult are highly complex, and an adult is not simply a scaled up version of a child. Figure 1.2.1/5 shows the changing proportions of the body during development using data for head, trunk and limb dimensions taken from Altman and Dittner (1962:333). It is clear that the shape and proportions of the adult body are altogether different from those of the child. This is achieved by intricately coordinated growth processes, and proper development of the adult form demands that some

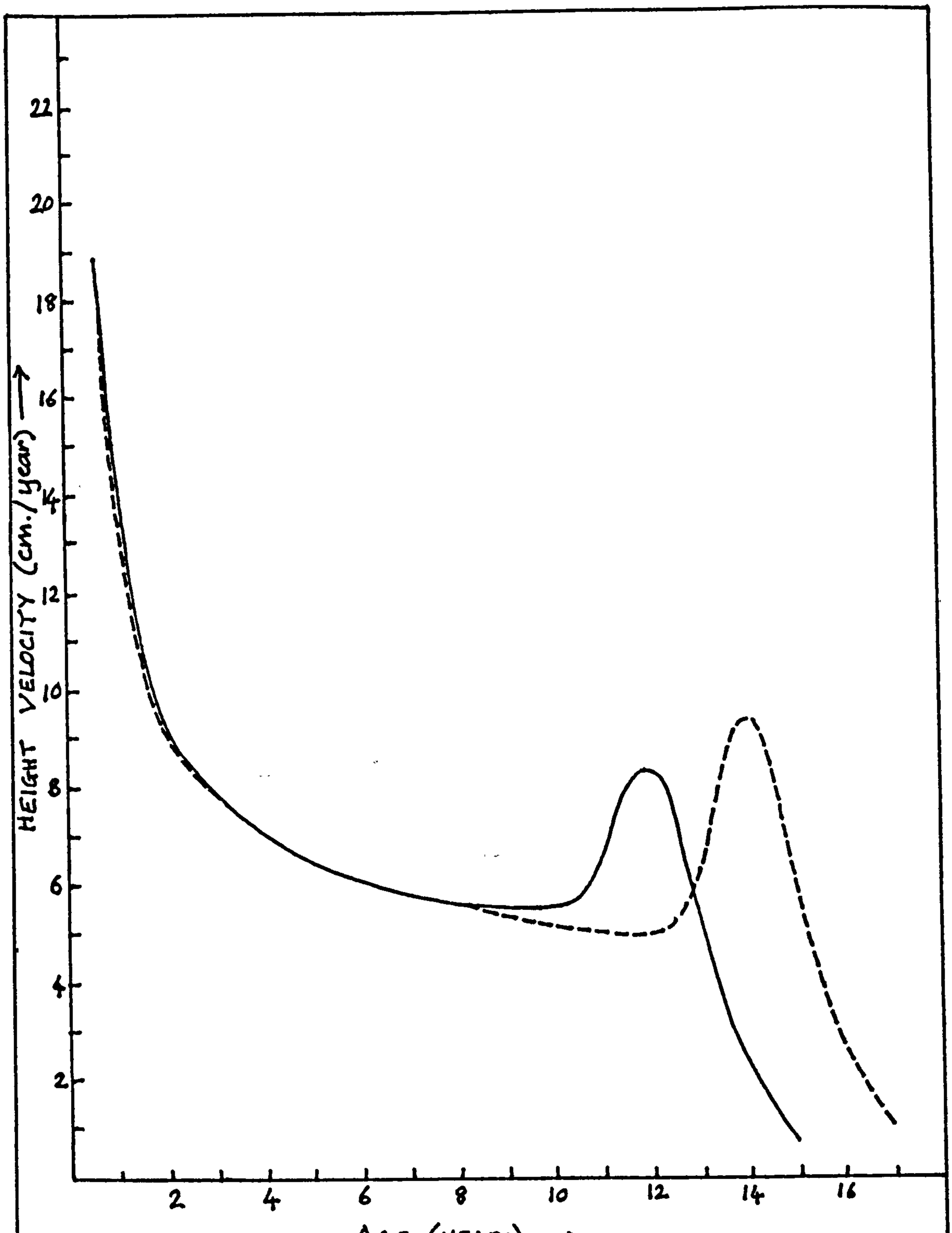


FIGURE 1.2.1/3: Height growth velocity curves for British children (adapted from Tanner 1978:177,178)

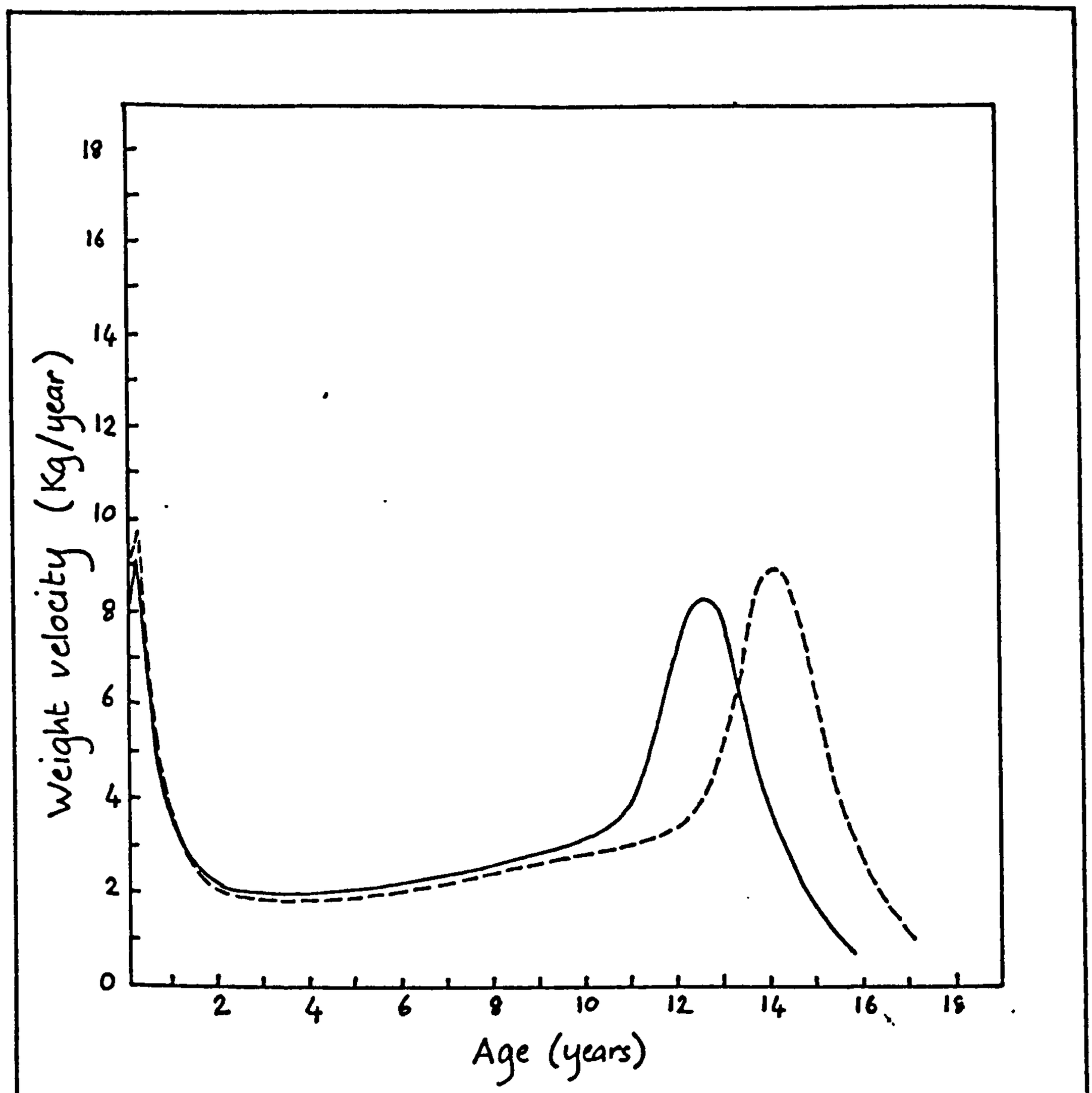


FIGURE 1.2.1/4: Weight growth velocity curves for British children (adapted from Tanner 1978:182-183)

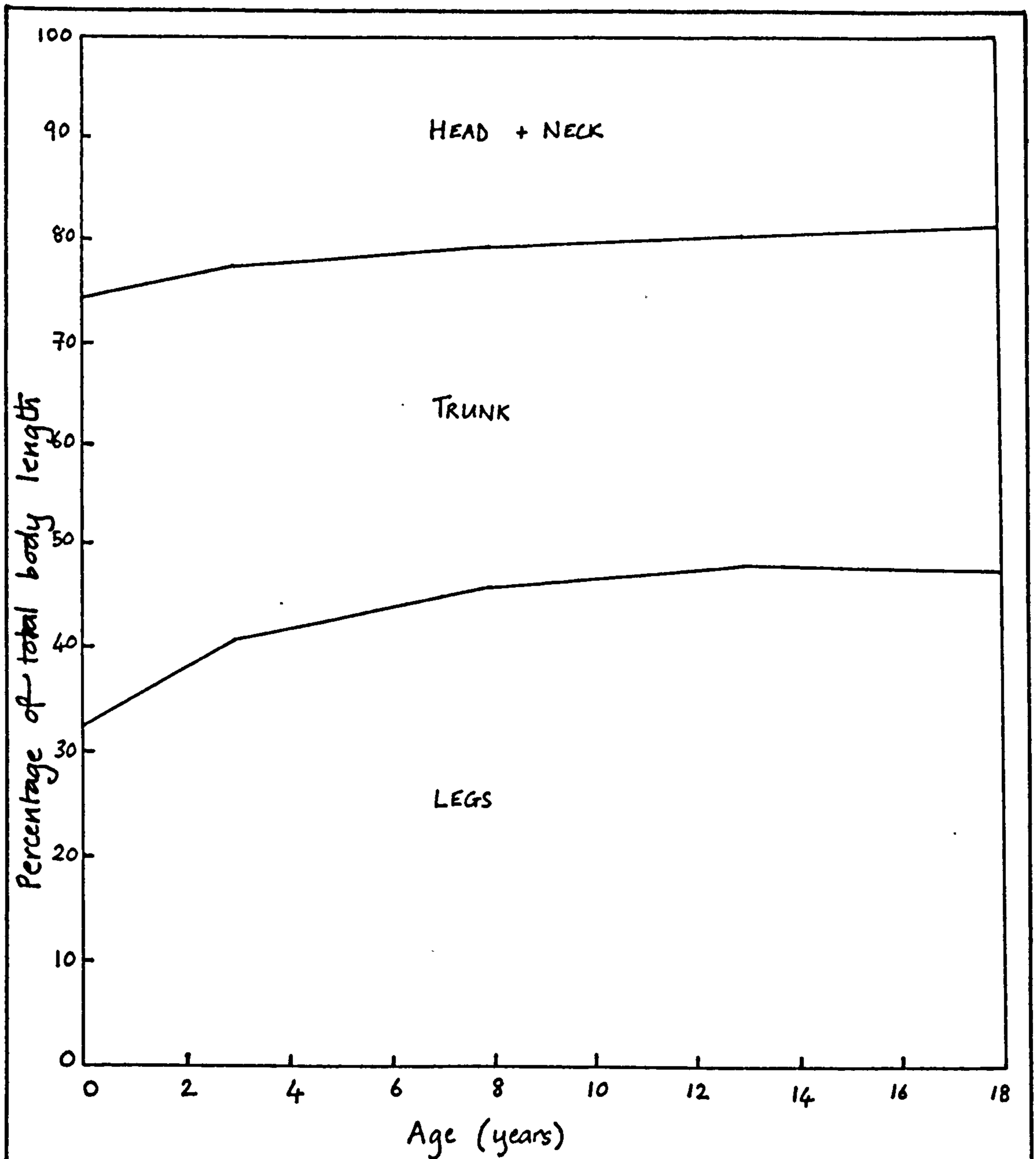


FIGURE 1.2.1/5: Graphic representation of changes in bodily proportions during development (using data from Altman and Dittner 1962:333)

areas and tissues should grow faster, or at different times, from others. As a result, some parts of the body show major deviations from the general growth curve. Tanner (1978:16) shows the different growth curves exhibited by the brain and head, the reproductive tissue and lymphoid tissue. These are shown in Figure 1.2.1/6. The brain and skull develop very early, and 80% of the increase in size from birth to 20 years is achieved by 4 to 5 years of age. Reproductive tissue, in contrast, shows little significant growth before puberty. Lymphoid tissue has a very unusual growth pattern, with a huge increase in size during the first ten or eleven years being followed by a decrease, so that the lymphoid mass at 12 years is about double that at 20 years.

Such differences may be easier to understand if we look at growth at the cellular and tissue level. The relative importance of cellular and intercellular material in the make-up of various tissue types has already been discussed in Section 1.1.2, and it follows from this that growth may occur in three ways: the number of cells may increase, the size of the constituent cells may increase or the amount of intercellular material may increase. Most specialized tissues enlarge by a combination of these processes, with an initial growth phase of rapid cell division being followed by a slowing, and finally a cessation, of cell division and an increase in cell size.

Growth by an increase in cell size is limited by the need for cytoplasmic functions to be controlled by the nucleus (see Section 1.1.2). This determines the maximum size to which a cell can grow. Similarly the maximum quantity of intercellular material which a tissue can contain is limited by the need for cells within that tissue to exchange nutrients and waste products with the blood or lymphatic system.

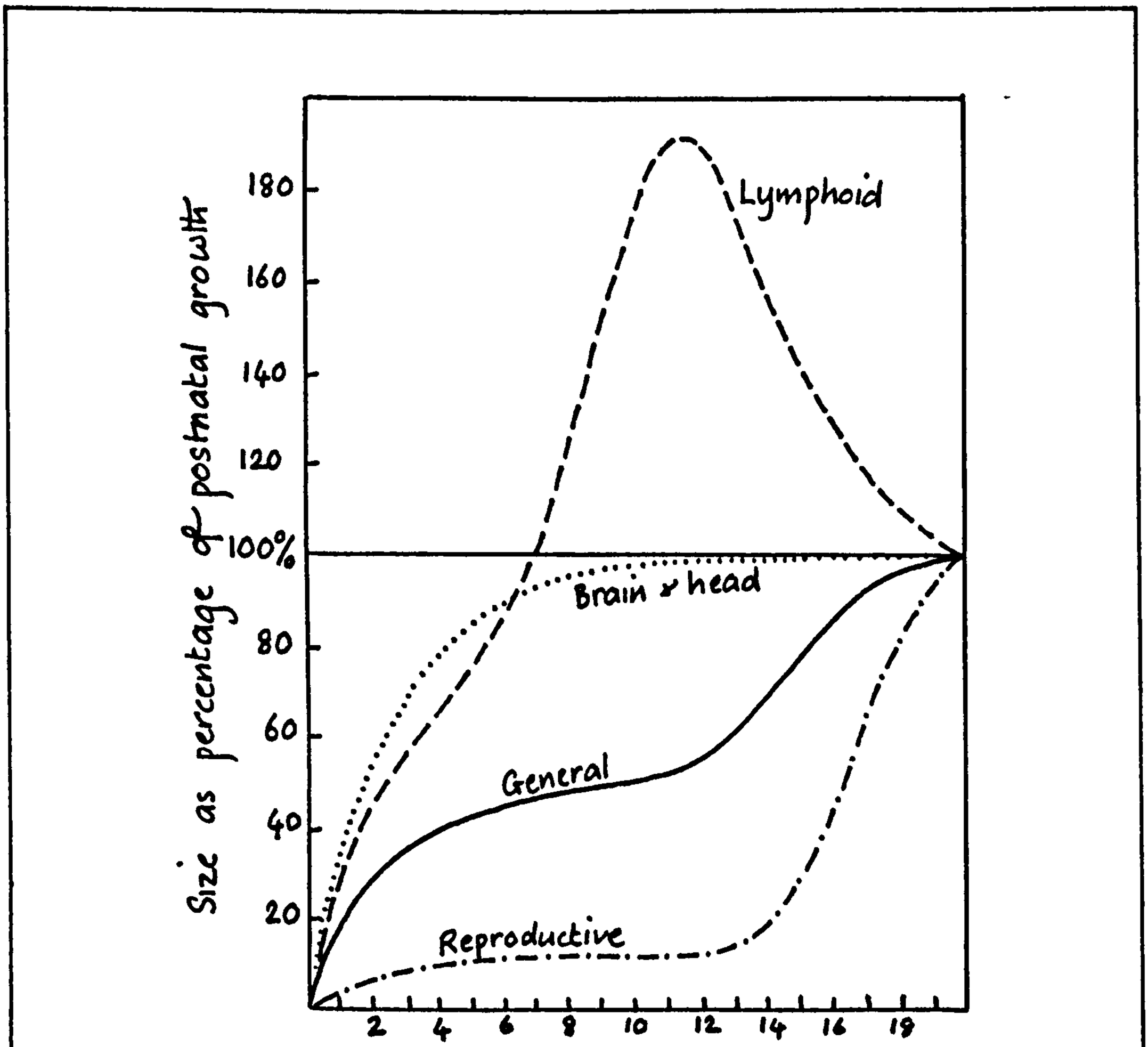


FIGURE 1.2.1/6: Growth curves of reproductive tissue, brain and head, and lymphoid tissue, compared with the general growth curve (redrawn from Tanner 1978:16)

A further convenient distinction may be made between interstitial and appositional growth (Sinclair 1978:4). In interstitial growth an increase in size is achieved by adding cells or intercellular material evenly throughout the tissue. In appositional growth existing tissue remains largely unchanged and new material is concentrated in one area. This is shown diagramatically in Figure 1.2.1/7.

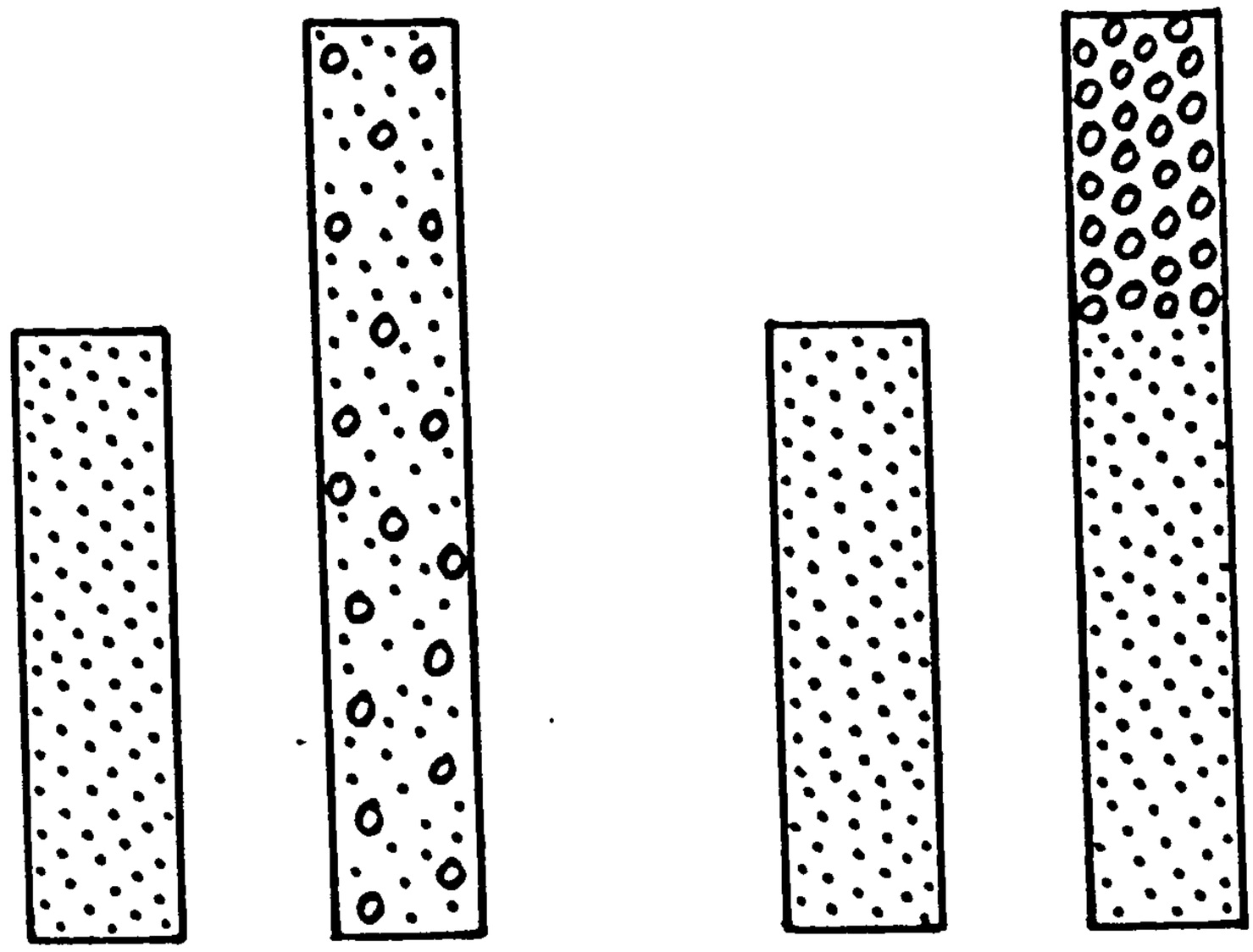
Since the details and timing of tissue growth depend on the function of each tissue type, and on the constraints imposed by the presence of intercellular material, the growth characteristics of the major tissue types which together form the vocal apparatus will be summarised in turn. First, however, it is necessary to consider briefly the mechanism by which cell multiplication occurs.

B. Cell division

Since proper cell functioning depends on the presence of correct genetic information in the form of DNA, accurate copying and redistribution of DNA is a crucial requirement in cell division. The mechanism which ensures that all daughter cells receive a set of genes which is identical to that contained in the parent cell is known as mitosis. The main steps of this are shown in Figure 1.2.1/8. These diagrams show only the changes which are clearly visible using a light microscope, although there are some crucial steps which cannot easily be observed.

Interphase

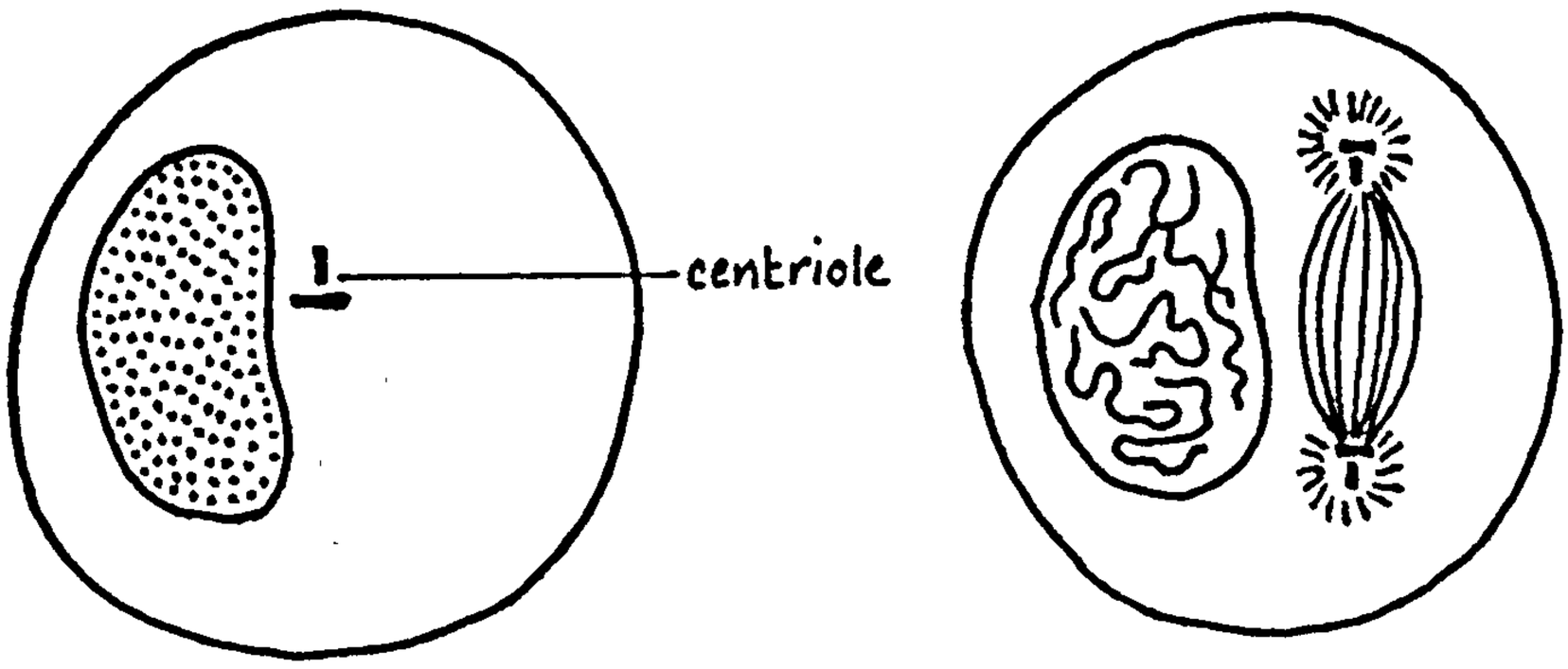
During interphase, when there is no visible sign of cell division, the structure of the cell is as described in section 1.1.2., and the chromosomes are so elongated as to be invisible. It is during this stage that the vital process of DNA replication takes place, so that by the



= Old material

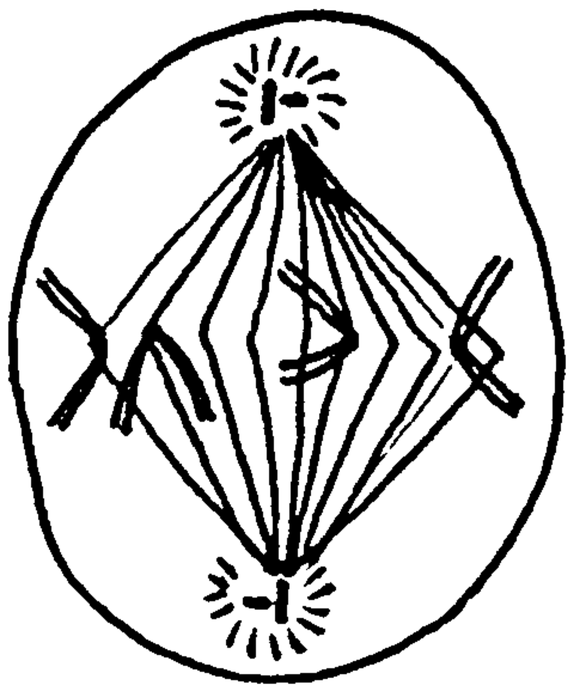
= New material

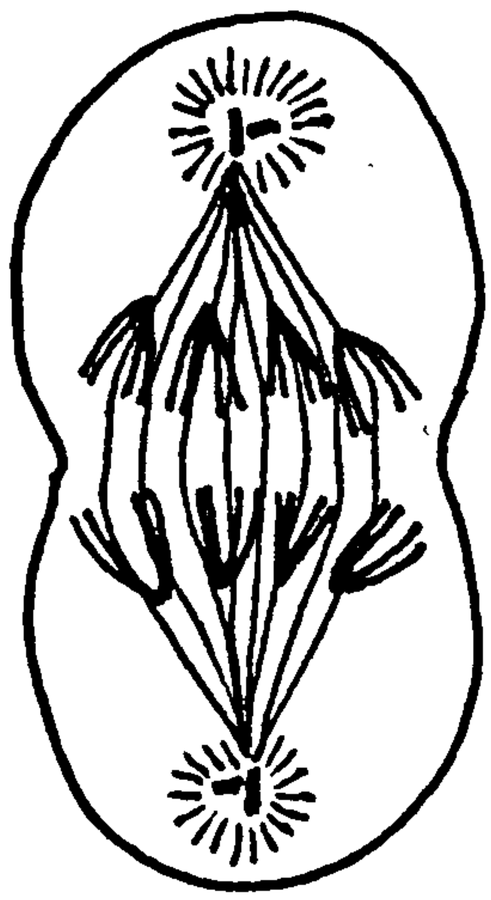
FIGURE 1.2.1/7: Diagrammatic representation of
A. interstitial and
B. appositional growth



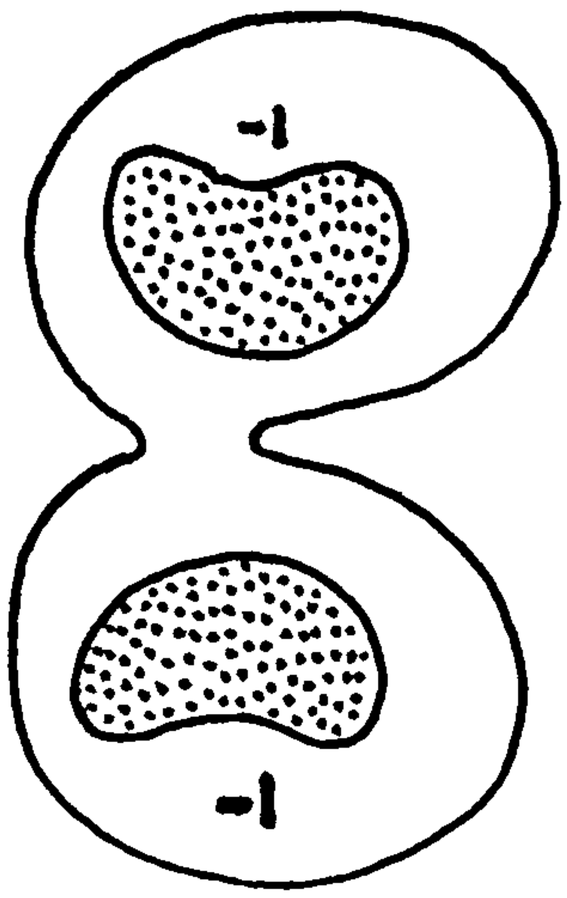
1. Interphase

2. Prophase





3. Metaphase



4. Anaphase

5. Telophase

FIGURE 1.2.1/8: Schematic diagram of mitosis

time the first obvious sign of cell division occurs each chromosome already contains two identical lengths of DNA.

Prophase

During this stage several changes occur.

i) The chromosomes shorten and thicken and eventually become visible as dark rod-like structures. Each one is split longitudinally into two strands, or chromatids, which have a point of close attachment called the centromere. Each chromatid is in fact a fully replicated daughter chromosome, as a result of the DNA doubling which took place during interphase.

ii) The nuclear envelope begins to disintegrate.

iii) The pair of centrioles within the cytoplasm duplicates, and the new pairs then move away from each other and position themselves at opposite ends of the cell. As they move apart microtubules begin to form between the two pairs. These are known as spindle fibres, since as the nuclear envelope disappears they form a spindle-shaped structure of continuous strands, linking the two pairs of centrioles.

Metaphase

During this stage the chromosomes move to the centre of the cell and attach themselves to the equatorial plane of the spindle.

Anaphase

Each pair of chromatids then separates at the centromere, and one member of each pair moves to each end of the cell. In this way the chromatids, which can now be called daughter chromosomes, are distributed so that a complete set of chromosomes (=46) goes to each half of the cell. The cell then begins to constrict around the centre.

EX BIBL UNIV. EDINBURGEN

Telophase

Various steps are completed in this stage which result in two separate cells in interphase condition.

i) The chromosomes detach from the spindle, elongate and become less visible.

ii) The nuclear envelope reforms around the chromosomes, probably using membrane material derived from vesicles breaking off the endoplasmic reticulum.

iii) The spindle fibres disappear, and cytoplasmic contents are distributed equally between the two halves of the cell.

iv) Constriction around the centre of the cell continues until the two daughter cells are separated.

The only exceptions to this pattern of cell division occur during the formation of reproductive cells (ova and spermatazoa). Here, the requirement that each daughter cell contains only one member of each chromosome pair, i.e. 23 chromosomes, necessitates a more complicated process of cell division called meiosis. This need not concern us here.

C. Tissue growth patterns

It has already been indicated that proper development of the adult body demands that some tissues of the body should grow faster, or at different times, from others. The details and timing of tissue growth depend on the function of each tissue type, and upon the constraints imposed by the presence of intercellular material. The growth characteristics of each of the major tissue types which contribute to the architecture of the vocal apparatus will be summarised in turn below.

Epithelial tissue

Covering epithelium

Since the cells which make up covering epithelium are relatively undifferentiated, and there is little intercellular material, growth is achieved simply by the division of existing cells. Covering epithelia undergo a constant process of regeneration throughout life, so that it is not possible to draw a clear distinction between developmental and regenerative growth (see section 1.2.2). The renewal rate of epithelium under normal conditions varies considerably in different geographical sites. In the intestine, epithelium may be renewed every 2-5 days, whilst in tissues such as the pancreas renewal may take 50 days (Junqueira and Carneiro 1980:74). Epithelial growth typically involves mitosis of the germinal cell layer, nearest the basal membrane. In stratified epithelia, cells progress to the surface of the tissue as they mature and age.

Glandular epithelium

All types of glandular epithelium are derived from sheets of covering epithelium, although endocrine glands lose their connection with the epithelial surface. Epithelial cells proliferate locally, and grow down into underlying connective tissue. The formation of various types of gland by this process is shown schematically in Figure 1.2.1/9).

Connective tissue

Connective tissue proper

Little seems to be known about the growth of connective tissue proper, although it is clear that the growth in

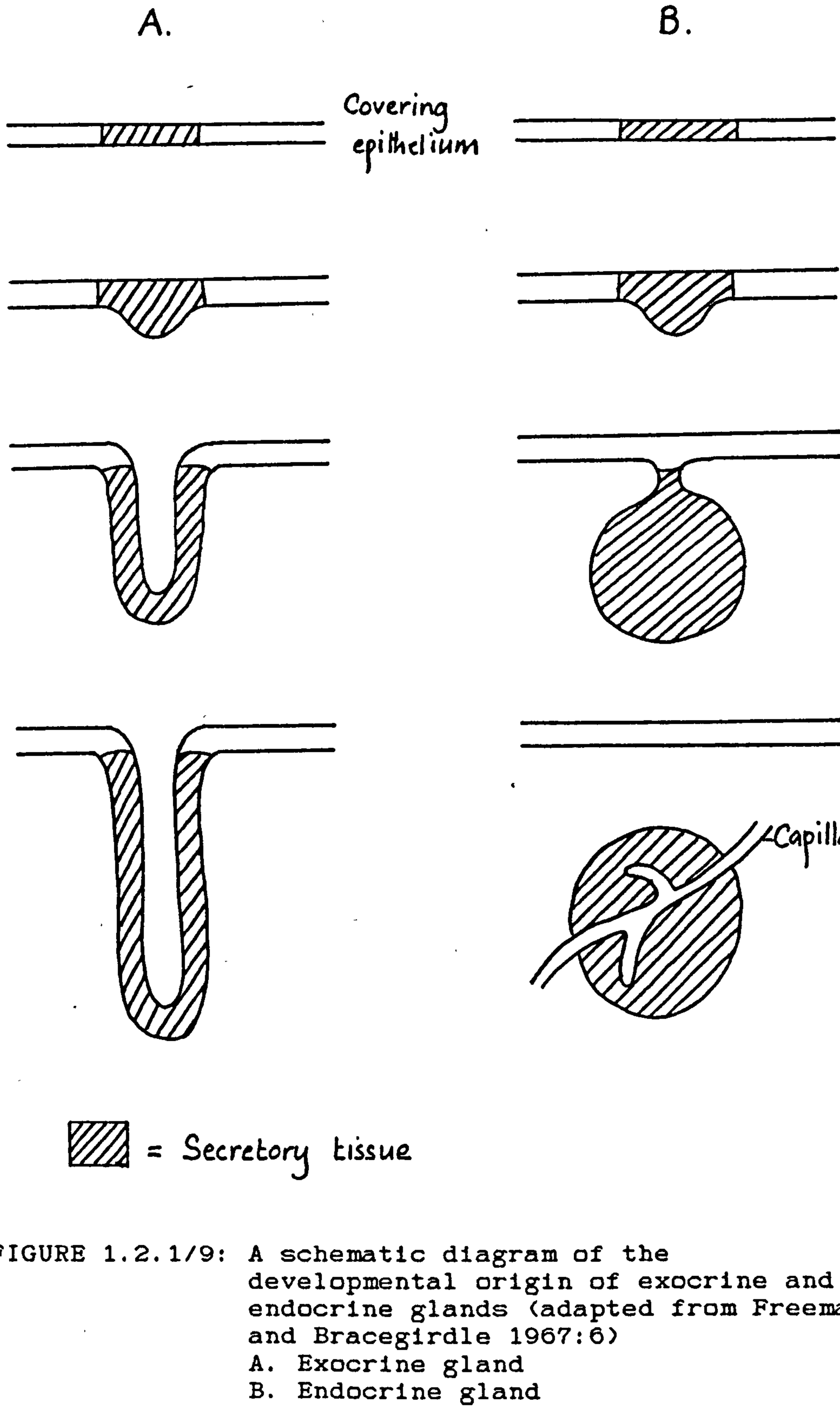


FIGURE 1.2.1/9: A schematic diagram of the developmental origin of exocrine and endocrine glands (adapted from Freeman and Bracegirdle 1967:6)
A. Exocrine gland
B. Endocrine gland

mass of connective tissue may result at least as much from increases in matrix and fibre content as from increases in cell number. It is thought that the amount of collagen which is laid down in a tissue depends on the amount of stress to which the tissue is exposed. The thickness of a tendon, which is determined largely by the number of collagen fibres within it, seems to vary with the magnitude and duration of stress applied to it. Collagen fibres tend to form along the lines of stress, and in studies on wound healing it has been shown that fibroblasts arrange themselves according to the tensional forces acting on the wound (Sinclair 1978:46).

Cartilage

The growth of specialized connective tissues such as cartilage and bone has been more thoroughly investigated, and cartilaginous growth is described in Leeson and Leeson (1976:132-133) and Junqueira and Carneiro (1980:126-127). Three stages are involved in the initial development of cartilage.

i. Undifferentiated cells become rounded, and multiply to form dense clusters of cells, which may then be classed as chondroblasts.
ii. The chondroblasts begin to synthesise matrix materials (fibres and amorphous ground material), which accumulate and begin to separate the cells from each other.
iii. Differentiation of the cartilage tissue progresses from the centre.

This results in a situation in which the cells at the centre are typical chondrocytes, whilst the cells at the periphery are typical chondroblasts. The undifferentiated surface cells develop to form the fibroblast-like cells of the perichondrium. After this initial developmental

stage, growth may progress by either of two processes. Interstitial growth involves the multiplication of existing chondrocytes, resulting in growth within the body of the cartilage as the newly formed cells synthesise new matrix materials.

In hyaline cartilage, this type of growth is important only during the early stages of cartilage development. As cross linkages between collagen fibrils and ground material chemicals increase rigidity of the matrix, growth becomes limited to the second process; appositional growth. This may be seen as a continuation of the initial stage of cartilage formation, as growth proceeds by differentiation of cells within the inner perichondrium. As the resulting fibroblasts are incorporated into the matrix as chondrocytes, and produce new matrix materials, the cartilage increases in size at its periphery.

Cartilage growth is shown diagrammatically in Figure 1.2.1/10.

Bone

Bone tends to be thought of by the layman as a hard and invariable substance. In fact, it shows considerable plasticity during development, in the sense that it is able to modify its size and shape as an adaptation to the stresses imposed upon it. It is, however, a rigid and inelastic tissue, so that changes in size and shape can only be achieved by the deposition and/or removal of surface bone. It should be obvious, following the description of bone tissue in Section 1.1.2, that intercellular material must play a major role in bone growth. Changes in bone architecture are brought about by the concerted action of three types of cell. The cells which are actively involved in manufacture of bony

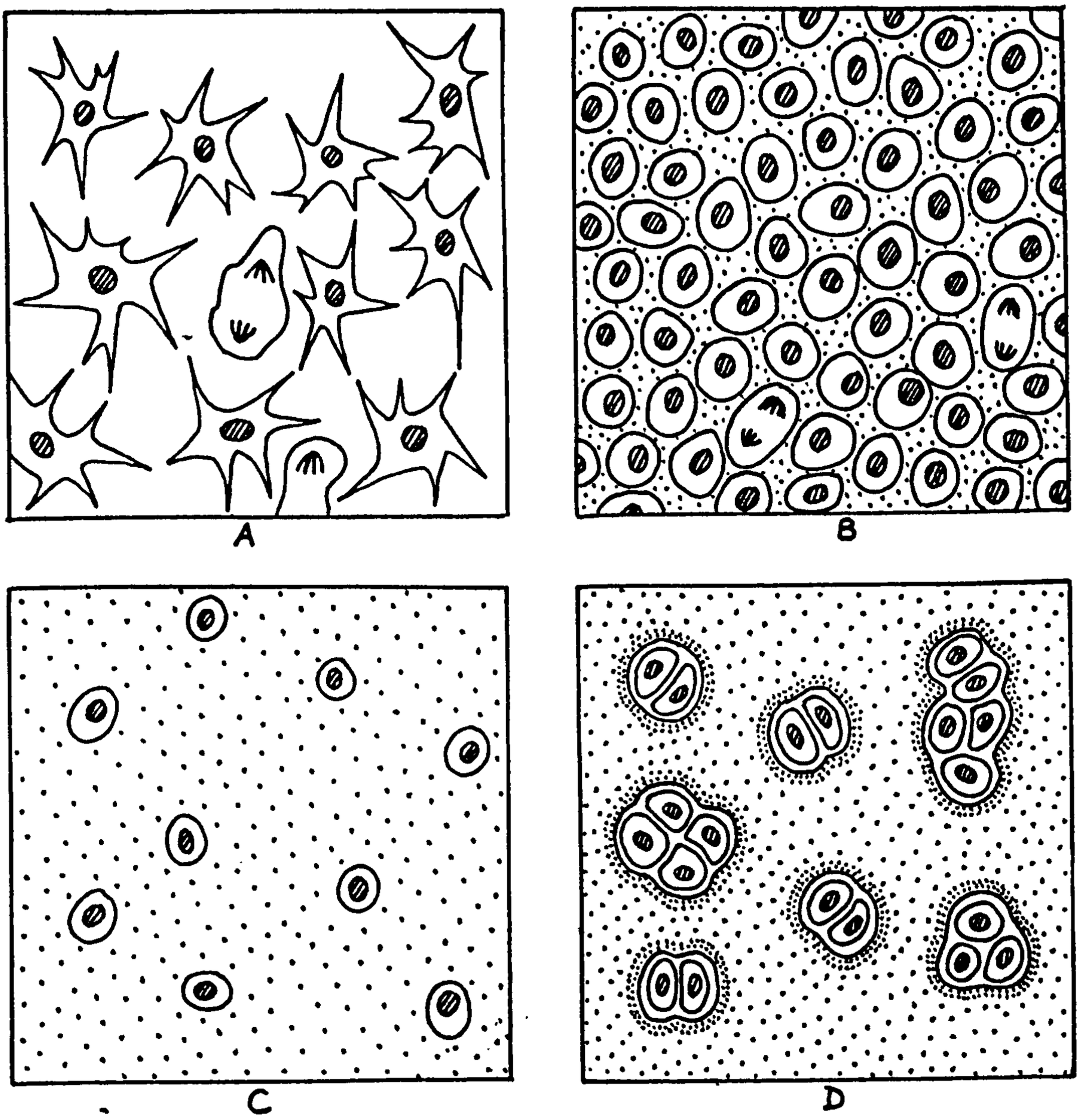


FIGURE 1.2.1/10 Schematic diagram of hyaline cartilage development (adapted from Junqueira and Carneiro 1980:127)

A. Primitive precursor cells.

B. Rapid mitosis leads to high cell density.

C. Cells become separated by large amounts of matrix.

D. Cartilage cells divide to form groups of cells surrounded by capsules of condensed matrix.

material are known as osteoblasts. Cells which are actually trapped within the matrix are known as osteocytes. The multinucleated cells which are responsible for the removal of bony matrix are osteoclasts. The initial formation of bone during development results from the conversion into bone of either fibrous tissue (= intramembraneous ossification) or cartilage (= endochondrial ossification). The development of fibrous membrane, cartilage and bone is therefore a very carefully orchestrated process during human development.

Intramembraneous ossification

Pritchard (1974) gives a very clear account of the early development of membraneous bone (i.e. bone resulting from intramembraneous ossification), using the bones of the face and cranial vault as typical examples. During foetal development of the facial bones, a network of bone matrix trabeculae develops from a framework of fibrous tissue. The growing bone then becomes encapsulated within a layer of collagen fibres, together with the fibroblasts which are responsible for collagen formation. Once this fibrous coat, the periosteum, is fully formed, the pattern of intramembraneous ossification is similar to that seen in post-natal growth.

The periosteum then has two distinct strata. The outer, fibrous layer consists of dense parallel fibres and fibroblasts. The inner, cambial (or osteogenic) layer contains a looser arrangement of fine fibres, blood vessels, osteoblasts, and progenitor cells which are capable of developing into osteoblasts. The bone grows as a result of multiplication of progenitor cells within the cambial layer, some of which are then converted into osteoblasts capable of bone matrix manufacture. Part of the vascular network of blood vessels becomes trapped

within the bony network, so that there is no sharp boundary between the cambial layer and the soft, vascular connective tissue (= primary marrow) between the bone trabeculae. Meanwhile, remodelling alters the interior of the bone to produce a mature bone structure, with a compact cortex and a hollow, marrow-filled medulla (see Figure 1.2.1/11). Osteoblasts within some trabecular spaces are replaced by osteoclasts, which resorb the bone matrix, whilst in other areas osteoblasts continue to thicken the trabeculae until very little space is left between them. In this way, bone development results in the correct balance of density within each bone. In addition to the formation of many of the skull and facial bones, intramembraneous ossification is the mechanism of growth and remodelling of most bones in post-natal life.

Endochondrial ossification

Most bones in the embryonic skeleton are laid down first as cartilage models, of similar shape to the adult bones. The bones then grow as a result of cartilagenous growth and endochondrial ossification, by which the cartilage is converted to bone. This process is most clearly seen in the long limb bones. The cartilage model is surrounded by a perichondrium, which is structurally and functionally analogous to the periosteum. The model grows partly by internal cell multiplication and matrix production, and partly by multiplication and conversion of perichondrium cells into chondroblasts, which are responsible for formation of the cartilage matrix. At some stage during foetal life the cartilage cells in the centre of the model cease matrix production and break down. The matrix becomes calcified, and the perichondrium surrounding the area becomes a periosteum. Osteoblasts deposit a layer of bone around the calcified cartilage, and the area is invaded by progenitor cells and blood vessels. This constitutes a primary area of ossification. The

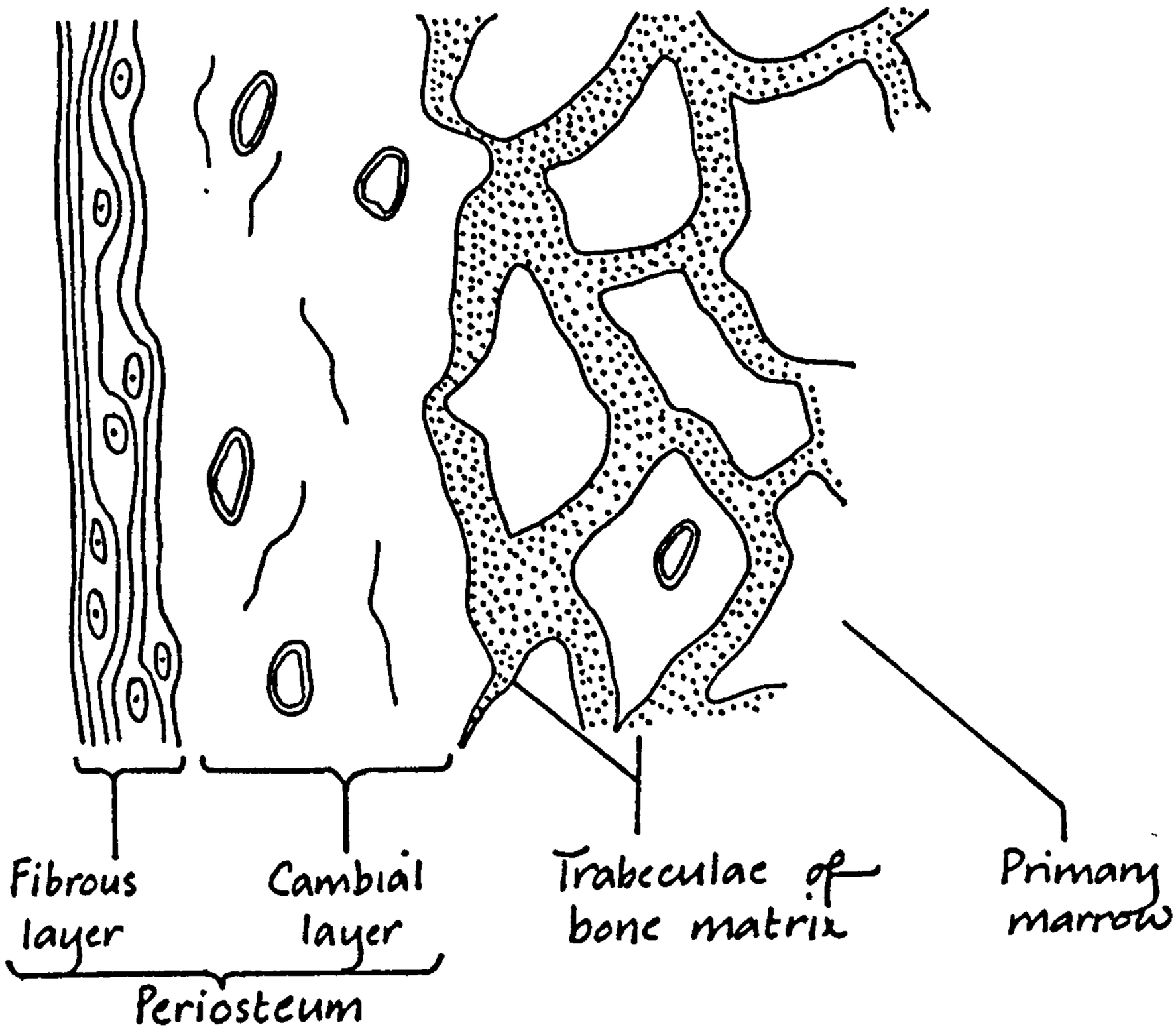


FIGURE 1.2.1/11: Schematic diagram of intramembraneous ossification

progenitor cells give rise to osteoblasts and osteoclasts, and ossification gradually progresses towards the two ends of the cartilage model.

Cartilage growth, meanwhile, continues at the ends of the bone, and as ossification progresses, an equilibrium develops between osseous invasion of cartilage and cartilage growth. There are bands of intense cartilage cell multiplication and matrix production at the ossification "fronts". These areas are known as growth cartilages, or epiphyseal plates. Beyond them, the cartilage grows radially, to form expanded cartilagenous pads, known as epiphyses. Epiphyses first appear in the skeleton shortly before birth, and Tanner (1978:32) states that new epiphyses may appear right up until puberty. The shafts of the long bones also grow radially, and this is achieved by intermembraneous ossification. Remodelling eventually causes resorption of much of the cartilaginous bone, together with some of the membraneous bone, and a dense bone cortex develops. As growth nears completion, centres of ossification appear within the cartilaginous epiphyses, and ossification occurs between the epiphyses and the articular cartilages (i.e. the cartilage pads which cushion bone joints). Eventually the epiphyses are eliminated, and growth ceases. This "closure" of epiphyses seems to be under the control of sex hormone activity in humans, and this feature is used as a gauge of "bone age" (see discussion of control of growth later in this section). By the time closure occurs, the cartilage cells have already ceased to proliferate to any great extent. A schematic diagram of long bone growth is shown in Figure 1.2.1/12..

Lymphoid tissue

Lymphoid tissue shows an unusual growth pattern, as mentioned earlier (see Figure 1.2.1/6), reaching a

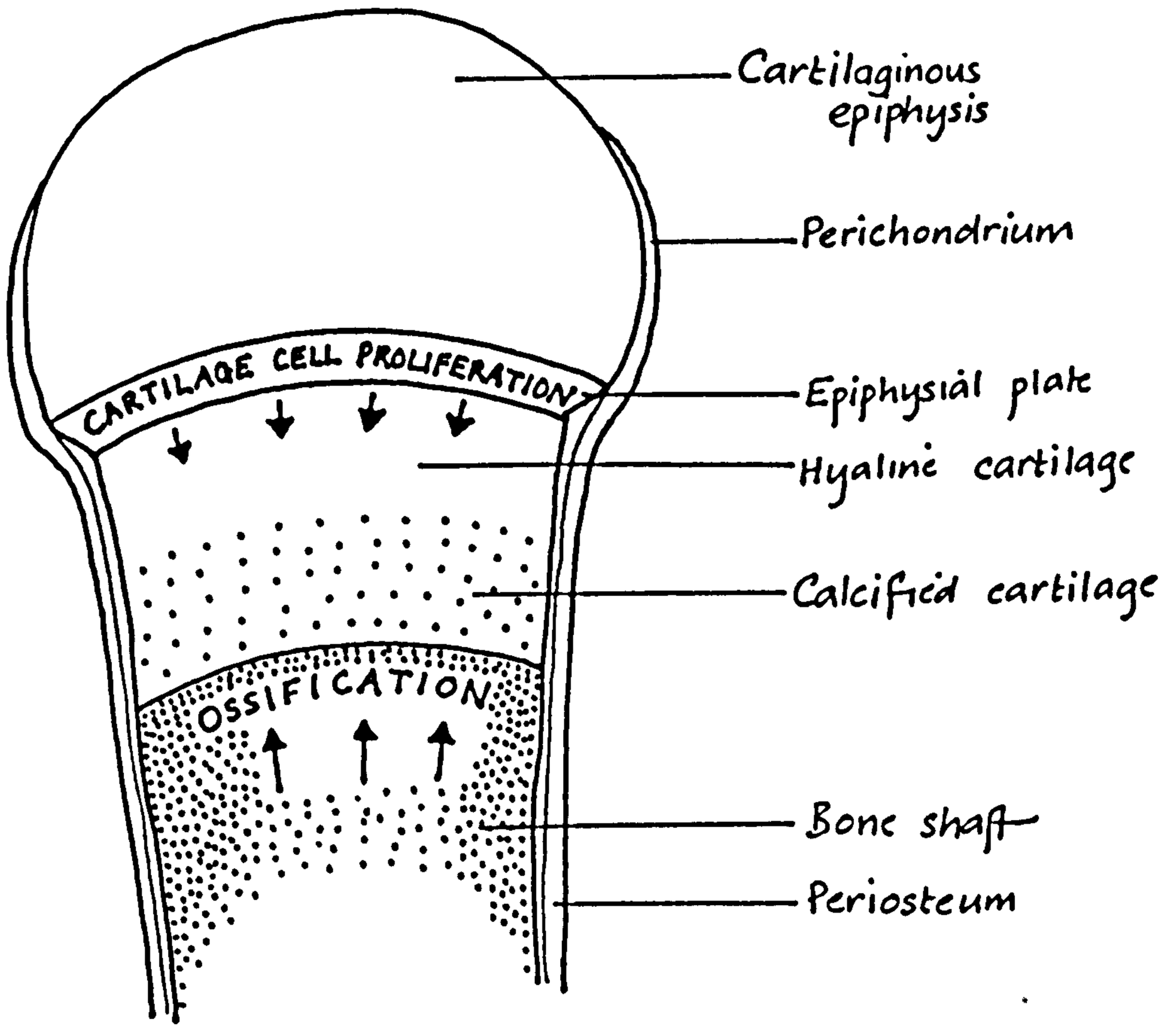


FIGURE 1.2.1/12 Schematic diagram of long bone growth (adapted from Tanner 1978:33 and Freeman and Bracegirdle 1966:27)

maximum before puberty, and thereafter declining in mass. Diffuse lymphoid tissue develops as an infiltration of the connective tissue of mucous membranes. Isolated lymph nodules seem to develop as a specific response to infection, and they are absent in new born infants or animals raised in sterile conditions. The size of the more organized lymph nodes increases greatly after birth, although their number may not increase more tham threefold (Sinclair 1978:88). Tonsils reach a maximum size at about 6 years, and then normally regress, becoming insignificant in adults.

Teeth

Tooth development begins long before any part of the tooth is visible, and the early stages of deciduous (milk) tooth development begin during the fifth week of foetal life. Growth of deciduous and permanent teeth proceeds in essentially the same manner. Primitive epithelium of the mouth grows down into the underlying tissue, and connective tissue begins to condense underneath this downgrowth. The epithelial downgrowth, now known as the enamel organ, becomes separated from the surface epithelium, and sits like a cap upon the differentiating connective tissue, which constitutes the dental papilla. The whole structure becomes encapsulated in a layer of connective tissue, the dental sac. Cell differentiation within the enamel organ results in the formation of ameloblasts, which will be responsible for enamel production. The peripheral cells of the dental papilla form a thin layer of odontoblasts, which will be responsible for dentin formation. By end of the fifth month of gestation the hard tissues of the tooth begin to be laid down, and by the time of birth the crowns of the first deciduous teeth are complete. Root development is acheived by the downgrowth of epithelial cells from the enamel organ, to form the epithelial root sheath.

Odontoblasts form adjacent to this, and produce dentine, and cementum develops from the enclosing membrane. Root development is only completed at the time of tooth eruption. Figure 1.2.1/13 is a schematic representation of incisor development.

Skeletal Muscle

Of all the soft tissues, the growth of muscle is perhaps the most important determinant of body shape and size. At birth, skeletal muscle forms 25% of total body mass, and this proportion increases to 40% or more (Brasel and Gruen 1986:60, Malina 1986:89). The rate of muscle growth is much the same in males and females up to the onset of puberty, after which there is a relatively larger rate of growth in males. Between 5 and 18 years the mass of skeletal muscle increases at least five-fold in males and four-fold in females. In females, the muscle mass doubles between the ages of 9 and 15 years, whilst in males the muscle mass doubles between 11 and 17 years (Brasel and Gruen 1986). Estimations of muscle mass are somewhat difficult in living subjects, and are based on biochemical measurements which may not be entirely accurate, but the general trends in muscle growth are fairly clear.

The precise manner by which this increase in muscle mass is brought about is somewhat controversial, and full reviews of research in this field can be found in Goldspink (1974), Brasel and Gruen (1986) and Malina (1986). In the embryo, cells known as myoblasts fuse with each other to form multinuclear myotubes, which seem to be precursors of muscle fibres (Goldspink 1974:73). Goldspink (1972) suggests that the postembryonic growth of muscle occurs in two stages. In the early stage, new muscle fibres are formed from myotubes, and the new fibres increase in girth and length. During the second

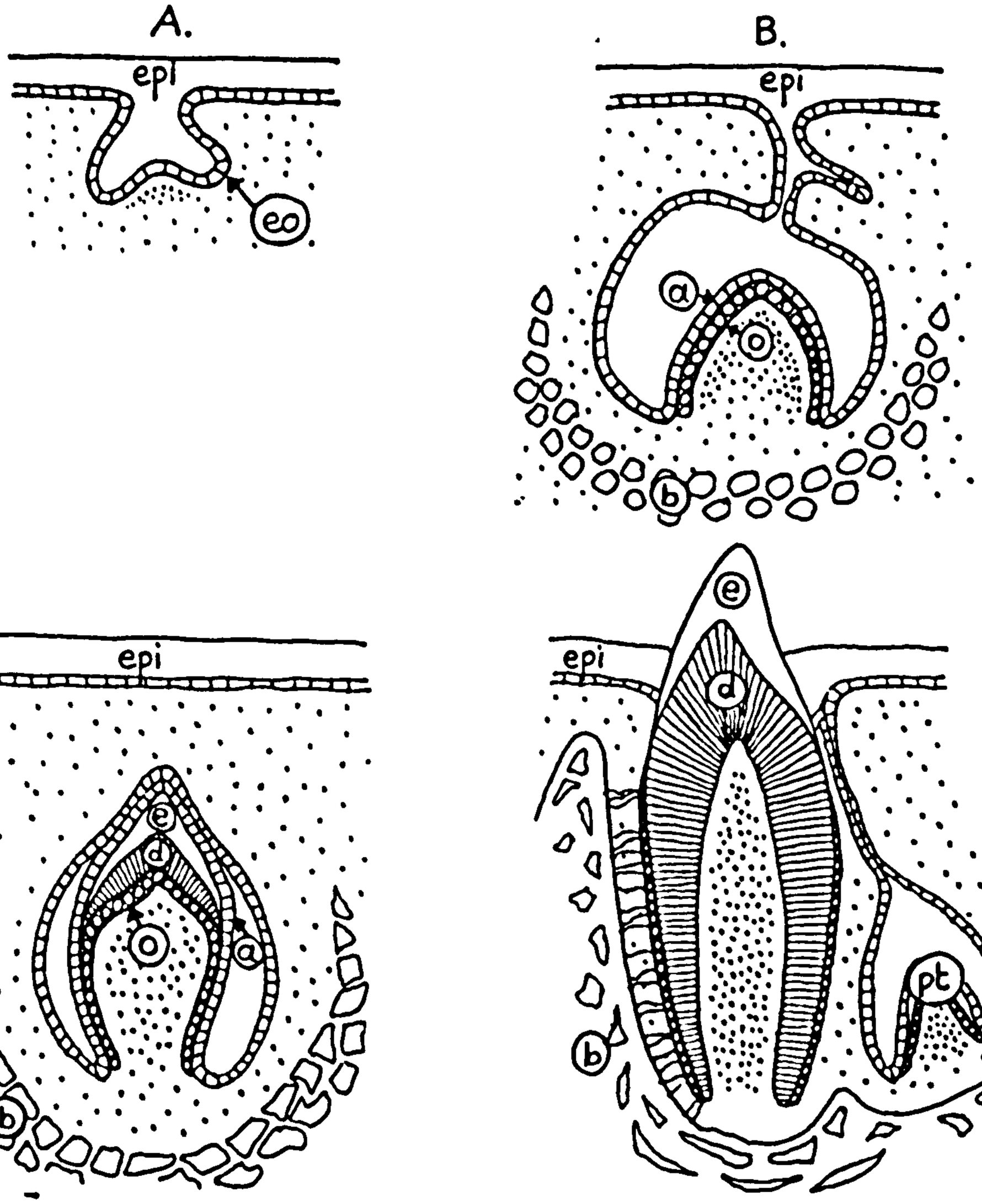


FIGURE 1.2.1/13: Schematic diagram of incisor tooth development (adapted from Clegg and Clegg 1963:256 and Bloom and Fawcett 1968:538)

growth stage, no new fibres are formed, and growth is achieved by an increase in length and thickness of existing muscle fibres. The age at which the early stage ceases, and no further muscle fibres are formed is not clear. Most authors agree that this occurs at, or soon after, birth, and Montgomery (1962) reports that the number of muscle fibres stops increasing some time between birth and four months of age. Brasel and Gruen (1986:60) report contradictory findings by Adams and De Rueck suggesting that the number of fibres may continue to increase up until the fifth decade of life, but these seem not to be generally accepted (Goldspink 1974:81).

Muscle fibres increase in length by an increase in the number of sarcomeres which are arranged sequentially along the myofibrils. The primary sites for longitudinal growth are at the junctions between muscle and tendon. Length increase also involves an increase in the number of nuclei contained in each muscle fibre. The new nuclei are thought to be derived from satellite cells, or residual myoblasts, which may be found alongside muscle fibres, and are most common in young muscle (Goldspink 1974:80, Malina 1980:77).

Muscle fibre girth also increases after birth, and the average diameter of fibres seems to increase more or less linearly with age (Malina 1980:82), although studies of mouse tissue suggest that individual fibres may show a discontinuous pattern of growth (Goldspink 1974:83). At birth, all fibres are approximately the same thickness, but in some muscles individual fibres seem to show a rapid transformation from a thin to a thick state. Older muscles therefore show a bimodal distribution of muscle thicknesses. Increase in muscle fibre girth involves an increase in the number of myofibrils within each fibre, which become more densely packed together. This can be related to an increase in the water content of muscle

tissue which occurs during growth and maturation (Malina 1980:77).

In most skeletal muscle it is possible to differentiate two types of muscle fibre, which differ in appearance and speed of response. "Slow twitch" fibres appear more irregular in cross section than "fast twitch" fibres, and the undifferentiated type of fibre which is found in the foetal and early post-natal periods may develop into either type. The ratio of slow twitch to fast twitch fibres which forms in any given muscle seems to be governed by the relative needs for strength and speed of muscle contraction.

Nerve tissue

Nerve tissue growth need not be considered here, beyond a reminder that most nerve tissue development is completed very early in life. The neurons (nerve cells proper) are thought to reach their maximum number by the fifth or sixth month of foetal life, and further growth and development of the nervous system depends on an increase in size of these cells, an increase in complexity of their connections, and the growth in size and number of the supporting cells of the nerve tissue.

D. Control and coordination of growth

It is not feasible to attempt a full discussion of the mechanisms by which the timing, amount and pattern of growth displayed by an individual are controlled. The aim of this section is simply to outline some of the factors which are known to have some influence on growth, as illustration of the complexity of the growth process and the many points at which it may be disturbed.

Factors which have been shown to influence growth fall into two classes: those which are endogenous to the individual, which generally means they are under genetic control, and those which can be loosely classified as environmental. Useful summaries of the genetic and environmental factors which may influence growth can be found in Sinclair (1978) Tanner (1978), and Rona (1981). The relative contributions of endogenous and environmental factors is much disputed, and as with any nature/nurture debate, the results of studies in this area will depend on which factors are held constant. If individuals with similar or identical genetic make-up are compared, then it may be shown that environmental factors are responsible for dramatic differences in overall growth. If, on the other hand, environmental factors are held constant, then the enormous contribution of genetic factors may be clearly demonstrated. Normally it is impossible to fully extricate the effects of endogenous and environmental influences, and both obviously play major roles in determining the final shape and size of an individual. Genetic factors will determine the maximum growth potential of each person, and environmental factors will determine the extent to which that potential is fulfilled.

Endogenous factors

Genetic factors

Whilst studies of genetically identical twins make it clear that the genetic make up of a person plays a major role in determining his or her overall size, shape and rate of growth and maturation, investigation of which genes are responsible is hampered by the fact that the growth process involves so many stages at which genetic control of cells may affect growth. Very many genes play a part in the process, by controlling such things as the

rate of cell division of a given cell type, the rate of intercellular matrix synthesis, the rate of production of some hormone, or the sensitivity of a cell to that hormone. Some single genes have been isolated by virtue of the fact that abnormalities in these genes cause drastic disturbances in growth, but far more remain unidentified. An example of a well mapped single gene which is crucial for normal growth is the gene which causes achondroplasia, where the long bones of the legs and arms fail to grow in proportion with the rest of the body.

One growth phenomenon which has a clear genetic basis is the differentiation between males and females. The timing of onset of the pubertal growth spurt is probably genetically determined (Sinclair 1978:142), and the earlier skeletal maturation of girls may be due to some difference in the genes carried by the X and Y chromosomes.

Hormonal factors

Hormonal factors are ultimately under genetic control unless there is medical intervention of some sort, but since their involvement in growth has been closely studied, they deserve some specific comments. A useful summary of hormonal control of growth can be found in Tanner (1978: Chapter 7).

Hormonal factors probably start influencing growth sometime between the second and fourth months of foetal life, by which stage at least the pituitary and thyroid glands are active. It is likely that all the hormones produced by the body play a part in growth control at some stage in the developmental cycle.

The most crucial group of hormones for growth control is produced by the pituitary gland, and includes growth hormone, thyroid stimulating hormone, and various hormones which control activity of the reproductive organs. The thyroid gland, the adrenal gland, the testis, the ovary and the pancreas all also produce hormones which are necessary for normal growth control. As our understanding of the role these hormones play in growth has increased, it has become possible to treat many of the growth disorders which may result from hormone imbalance, as long as they are detected sufficiently early. Some unfortunate individuals remain untreatable, however, where the growth disorder results not from inadequate hormone production, but from an inability of the cells to respond appropriately to the hormone.

Environmental factors

Nutrition

It is clear that malnutrition is deleterious to growth, and that there may be consequences for rate and timing of growth, for adult size and shape, and for relative tissue proportions. Famine associated with war and deprivation has been shown to cause marked delays in growth of children (Tanner 1978:127,132), but short periods of malnutrition during childhood seem to have little or no effect on adult size, as growth regulatory mechanisms seem to ensure a compensatory period of catch-up growth once an adequate diet is resumed. Chronic malnutrition during childhood, however, may mean that individuals never approach their full growth potential. The growth disturbances which may follow from lack of some specific dietary components at crucial stages in development are well known. Vitamin D deficiency, for example, causes rickets, where bone growth at the epiphyseal plates is distorted as a result of faulty calcification. Vitamin D

is an important factor in normal calcification because it stimulates intestinal absorption of dietary calcium.

Socioeconomic factors

Socioeconomic status seems to be related to size and rate of growth in almost all societies (Garn and Clark 1975, Tanner 1978:146), with children of parents with higher educational or occupational status typically being taller than others. In Britain, the height difference between children of professional or managerial fathers and those of unskilled manual workers averages about 2 cm. at two years of age and 5 cm. at adolescence (Tanner 1978:146). Some of this difference may be due to more rapid growth and maturation in the wealthier groups, but there is evidence that at least some of the height difference persists into adulthood (Schreider 1964, cited by Tanner 1978:233). It is not clear whether the effects of socioeconomic status are mediated through diet, or other environmental factors. The observation that the weight of children of lower socioeconomic status is higher relative to their height has been taken as evidence that the higher proportion of carbohydrate and lower proportion of protein in the diet of these children is a major factor in growth retardation (Tanner 1978:146). A high weight to height ratio is also, however, seen in children with low growth hormone production, and the possibility that some other factors are depressing growth hormone production in the lower socioeconomic groups should not be ignored. It is known that psychological stress may interfere with growth hormone production (see below), so it does not seem too implausible that general deprivation could have similar effects.

Family size and sibling order

The number of children within a family is inversely related to height, presumably as the amount of food or attention available is rationed more thinly in larger families. First born children grow more rapidly than do their younger siblings, although adult height does not seem to be related to birth order (Tanner 1978:147).

Emotional factors

There are a number of studies which show that extreme psychological stress may cause short stature in children (Friend and Bransby 1947, Widdowson 1951, Tanner 1978:144,217-9). Growth hormone deficiency may be found in such children, but removal of psychological stress is followed by a resumption of normal growth hormone production and a period of catch-up growth. Emotional factors are also implicated in such eating disorders as bullimia nervosa and anorexia nervosa, which may have marked secondary effects on.growth and the proportion of fat within the body.

Disease

Even quite minor diseases may cause temporary disruptions in the normal growth curve, as may the administration of certain drugs, but catch-up growth after cure of the disease normally compensates for any growth delay. Chronic serious disease may have more permanent effects, but this is relatively rare. Specific growth disorders will be considered briefly at the end of this section.

Secular growth trends

There is clear evidence that a trend towards increased size and earlier maturity has been operational in many

countries over at least the last century (Tanner 1978:150-151, Rona 1981). This trend seems to have slowed or stopped in Britain and some other countries, but is still continuing elsewhere. Various factors have been proposed as explanations for this phenomenon, including climatic change, a reduction in disease, improved nutrition, and genetic factors. The observation that the trend was more obvious in the more industrialized areas of Scotland than in less industrialized areas (Grant Keddie 1956, cited in Rona 1981:270) has been used to support the hypothesis that changes in social and material conditions are the most important contributory factors. An alternative explanation for this observation is related to the phenomenon known as hybrid vigour. In many plants and animals, the offspring of individuals with very different genetic make up are often larger and more vigorous than either parent. Increased mobility of the population has no doubt increased the incidence of outbreeding (i.e. marriage to unrelated individuals) in most countries of the world, and this could well be contributing to the secular trend. In the Grant Keddie study, it is possible that increased mobility in industrial areas could have been associated with increased levels of outbreeding.

Integration and coordination of growth

The growth process is something of an organizational miracle, and the resilience of development to adverse factors is extraordinary. Waddington (1957) used the term "canalization" to describe the strong tendency for the development of a young animal to return to its original course if anything had caused a temporary diversion in the normal stream of development. It is as if the architectural plans of the adult body are laid down in the genes, but the exact timing and sequence of the building stages needed to produce the adult form are

fairly flexible. If something interferes with development for a while, later stages of growth and development can usually be modified to make up for lost time. The canalization phenomenon is evident both in overall growth curves and in local growth of tissues and organs. When overall growth is measured, it is well documented that periods of growth delay during starvation or disease are usually followed by periods of increased growth activity which continue until the growth curve returns to the level which would have been expected had there been no growth restriction. This rapid compensatory growth is known as catch-up growth, and it is only absent or incomplete if growth restriction occurs very early in life or for a prolonged period. If the rate of catch-up growth is inadequate to allow full compensation for growth delay by the normal time of maturity and cessation of growth, then maturity may be delayed to allow a longer period of growth. One interesting feature of catch-up growth is that it is more efficient in females than in males, but the reasons for this are not clear (Sinclair 1978:158).

The mechanisms by which canalization and associated phenomena such as catch-up growth are controlled are very poorly understood, although it has been suggested that the pattern of growth and development is to some extent under neural control (Tanner 1978:159). The proposal is that a growth centre in the brain, possibly in the hypothalamus, has a representation of the ideal growth curve as laid down by the genes, and somehow monitors any discrepancy between actual and ideal growth, and initiates corrective measures.

The widely varying growth patterns of different parts of the body and different tissue types must be coordinated most exactly if a properly proportioned body is to develop. Some physical characteristics can be clearly

linked to specific gene effects, but a certain amount of plasticity is necessary if these physical traits are to harmonise properly. Different parts of the face, for example, must exert some kind of mutual growth control if they are to fit together adequately. In general, the ability of parts of the body which are under different genetic control to grow in such a way as to form an integrated whole is remarkable, although major genetic imbalances may prevent normal development and integration. Down's Syndrome is an obvious example of such a major, global imbalance in growth and development, and this is discussed in detail in Section 2.3 below.

Although all cells possess identical genetic information, and therefore have the potential ability to manufacture all the proteins coded in the genes, cells tend to lose this general competence as they become differentiated. During development of a cell type, some genes are turned on, and others are switched off. In many cases it seems to be impossible to switch genes on again once a critical stage in development has been passed. There seem to be some stages during development at which a tissue or organ is especially sensitive to some controlling factor, such as a hormone, and this is presumably related to the sequence of switching genes on and off. If the necessary stimulus is not present during this sensitive period, or if some agent interferes with the normal developmental response, then later growth and development may never be able to compensate for the missed moment. An example which is relevant to phonetics and speech therapy is the failure of the two sides of the palate to fuse with one another during the second and third months of foetal development, causing cleft palate. In some cases there seems to be a genetic basis for this, whilst in others it seems likely that drugs or nutritional factors have interfered with growth at the sensitive stage of palatal closure.

E. Growth disorders

Short stature or disproportionate growth may be due to specific abnormalities, mostly of genetic origin. Since many of these have global effects on growth and development, and may thus have consequences for vocal tract size, the most common growth disorders will be summarised below. It should be stressed that many people who are considered to be unusually short or tall simply represent the edges of the normal population distribution, and are perfectly normal. Some children may cause concern because they appear to be smaller than normal, but it will often be found that this is because of growth delay, which is to some extent genetically determined. In other words, it may be found that growth delay, associated with late puberty so that growth continues longer to compensate for early small stature, is a familial trait.

Bone and cartilage disorders

There are a number of genetic disorders of bone and cartilage growth, but most are very rare. The most easily recognised form of dwarfism results from one such disorder, achondroplasia, mentioned earlier. In this disorder a single gene defect causes a marked reduction in limb bone length and characteristic facial features, although trunk development is fairly normal (Sinclair 1978:186, Tanner 1978:215).

Endocrine disorders

Because of the many hormones involved in growth control, there are many possible hormonal growth disorders. Children with growth hormone deficiency are usually of normal size at birth, but grow slowly thereafter (Tanner 1978:212). Most cases respond well to treatment with

growth hormone, if detected early enough. Thyroid deficiency is another common cause of short growth, which also responds well to treatment. A rather different type of growth disorder is seen in children with hormonal disorders which lead to precocious puberty. Early growth is fast, but as puberty occurs very early, and is associated with an early cessation of growth, final stature is rather small. Excessive growth may be as much of a problem as lack of growth, and one of the most obvious causes of gigantism is the overproduction of growth hormone. If this continues beyond puberty, the hormone produces disproportionate growth of parts of the body which are still capable of growth (i.e. where epiphyses have not closed). Excessive growth of hands, feet and jaws is especially marked.

Chromosome disorders

Many chromosome abnormalities influence growth, but two are most commonly associated with short stature. The first of these is Down's Syndrome, which will be discussed in more detail in Chapter 2.3. The second is Turner's Syndrome, which is associated with the presence of only one X chromosome instead of the normal two. These girls tend to be very short, and may have a variety of other physical abnormalities.

1.2.2 GROWTH MECHANISMS IN MAINTENANCE AND REPAIR

In the adult there are dramatic differences in the ability of tissues to regenerate and replace material which is damaged or lost by injury or disease. These variations are related in part to the degree of trauma which a tissue normally has to withstand, and partly to the degree of differentiation displayed by a cell type. Cells which are highly specialized in form and function are usually much less able to multiply themselves than are less differentiated types.

Epithelium is an example of a tissue with very great powers of regeneration. The epithelium lining the respiratory and digestive tracts and covering the external surfaces of the body is subjected to constant mechanical and chemical irritation. In many sites it has to withstand considerable friction and frequent minor injuries. It therefore needs to be continually replaced, and this form of regenerative growth continues throughout life. Most epithelium contains cells which are relatively undifferentiated, looking fairly similar to the generalized animal cell described in Section 1.1.2, and it seems that the process of growth by multiplication of cells is therefore relatively simple.

At the other end of the scale are tissues such as nerve and muscle, where the cells are highly differentiated, but which are not normally subject to regular damage. Nerve tissue is unusual in that after development of the nervous system is completed, at a relatively early age, no new cells can be formed. Broken sections of nerve cell may regenerate under some circumstances, but this is the limit of its regenerative capability. Muscle tissue also shows rather limited powers of regeneration.

Between these extremes is a range of tissue types which are not normally faced with continuous demands for repair, but which are able to respond to injury or increased tissue activity by making new cells. Most types of connective tissue and glandular epithelia fall into this class.

Inflammation and scar formation

Inflammation is a complex, coordinated response to tissue damage, which acts to limit infection and to repair injured tissue. It is common to many tissues, although it normally occurs principally in connective tissue. Since it may result in temporary or permanent increases in tissue bulk it is appropriate to consider it as a general growth process. In tissues which are capable of full regeneration, inflammation may be a fairly short phase preceding regeneration of the original tissue. In tissues such as muscle, where full regeneration is not possible, or in cases where damage is prolonged or extensive, connective tissue may develop as a permanent replacement for the original tissue, forming an area of scar tissue. More detailed accounts of inflammation can be found in, for example, Sandritter and Wartman (1969:20-27).

It is convenient to view inflammation as a two stage process.

a) The acute stage

The acute stage of inflammation can be thought of as an emergency reaction, which marshalls together the elements necessary for defence and repair. This stage exhibits certain common features, regardless of the size, site or type of injury. The three predominant signs are listed below.

i) Hyperaemia.

This simply describes an increase in blood flow to the area, which is usually acheived by capillary dilation.

ii) Leucocyte infiltration.

The capillaries become more permeable, and allow white blood cells (leucocytes) to pass into the affected tissue. Some of these cells are active in limiting infection, by engulfing foreign bodies, or by antibody production.

iii) Swelling due to fluid exudation (oedema).

Fluid also passes out of the dilated capillaries and collects in the intercellular spaces of the tissue.

b) The chronic stage

The chronic stage of inflammation follows a much more variable course, depending on the extent, duration and type of damage. Necrotic (dead) tissue and blood clots are resorbed by specialized cells, and the damaged area may be localized and walled off by the deposition of collagen fibres (fibrosis). Active repair of damaged tissue is brought about by the proliferation of new connective tissue and blood vessels. This proliferative repair tissue is generally known as granulation tissue, but its exact morphology may vary considerably. In some cases fibrosis predominates, with a progressive increase in collagen density, and eventually hyaline may also be deposited in the fibrosed tissue. Hyaline is the firm, glassy substance which forms the matrix of some cartilages (see Section 1.1.2), so that this type of granulation tissue will form areas of greatly increased stiffness. Other cases may show no sign of fibrosis, but have marked capillary growth with much lower consequent stiffness.

Granulation tissue may be a precursor of full regeneration, or it may develop into a permanent scar. Scar tissue is usually very rich in collagen fibres, and may appear whitish because of a limited blood supply.

Regenerative powers of tissues

The following descriptions of regenerative patterns in different tissue types are based on comments in several texts on histology and growth (e.g. Leeson and Leeson 1976, Sinclair 1978, Junqueira and Carneiro 1980), and is intended to be no more than a brief summary. In general, the ability of all tissues to repair themselves seems to decrease with age, and is dependent on a reasonable level of overall health, a good blood supply, and adequate levels of vitamins and minerals (Sinclair 1978:175).

Epithelium

The unusual powers of regeneration displayed by covering epithelia have already been mentioned, and in some parts of the digestive tract, for example, the constant injurious effects of mechanical and chemical irritation lead to the normal replacement of tissue as rapidly as once every two days.

Epithelium is particularily prone to metaplasia (Junqueira and Carneiro 1980:74). This is the process whereby one type of epithelium may respond to certain physiological or pathological stimuli by transformation into another type of epithelium. For example, chronic irritation of the larynx by smoke or chemicals may lead to the transformation of ciliated columnar epithelium into stratified squamous epithelium (see Section 2.5). This is often an adaptive response to environmental conditions, replacing a less resilient type of epithelium with one better able to cope with the unusual stimuli.

Connective tissue

Connective tissue proper

Connective tissue proper shows considerable powers of regeneration, and in addition to ready repair of damage within connective tissue itself, formation of new connective tissue is involved in scar formation within tissues which are less able to regenerate.

Cartilage

As with other tissues, the regenerative ability of cartilage is greatest in early childhood. Later in life, it often regenerates incompletely, so that in areas of extensive damage scars of dense connective tissue may replace lost cartilage. Such regeneration as does occur arises from activity of the perichondrium, from where cells migrate into the damaged area, forming new cartilage.

Bone

Throughout development, the power of bone to remodel itself in response to various stimuli is surprising for such a rigid material, and its regenerative response to injury is also remarkable. The rigidity and strength of bone means that if it is injured, the effect is likely to be quite dramatic, with complete fracture of the structure.

One of the most marked complications of a fracture results from tearing of the blood vessels within the bone. This leads to blood clot formation, and death of bone cells around the fracture. Bone matrix is also destroyed. An early stage of repair therefore has to be the removal of blood clot, and damaged tissue, by

osteoclasts and other cells. There is proliferation of fibroblasts at the periosteum and endosteum, and these new cells migrate into the damaged area, forming a cellular tissue. Small areas of cartilage then form within the new connective tissue, so that new bone growth can proceed both by endochondral ossification of these patches and intramembranous ossification. The fracture is thus temporarily repaired by the development of irregular trabeculae of immature bone, forming a bone callus. Remodelling of this bone callus occurs in response to the stresses imposed upon it, just as in normal bone development. In this way, normal bone structure is eventually regained. The primary bone of the bone callus is gradually resorbed and replaced by bone which is able to resist the forces it is subjected to. If the fragments of bone do not align in their original form, there may be unusually high stresses imposed at the fracture point, and the fully healed bone may actually be stronger in this area than before. Complete repair of bone may take months or even years in an adult, although it is usually much faster in children.

A schematic diagram of bone fracture repair is shown in Figure 1.2.2/1, adapted from diagrams shown in Sinclair (1978:176) and Junqueira and Carneiro (1980:144).

Lymphoid tissue

The regenerative power of loose lymphoid tissue and lymph nodules seems to be considerable, as they may develop throughout life as part of an immune response. The regenerative power of more organised lymphoid organs such as the tonsils seems to be limited, as illustrated by the fact that surgical removal of infected tonsils is seldom followed by significant regrowth.

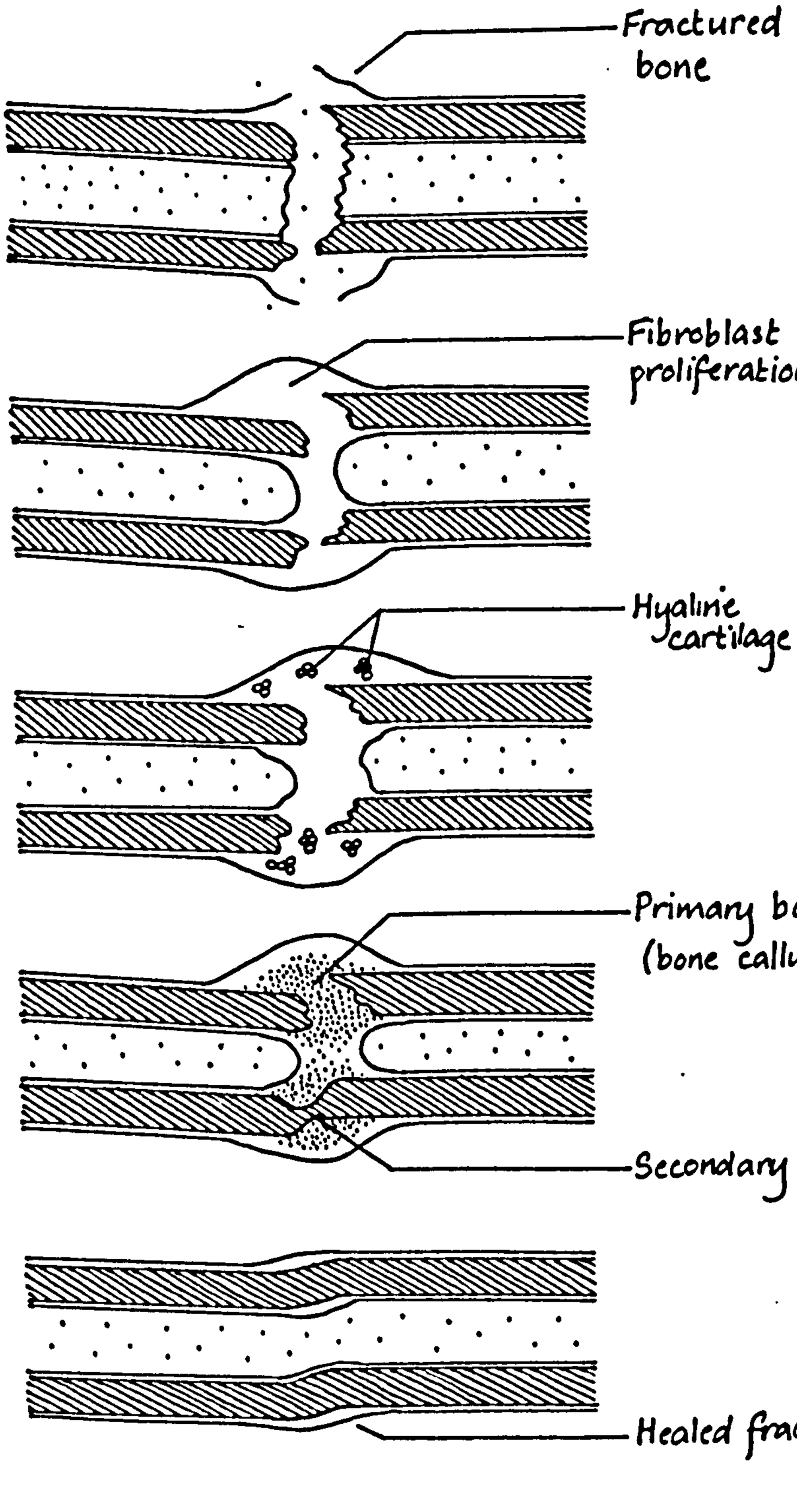


FIGURE 1.2.2/1: Schematic diagram of bone fracture repair (adapted from Sinclair 1978:135 and Junqueira and Carneiro 1980:144)

Teeth

If a whole tooth is lost through injury, or through disease of the tooth itself or the tissues within which it is embedded, no regeneration is possible. Repair of damage to small parts of the tooth depends on which tissue is affected, and on whether the nerve and blood supplies remain intact.

Skeletal muscle

Skeletal muscle is able to respond to prolonged periods of increased activity by increasing its mass, but this is done by the enlargement of existing muscle fibres (cells) rather than by an increase in cell number. If small areas of muscle are damaged as a result of injury there may be some regeneration of muscle fibres. Undamaged fibres grow out towards the injured area, injured fibres are digested by macrophages or other cells, and new fibres may develop within the framework of old fibres. Regeneration is most likely if the nuclei and some surrounding cell constituents remain alive. These can form separate cells and may then multiply and fuse to form the new muscle fibres. The importance of nerve activity in maintenance and growth of muscle is shown by the fact that proper regeneration seems to be possible only if the nerve supply remains intact. In larger injuries, damaged muscle is replaced by connective tissue scars, with a consequent impairment of function.

Nerve tissue

Nerve tissue is not, for the purposes of this thesis, being considered as a major structural component of the vocal apparatus, and its regenerative powers will not therefore be considered in any detail. It is worthy of some comment, however, because it represents the extreme

example of a highly specialized tissue in which some cells lose the ability to divide and regenerate at a very early stage in development. After birth, the principle cells within the central nervous system are unable to divide, and any injury is repaired by non-functional connective tissue. The nerve cells of the peripheral nervous system are capable of limited regeneration only if the nuclei of the cells are not damaged.

1.2.3 DEGENERATIVE CHANGE AND NEOPLASTIC GROWTH

If growth and development during childhood are seen as processes working towards an ideal mature organic state, and growth processes involved in maintenance and repair are seen as working to maintain this state, then the sorts of degenerative change which accompany old age, and neoplastic growth can both be seen as processes which tend to cause deterioration and disturbance of the mature organic state. Both will therefore be considered together in this section as adverse types of change.

General trends in aging and degeneration

The bodily changes which accompany aging are not well understood, and it is difficult to separate the changes which are an inevitable consequence of the passage of time from those which are the consequence of chronic disease. This is because one of the characteristics of old age is a progressive decrease in efficiency of the immune system, which leads to an increase in disease in the elderly. The universality of some organic changes in old age does, however, suggest that they are general features of aging tissues. In the same way that different tissue types show different growth patterns, and different patterns of repair, so they show different patterns of degeneration with age. The term degeneration will be used here to describe any organic change which has adverse effects for the function of a tissue or organ. Such changes may involve the loss of tissue mass, or deleterious alterations in tissue consistency. The susceptibility of a given tissue to degeneration with age is probably closely linked to its ability to regenerate and repair itself following injury. Comments on the types of degenerative change commonly seen in different tissue types may be found in such texts as Bourne (1961),

Comfort (1965), Leeson and Leeson (1976), Sinclair (1978), and Junqueira and Carneiro (1980).

As in growth, degenerative changes may be due either to alterations in the number or type of cells within a tissue, or they may be due to changes in the intercellular material. In early adulthood, cell division in most tissues balances the loss of cells through wear, tear and aging, but the rate of cell division gradually decreases in later years so that there is a progressive decline in cell number within most organs of the body.

Connective tissue

Connective tissue proper

Within the matrix of connective tissue, a progressive reduction in water content, and an increase in fibre content is seen throughout life. In old age, collagen fibres may increase in number, but they also change their properties somewhat, forming more cross linkages, and showing increasing signs of damage. Elastic fibres tend to become thicker, and then to split, and they lose their elasticity. Calcium salts may be laid down around collagen fibres, causing major loss of flexibility, and this is especially obvious in cartilage and in the connective tissue in blood vessel walls and in the dermis of the skin. Fatty degeneration of connective tissue is also common, as cells are lost, and fatty deposits laid down within the matrix.

Degenerative changes in cartilage and bone are sufficiently important in terms of their consequences for vocal apparatus configuration to merit some expansion below.

Cartilage

Hyaline cartilage loses its translucency in old age, as the matrix changes its composition, and the cell content decreases. Coarse fibres may be deposited, in a process known as asbestos transformation (Leeson and Leeson 1976), which leads to softening or loss of matrix. The most obvious degenerative change in cartilage is calcification. Calcium compounds are laid down within the matrix, as in bone formation, so that diffusion of nutrients is limited and cells die. There may then be gradual resorption of the calcified area so that overall tissue mass is reduced..

Bone

One of the main degenerative changes observed in bone is the loss of calcium. This tendency, known as osteoporesis, is most marked in women, and is thought to be exacerbated by hormonal changes following menopause. A further causal factor may be the calcium-deficient diet of many old people, together with less ability to absorb what calcium is consumed.

The reduction in the volume of bone tissue per unit volume of bone structure between youth and old age may be as much as 15%, and this is most marked in spongy bone. Bones become progressively more porous and brittle, as the number of trabeculae in spongy bone decreases, and the thickness of dense bone areas is eroded. The Haversian canals may become larger, and fill with fibrous or fatty tissue, as bony matrix is lost.

Skeletal muscle

The mass of muscle in the body is estimated to fall by about a third between the ages of 30 and 90 (Sinclair

1978:215), but it is not clear how far this is due to loss of muscle fibres, and how far it is due to changes in individual muscle fibres. The collagen and elastin content of muscle seems to increase, but it seems that the major cause of impairment of muscle function in old age is probably related to degeneration of nerve tissue.

Nerve tissue

The earliest degenerative changes are apparent in the nervous system, which accords with the lack of regenerative ability in nerve tissue. Accurate tests of the special senses and such measures as reaction times show the onset of functional deterioration soon after completion of the pubertal growth spurt (Sinclair 1978:211), and a steady deterioration continues throughout life. In the elderly, the results of nervous degeneration are widespread and obvious, including loss of muscular control and hence impairment of posture, loss of learning ability and memory, and poor physiological regulation of such factors as temperature and blood pressure

Neoplastic growth

Growth processes within the body are normally accurately coordinated and controlled (see Section 1.2.1), appearing to conform to some overall programme of development. Some forms of what may appear to be non-programmed growth may occur as specific responses to trauma or disease, as exemplified by the formation of granulation or scar tissue (see previous Section, 1.2.2). These are, however, appropriate growth responses to specific abnormal events, and as such they can be seen as obeying the rules of an overall "maintenance programme". Sometimes a tissue, or group of tissues, may begin to grow in a totally

inappropriate and non-programmed manner. In some way the normal mechanisms for organizing and restricting tissue growth seem to be defeated, and the result is the formation of tumours. Neoplastic growth may be defined as any inappropriate, non-programmed growth which may lead to the development of tumours (neoplasms). Neoplastic growth becomes more common with increasing age, possibly as a result of decreasing efficiency of the immune system. The speed of growth of tumours seems, however, to be linked to some extent with the overall growth potential of the body, and cancers tend to be slower growing in the elderly.

Tumours may develop at almost any site in the body, although they occur much more frequently at some sites, and within some tissue types, than in others. They vary widely in their morphology, and in their level of malignancy. Neoplastic growth may involve various basic growth patterns, which are shown schematically in Figure 1.2.3/1.

If the neoplasm originates at the surface of the body, or adjoining one of the internal cavities, then growth may proceed by the protrusion of a tissue mass from the surface. Such protrusions may be stalked (pedunculated) or broad based (sessile).

Growth may also involve displacement, but not invasion of other tissues. In this pattern of growth, adjacent tissues may be distorted and mechanical compression may eventually lead to necrosis (local tissue death) and loss of normal tissue, but there is no intermingling of tumour cells and neighbouring normal tissue cells.

Growth by invasion and infiltration of other tissues implies a breakdown of the boundary between the tumour and adjacent tissues. Tumour cells may actively migrate

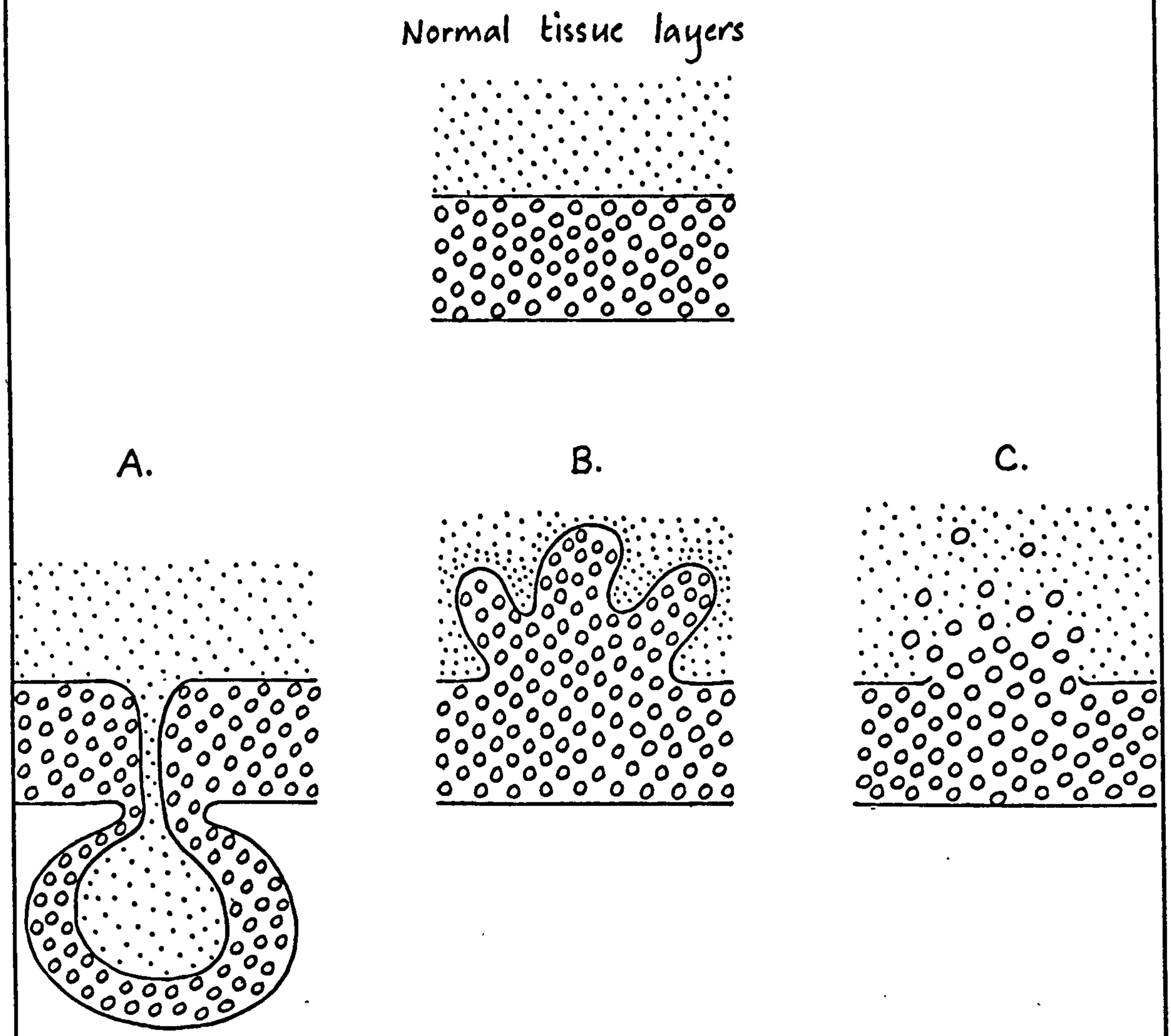


FIGURE 1.2.3/1: Schematic diagram of neoplastic growth patterns
A. extrusion through another tissue layer, with protrusion.
B. displacement of adjacent tissues
C. invasion within adjacent tissue

into other tissues, so that tumour cells become intermingled with normal tissue.

Tumours which are described as benign typically show one or both of the first two growth patterns. Although a large abnormal mass of tissue may develop, benign tumour cells do not actively invade neighbouring tissues. Neither do they metastasise, forming secondary tumours elsewhere in the body (see below). This is not to say that benign tumours are necessarily without risk for the patient. The sheer bulk of unwanted tissue may cause serious problems, either by compression of other tissues or by obstructing internal cavities. Laryngeal papilloma, for example (see Section 2.5), is described as a benign tumour, but it may nonetheless become life threatening if it grows sufficiently large to block the airway.

It is misleading to suggest that there is a clear distinction between benign and malignant tumours. Rather, there is a continuum from thoroughly benign, non-invasive tumours to highly invasive malignant tumours. Some forms of benign tumour may also, under some circumstances, develop into malignant forms. The distinction between neoplastic growth and appropriate growth responses to trauma may also be somewhat unclear (Wahi et al. 1971:19). Vocal polyps, for example, are thought by some authors to be the result of a chronic inflammatory response to chemical or mechanical irritation of the vocal fold, and by others to be examples of laryngeal neoplasm (see Section 2.5). Certainly the histology of well defined benign tumours is often virtually indistinguishable from that of granulation tissue formed around the site of an injury (Sandritter and Wartman 1969).

The degree of malignancy displayed by a tumour can be defined in terms of its ability to invade and infiltrate

other tissues, and to metastasise. Metastasis is the formation of secondary tumours, resulting from the dissemination of primary tumour cells to other parts of the body, where they settle and multiply. Dissemination may be due to active migration of tumour cells, or to passive spread, when cancerous cells enter the circulatory system and are carried around the body by blood or lymph.

A useful summary of the biology of malignancy can be found in Currie and Currie (1982).

Cell features in malignancy

There is no simple diagnostic sign of cell malignancy. Currie and Currie (1982:79) comment that "Structurally, the most remarkable feature of malignant cells is their unremarkability." The only reliable way of identifying a malignant cell is to show that it is capable of giving rise to a malignant tumour when injected into a suitable host.

Despite this lack of consistent malignant features, there are various cell characteristics which may be taken as useful indicators of malignancy. In terms of gross morphology, abnormal observations may include an increased incidence of cells undergoing mitosis, and the presence of abnormal mitoses. The nucleus of malignant cells may be large relative to the volume of cytoplasm, and may stain more readily than normal. Various other structural abnormalities may sometimes be visible. The chromosome content of malignant cells is often unusual, with considerable variation in chromosome number and structure.

Cell changes which are thought to be typical of malignancy can be induced by a variety of agents,

including ionizing radiation, chemical carcinogens and oncogenic viruses, and tissue culture offers a convenient means of examining the behaviour of such transformed cells. The most obvious behavioural changes exhibited by these cells concern patterns of cell proliferation and growth control. For example, most normal cells will only grow in culture if they are allowed to settle on a solid surface. Malignant cells, in contrast, will grow readily even when prevented from anchoring themselves to a surface. They are described as displaying anchorage-independent growth.

Another characteristic of malignant cells is the loss of density-dependent growth inhibition. Normal cells grown in culture continue to multiply until they form dense sheets of cells, one cell thick, but intercellular contact then seems to act as a signal, preventing further cell division. Malignant cells, however, continue dividing far beyond this point, forming multilayered masses of tissue. There is some controversy over the factors involved in normal density-dependent inhibition of cell division, but it seems clear that malignant cells no longer respond to such controls.

The ability of malignant cells to move around also seems to be increased relative to normal, and they lack another of the behavioural features of normal cells which is presum ably involved on normal tissue organization. Normal cells exhibit contact inhibition. In other words, if, as they move about, they come into contact with other cells, they stop moving. Malignant cells do not share such inhibitions, and seem much more prone to apparently aimless wandering.

Malignant tumour progression

Tumour progression is a term coined by Foulds (1954, cited by Currie and Currie 1982:60) to describe the way in which malignant tumours evolve. The evolution of a fully malignant tumour may involve three stages; initiation, latency and promotion.

Initiation of a tumour usually seems to involve the multiplication of a single abnormal cell, to form a nest of potentially cancerous cells. Initiation may be due to exposure to carcinogens of various sorts, which in some way alter the genetic material of susceptible cells. There may then be a period of latency, which can continue for a considerable length of time. The next step in tumour progression is promotion. Once a group of cells has been initiated, a variety of chemical and physical stimuli seem to act as triggers, promoting the development of a malignant tumour.

Whilst this idea of tumour progression represents only one possible model to explain clinical findings, there are many histological abnormalities which are consistent with a view that initiated cell populations may remain in a latent phase for many years, and that only under some conditions does promotion to active cancerous disease occur. Examples include various lesions which are commonly described as "pre-malignant" or "pre-cancerous", because a certain proportion of such disorders eventually develop into active cancers (Wahi et al. 1971:19). Keratosis and hyperplasia of the squamous epithelium of the vocal fold (see Section 2.5) fall into this group. Carcinoma-in-situ (also described in Section 2.5) may be another example of initiated, but latent, malignancy. The histology of carcinoma-in-situ is highly suggestive of malignant change, but it remains delimited by the

basement membrane of the epithelium, and may not become actively invasive for many years, if at all.

Histologically, malignant tumours are characterized by a lack of normal tissue organization. Some may present as an apparently totally haphazard arrangement of cells. Others display some organizational features of their tissue of origin, but arranged in an abnormal way. Cancerous tissue varies, too, in the extent to which it retains the functions of its parent tissue. In some cancers, cells lose almost all differentiation of both form and function. Many cancers of the endocrine glands, however, continue to produce hormones, but one feature of malignany in such semi-differentiated tumour tissue is that it no longer responds to the normal mechanisms for controlling hormone production.

Invasiveness

Clinically, the ability of cancers to invade and destroy normal tissues is their most disturbing characteristic. Both benign and malignant tumours tend to expand first along the lines of least mechanical resistance. For example, epithelial tumours will tend to spread laterally, rather than breaching the basement membrane. Connective tissue tumours will expand through loose areolar tissue layers. Once a mass of tissue has formed, oedema and compression-induced necrosis of normal tissue may allow new pathways for easy expansion.

Active invasion along less open pathways seems to be associated with increased motility and the loss of some of the normal cell control features mentioned above, i.e. contact inhibition and density-dependent growth inhibition. Another factor in increased invasiveness may be associated with the observation that malignant cells often show less intercellular adhesiveness. Normal

epithelial cells, for example, are very firmly attached to one another, but malignant cells derived from epithelium detach very easily from their neighbours, and may thus be more readily able to infiltrate other tissues. The loss of normal intercellular behaviour may be related to various changes in the cell surface structure which can be detected in malignant cells.

Many other abnormal properties have been detected in malignant cells, but since one of the problems faced by researchers in the field of malignancy is the huge range of morphology, biochemistry and behaviour of cancers, it is not possible to discuss the implications of these. The above comments must stand as a very brief summary of some of the principle features of malignancy which may illustrate the problems which may ensue when the body's highly complex and incompletely understood growth control mechanisms are disturbed. Some examples of laryngeal neoplasms and their consequences for phonation will be discussed in Section 2.5.

1.2.4 GROWTH AND CHANGE OF THE VOCAL APPARATUS

Introduction

The aim of this section is to present a summary of the normal patterns of growth and change of the vocal apparatus which occur between birth and old age, and to show how the growth processes outlined in the previous sections are coordinated to produce systematic changes in the configuration of the vocal organs.

Age related changes in the vocal apparatus can be seen as falling into three main phases. During the first phase, which corresponds to the period between birth and puberty, there are major changes in the vocal apparatus which accompany general patterns of growth and development. There are no major differences between the sexes in terms of organic factors during this phase. The second phase, from puberty to maturity, is characterised by the fact that male and female patterns of growth and development are rather different, and it is during this phase that the major differentiation between the male and female vocal apparatus emerges. During the final phase, from maturity to senescence, growth processes are active in maintenance and repair only, and the major changes which occur are the result of degenerative change.

Major organic changes within the vocal apparatus

A. Skeletal development

It is easy to think of the skeleton as a constant structure, underlying the more flexible and changeable soft tissues. In the long term view of development, the plasticity of the skeleton makes it a far from constant structure, but it is reasonable to say that at any given point in the life cycle, the skeleton does behave as a

rigid framework, around which the soft tissues are arranged. Soft tissues are subject to constant observable distortion during normal movement of the body, whereas bones are not. Infection, hormonal or physiological state may have immediate and significant effects on the size and consistency of some soft tissues, but not on the skeleton. When an individual is studied over a short time period, we can therefore state that the overall shape and size of that person's vocal apparatus will be limited primarily by his or her skeletal structure. For this reason, the growth patterns of the skeletal structures which underlie the vocal apparatus will be considered first. The most important of these structures is probably the skull, together with the cartilages of the facial skeleton.

The skull is usually described as consisting of two parts; the cranium, which encloses and protects the brain, and the facial skeleton. Structurally, these two parts form a cohesive whole, but functionally they are rather different, and this difference is reflected in their growth patterns. The cranial and facial proportions of the skull grow disproportionately throughout childhood. At birth, the cranium is 8 to 9 times the size of the face, and during the first 6 to 12 months of life the cranium grows more rapidly than the rest of the skull, thus increasing the relative size of the cranial portion. Thereafter, facial growth is greater, and continues longer, so that in an adult the cranium is only 2-3 times the size of the face (Watson and Lowrey 1967). Figure 1.2.4/1, which compares front views of newborn and adult skulls, illustrates this change. Growth of the base of the skull, which provides points of articulation with the vertebral column and allows passage of the respiratory and digestive tracts and the spinal cord, is allied with the facial skeleton in terms of its growth behaviour.

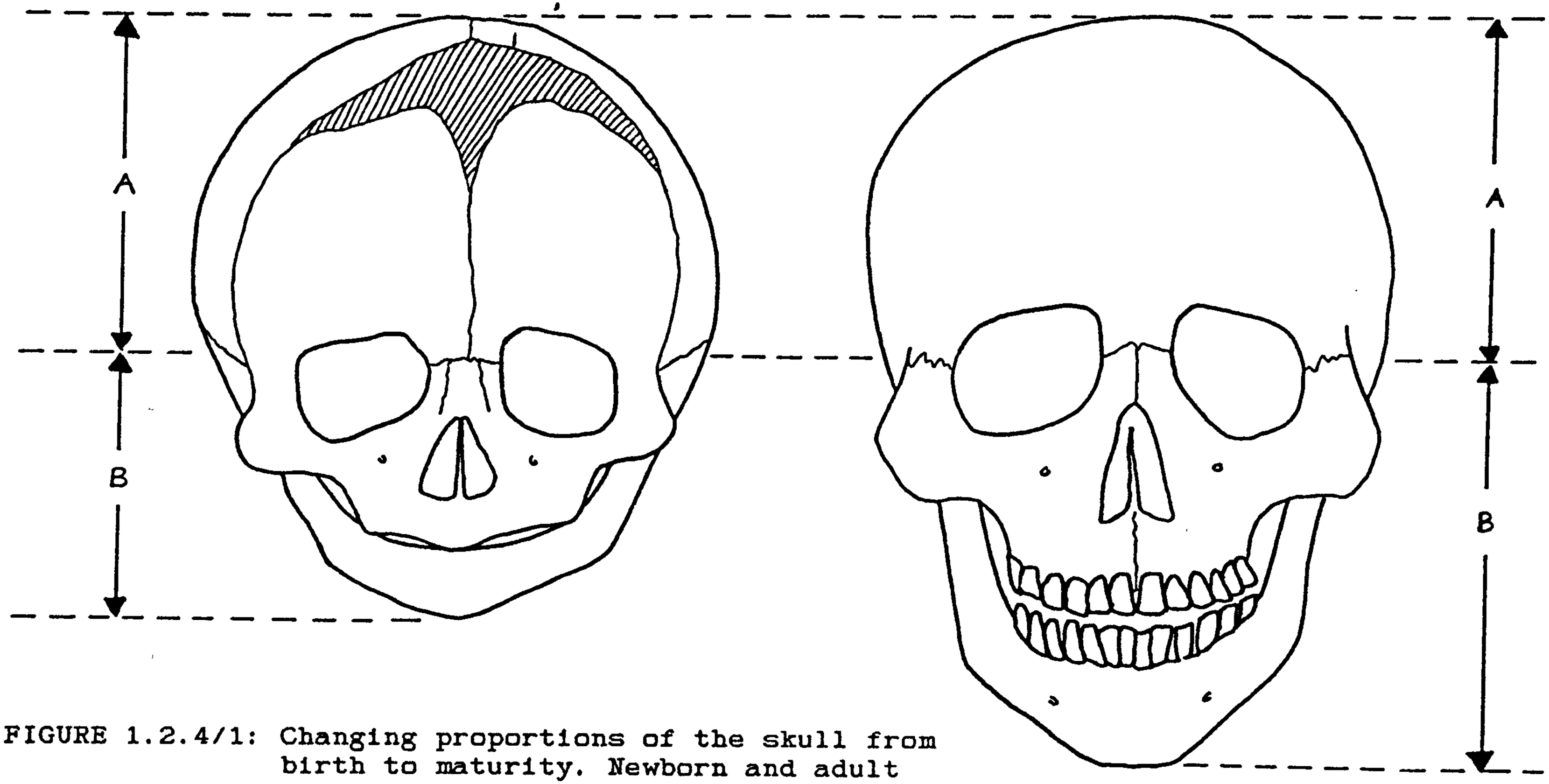


FIGURE 1.2.4/1: Changing proportions of the skull from birth to maturity. Newborn and adult skulls are drawn so that cranial height is the same. Note the proportionately larger face in the adult (Adapted from Sinclair 1978:94)

The cranium

The cranium, as might be expected given its role as protector of the brain, tends to follow a neural growth pattern (see Section 1.2.1.A), reflecting quite accurately the growth of the underlying brain (Bambha 1961, Watson and Lowrey 1967, Sinclair 1978, Tanner 1978). Growth rate is very high for the first 1 or 2 years of life, and then falls. 90% of adult size is attained within 4-5 years of birth, and growth is virtually complete by 10-12 years of age. The volume of the cranial vault is about 400 ml at birth, increasing to 950 ml at two years of age, compared with an adult volume of 1300 to 1500 ml (Sinclair 1978:72).

The cranium develops by the process of intramembranous ossification, and at birth there are quite large areas of fibrous connective tissue still separating the bones of the cranium (see Figure 1.2.4/2). During the first two years after birth, ossification gradually fills in these fontanelles, and the bones of the cranial vault come into contact with each other. The junctions between the individual bones, which are known as sutures, are lined with fibrous connective tissue. These are the principle sites of rapid cranial growth during the next few years of childhood. Growth also proceeds by the apposition of new bone on the external surfaces of the cranium. Simultaneous removal of bone from the inner surfaces by osteoclasts ensures that the correct thickness of the cranial bones is maintained. Remodelling of the bone also continues, to produce a bone structure consisting of spongy bone sandwiched between two layers of dense bone. In this way, growth at the sutures progressively pushes the individual bones apart, whilst external apposition of bone and remodelling ensure that radial growth results in the correct cranial contour. These two major

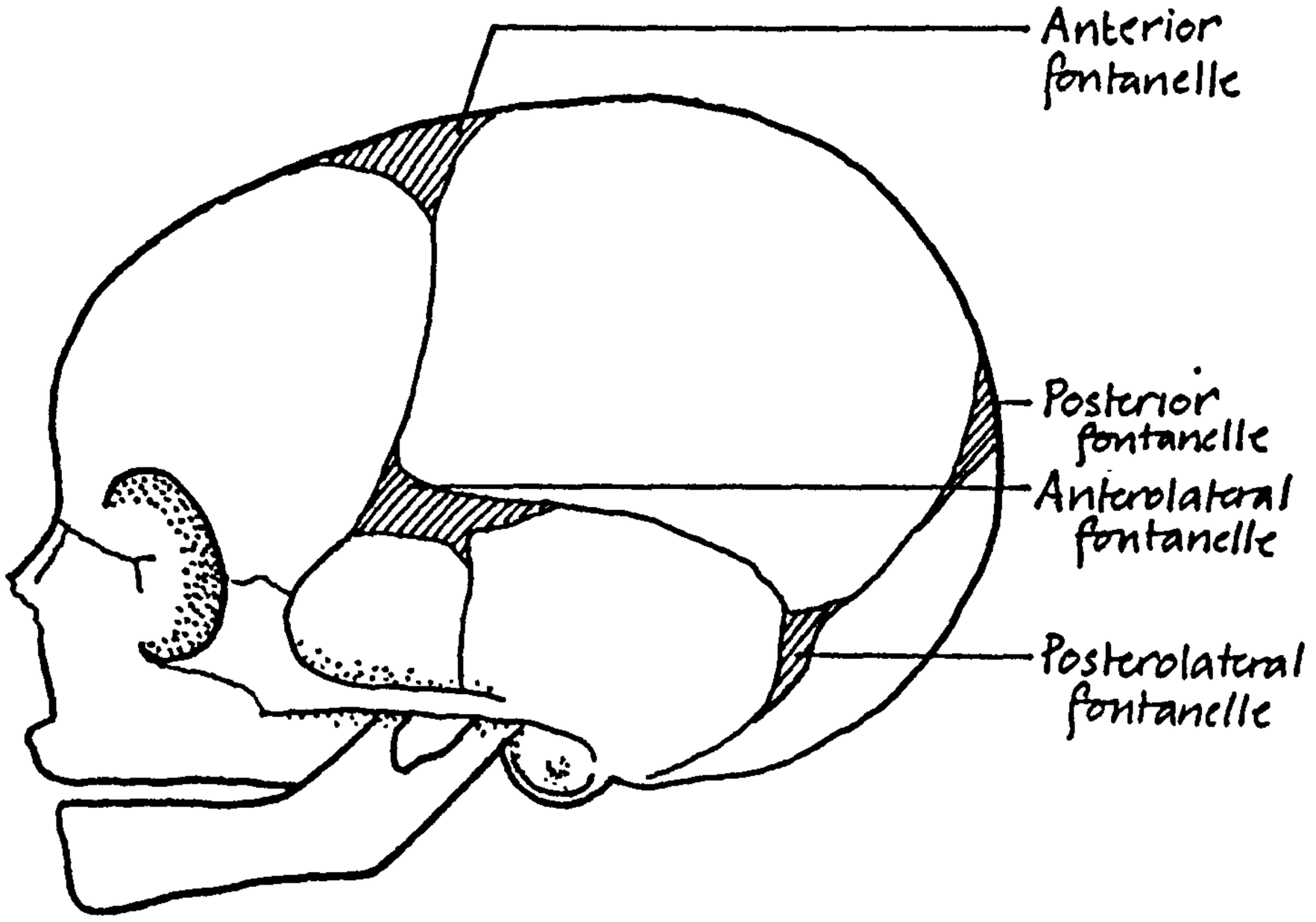


FIGURE 1.2.4/2: Sagittal view of the skull at birth, showing fontanelles (adapted from Sinclair 1978:55)

mechanisms of cranial growth are shown schematically in Figure 1.2.4/3.

Growth at the sutures slows down dramatically after early childhood, and is complete by puberty. The sutures form a closely serrated interlocking pattern, and eventually ossification at the suture lines fuses the bones together. This may begin during the third decade of life, continuing on into the fifties.

The cranium is one of the few parts of the body which is not significantly affected by the adolescent growth spurt. Bambha (1961) found an occasional growth increase at this time, and Tanner (1978:69) reports a small increase in head diameter in most persons, which may be largely accounted for by a 15% increase in bone thickness, and by a thickening of scalp tissue.

Ingerslev and Solow (1975) found that in the Swedish population the cranium is significantly smaller in females, and these findings echo other reports of sexual differentiation in cranial size (Wei 1970). Shape seems to show little sexual differentiation, except that the frontal bone may be more prominent in women (Ingerslev and Solow 1975). Figure 1.2.4/4 shows average male and female differences in cranial measurements.

The facial skeleton

The main constituent parts and landmarks of the facial skeleton are shown in Figure 1.2.4/5.

There is a large and often controversial literature concerning development of the facial skeleton. Disagreement about normal patterns of growth arise partly from the high degree of real variability in facial

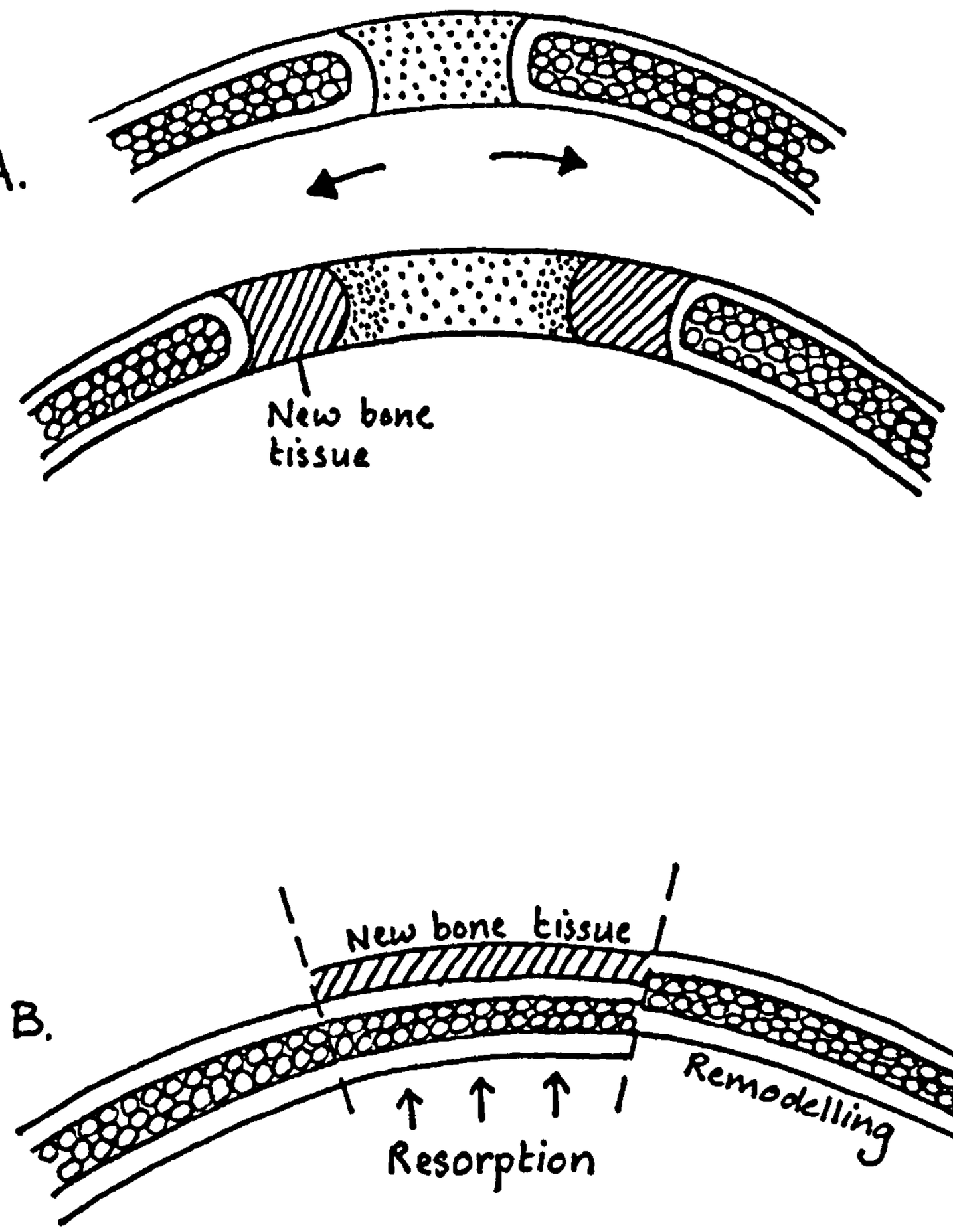


FIGURE 1.2.4/3: Schematic diagram of cranial growth mechanisms (adapted from Sinclair 1978:56)

A.= Sutural growth
B.= Appositional growth

	MALE		FEMALE	
	MEAN	RANGE	MEAN	RANGE
Bizygomatic width	124.7mm	11.5	120.58mm	13.0
Bigonial width	93.2	11.5	87.7	13.0
Maxillary canine width	36.6	11.0	38.3	6.0
Mandibular canine width	31.0	4.0	30.3	5.5

FIGURE 1.2.4/4: Sex differences in cranial width measurements. Data for 14- and 15-year old subjects from Woods (1950), cited in Wei (1970:144)

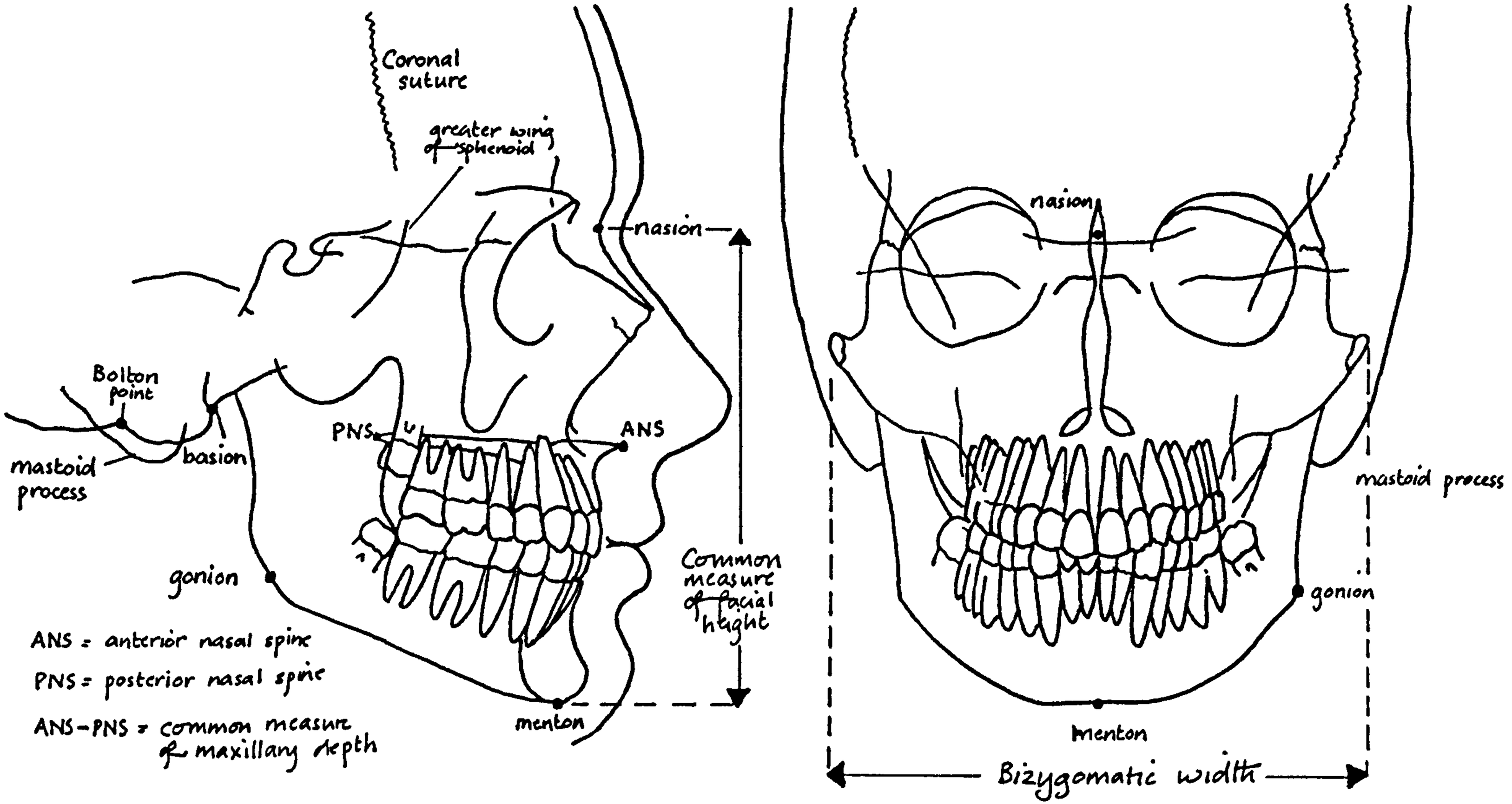


FIGURE 1.2.4/5: Landmarks of the facial skeleton

Adapted from several sources

SEE ALSO FIGURE 1.1.1/6

morphology and growth, and partly from the variety of cephalometric techniques used.

Variability in facial structure obviously has a large genetic component, as evidenced by the observation that different ethnic groups show very different facial characteristics, but facial growth patterns also display a high degree of plasticity, responding quite readily to environmental factors. A certain amount of flexibility in the growth patterns of the various parts of the facial skeleton is presumably an adaptive response to the need for very complex coordination of growth of the many bones and cartilages which make up the facial skeleton. The growth of each part must be carefully timed so as to maintain harmony of the overall facial structure, and it may be that the best way of achieving this harmony is for each growth area to be especially sensitive to its skeletal and soft tissue environment. The problem of coordinating growth is not, of course, unique to the face, but the complexity of the skeleton in this area makes it particularly crucial. The observation that facial characteristics are highly prone to disturbance by a wide variety of genetic and environmental abnormalities (Martin 1961), ranging from Down's Syndrome to foetal alchohol syndrome (where the foetus is exposed to high levels of alcohol in the maternal blood stream), is indicative of the level of sensitivity to general growth disturbance displayed by the facial skeleton. One virtue of this plasticity is the success of orthodontic treatment and plastic surgery in the treatment of such disorders as cleft palate.

Cephalometry conventionally involves the measurement of distances and angles between skeletal registration points. Registration points are chosen for their supposed stability during bone growth and movement (Scott 1967). The complexity of facial growth, and the ability of bone

to remodel itself, make it very difficult to find fixed points. Different workers tend to choose different registration points, depending on the emphasis of their research, or possibly on the populations under study. This makes the comparison of different cephalometric studies very difficult. Bjork (1966) demonstrated very clearly the way in which dependence on registration points can lead to false conclusions about bone growth. He inserted metallic implants into the mandible at conventional registration points, and followed their movements during facial development. The amount of movement showed that some registration points are far from stable. What appears to be a linear consistency of growth pattern and direction in an individual may actually result from rotation of the mandible, with remodelling along its lower border. Remodelling tends to recreate former spatial relationships. Whilst a detailed analysis of the mechanism by which a bone grows is less important to a phonetician than the absolute size and shape of a bone at any one time, it is as well to be aware that the formation of theories about facial development may be influenced by the choice of measurement techniques.

In spite of these difficulties, it is possible to make some generalizations about facial growth. These generalizations may be biased towards an ideal view of growth, however, since many studies use children of "good dental health" (e.g. Walker and Kowalski 1972:111) or normal occlusal relationships (Knott 1961, Baber et al. 1965, Andria and Dias 1978, Shah et al 1980).

The facial skeleton and the cranial base follow the general body growth curve (see Section 1.2.1.A) much more closely than does the cranium. In early childhood, growth is closely related to development of the muscles of mastication, the tongue and the dentition. There is a

pronounced adolescent growth spurt in most measurements (Rose 1953, Bambha 1961, Hunter 1966, Bergerson 1972, Tofani 1972, Dermaut and O'Reilly 1978, Shah et al. 1980), but the precise timing of the growth spurt relative to bone age and overall stature increase seems to be somewhat variable, and may depend on the measurements used, the sex of the subjects, and the racial group. Bambha (1961), for example, found that the adolescent growth spurt lags a little behind the body height growth spurt, but Hunter (1966) found it to be coincident. He also found that females showed more heterogeneity in chronological and skeletal age at the time of maximum facial growth. In females, facial growth is usually almost completed in the late teens, by the time that maximum body height is attained, but in males facial growth continues to be a marked feature after cessation of overall growth of height (Bambha 1961, Hunter 1966), and may continue into the mid-twenties. The growth of the mandible seems to show the closest correlation with overall body growth curves (Hunter 1966). Generally, growth in facial width is completed earlier than growth in the anteroposterior dimension, and vertical growth of the face may continue into the third decade of life.

The various component sections of the facial skeleton will be considered separately, although vocal tract configuration depends as much on the relationship between these sections as on the shape or absolute size of each.

Palate and maxilla

Growth in size of the maxilla and palate is quite complex. Watson and Lowrey (1967) differentiate three anatomical regions of the nasomaxillary complex, which all show different growth patterns. During the first year of life the maxilla and palate increase in size primarily by generalised appositional growth, as new bone is laid

down around the bone surfaces. After this period, growth becomes localised to specific areas.

i. In early infancy the premaxillo-maxillary suture closes, and the length of the anterior portion of the palate and maxilla becomes fixed. At 4 to 5 years of age the sagittal suture begins to fuse, so that palatal width becomes fixed. Thereafter, alveolar width is increased by apposition of bone at the external surface of the alveolar bone.

ii. Bizygomatic width (see Figure 1.2.4/5) has a very different pattern of growth, increasing at a smoothly and steadily diminishing rate until adulthood. Growth in this dimension is particularly pronounced in males.

iii. Maxillary width increases by surface apposition of bone, keeping pace with palatal and bizygomatic widths. Height and length of the maxilla increase concurrently, as growth proceeds in a forward and downward direction.

The first good metrical data on palatal size was acquired by Redman, Shapiro and Gorlin (Shapiro et al. 1963, Redman et al. 1966), using a specially designed measurement device. They measured palatal height, width and length in more than a thousand caucasian Americans, aged from 6 years to adulthood, and these measurements give a good indication of trends in palatal growth in normal individuals. The results are summarised in Figure 1.2.4/6. Unfortunately these findings were not related to measurements of any other part of the craniofacial skeleton, or to overall body growth.

There seems to be a steady increase in all palatal dimensions until 10 or 11 years. After this point, mean width and length increase only slightly, but palatal height increases more rapidly until 16 to 18 years. This height increase is more marked in males. The relative height of the palate thus increases between the ages of

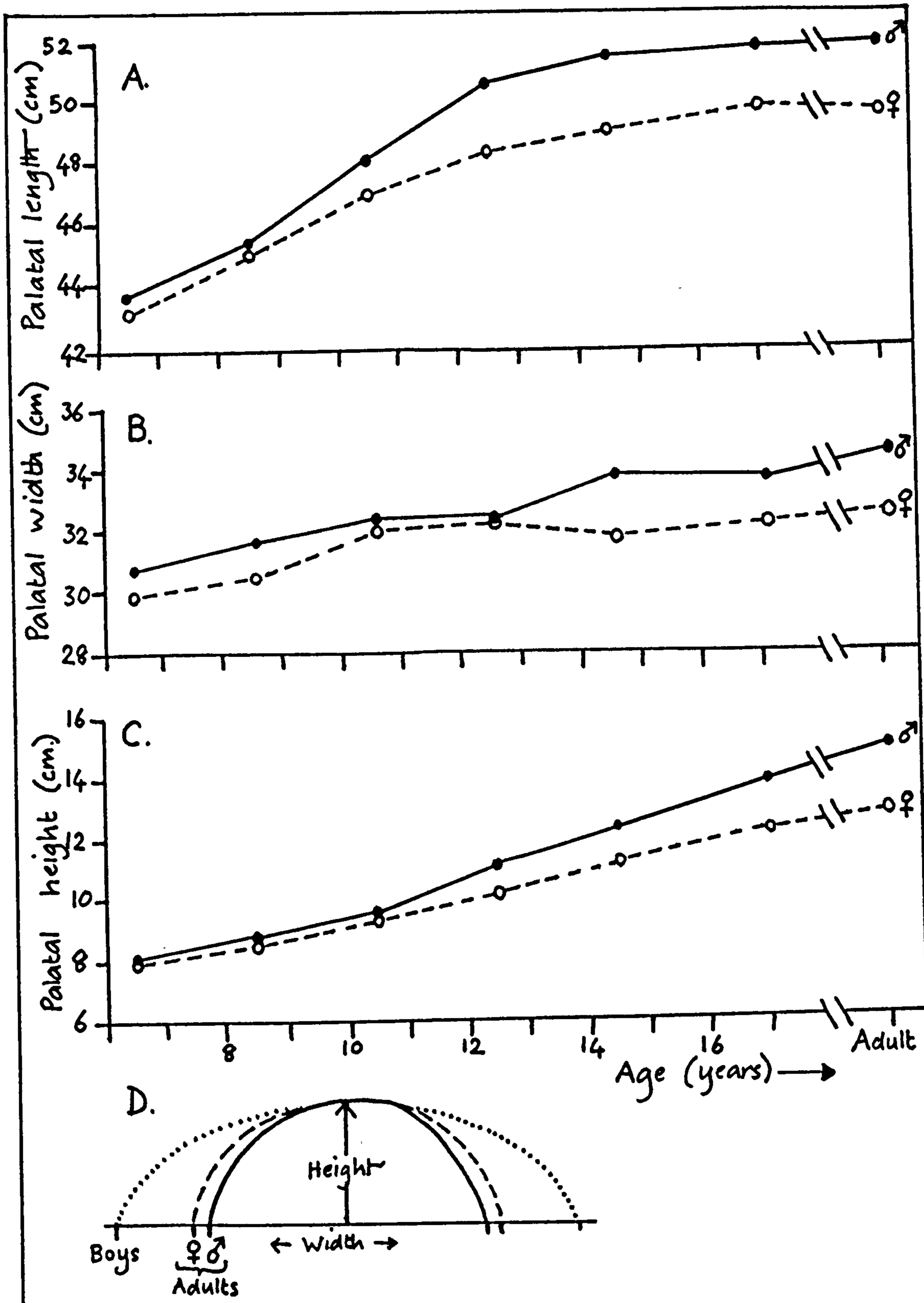


FIGURE 1.2.4/6: A.,B.,C.=Graphic representation of changing palatal dimensions with age D. shows relationship between height and width for 6-7 year old boys, women and men (data from Redman et al. 1966)

10 and 18. There is no significant sex difference in palatal measurements before 10 or 11 years, but after this the relatively greater height of the male palate becomes progressively more significant as adulthood is approached. All palatal dimensions are significantly larger in males from 14-15 years onwards. This accords with the finding of Ingerslev and Solow (1975) that in adult Danish subjects both length and width of the maxilla as a whole are significantly larger in males than in females.

O'Reilly (1979) studied the relationship between the timing of menarche and maxillary length in females, and found considerable variability, both in the timing of the maxillary growth spurt in relation to menarche and chronological age, and in the absolute length increase. The maxillary growth spurt typically lasts from 2 to 3 years, and occurs at some time between the ages of 11 and 15 years.

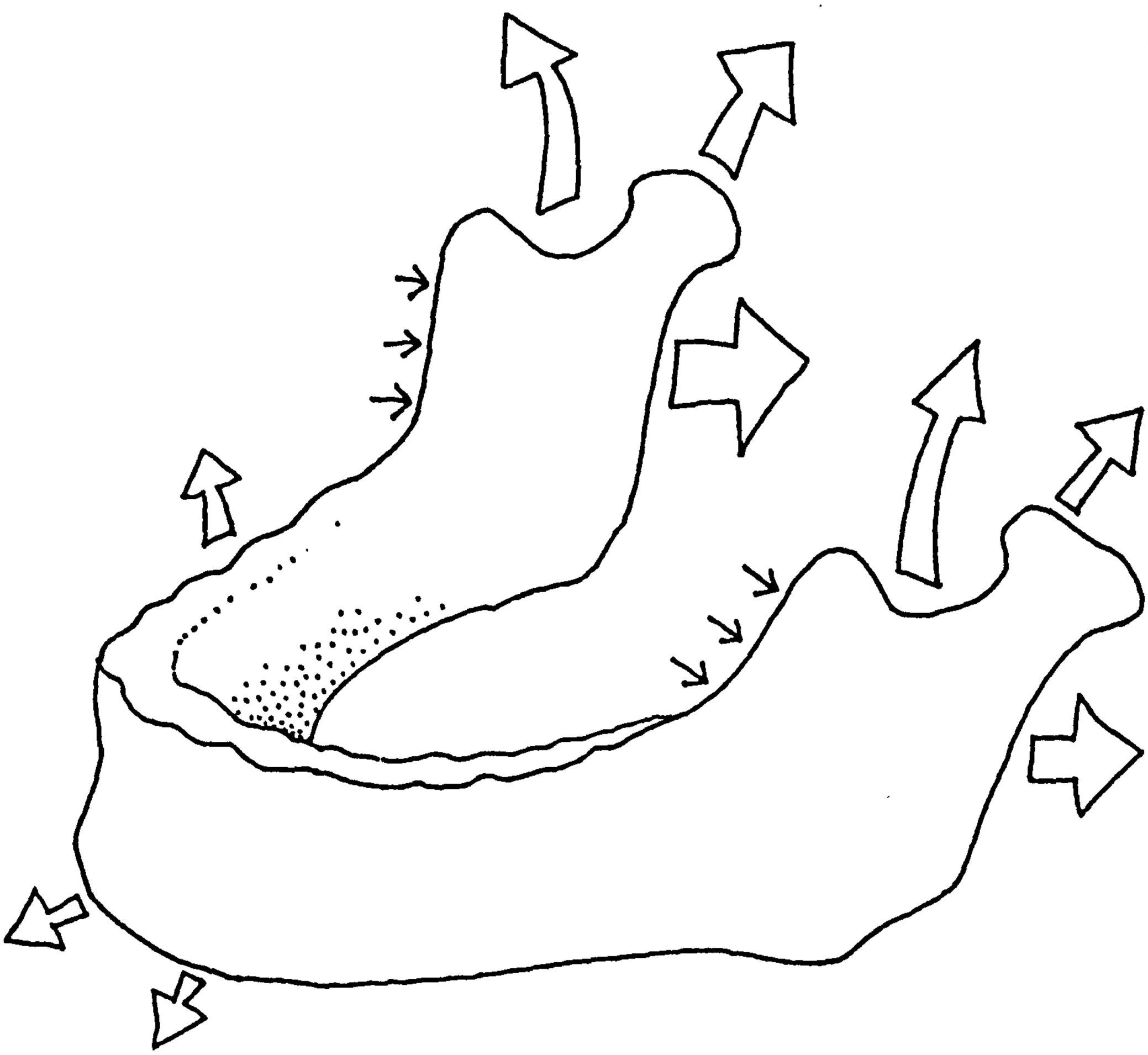
The maxilla shows some degenerative changes in old age, especially in the area of tooth insertion. As teeth are lost, the requirement for bone thickness in the tooth socket area is reduced, and bone tends to be lost.

The mandible

Growth of the mandible seems to be highly sensitive to a variety of factors. It seems to respond more to growth hormone than most other bones (Bevis et al 1977), and may also be more responsive to testosterone. It is also very sensitive to the muscular forces imposed upon it (Watson and Lowrey 1967). In terms of overall coordination, mandibular growth seems to be subordinate to maxillary growth. In other words, growth of the mandible seems to follow growth of the maxilla in such a way as to produce adequate occlusion.

The mechanism of mandibular growth is complex and very variable (Bjork 1966, Enlow and Harris 1964, Sinclair 1978:77). At birth, the mandible is very small, and is made up of two halves, separated by a layer of fibrous tissue in the midline. These two halves fuse during the first year of life. Increase in length follows the general bodily growth curve quite closely, with a greater and longer lasting growth spurt in males than in females, so that sexual dimorphism in mandibular length becomes quite marked by adulthood (Hunter 1966, Walker and Kowalski 1972, Ingerslev and Solow 1975). Figure 1.2.4/7, adapted from Enlow and Harris (1964) and Sinclair (1978:77), shows the main areas of mandibular growth and remodelling. Growth results primarily in a length increase, although width also increases to allow proper articulation with the skull. During the prepubertal phase, there is considerable appositional growth at the head of the mandible. There is also bone growth behind the ramus, accompanied by bone resorption at the front of the ramus, so that the space available for the dentition gradually increases. The angle between the ramus and the body of the mandible is gradually reduced from about 140° in infancy to 120° in adulthood (see Figure 1.2.4/8). The greatest contribution to overall facial growth at the time of puberty is made by the mandible. During this period, most growth continues in the ramus, but there are also marked increases in the length of the body and the vertical distance between the chin and the incisors.

As the mandible grows, the teeth move forwards to create space for the eruption of the molar teeth. This movement is achieved by resorption of bone from the anterior walls of the tooth sockets, and the addition of bone behind. When the deciduous teeth first erupt the upper and lower incisors are almost vertical, but the permanent incisors incline forwards to meet at an angle (Sinclair 1978:78 - see Figure 1.2.4/9).



→ = Resorption

⇨ = Direction of growth

FIGURE 1.2.4/7: Schematic diagram of mandibular growth (adapted from Enlow and Harris 1964:50 and Sinclair 1978:58)

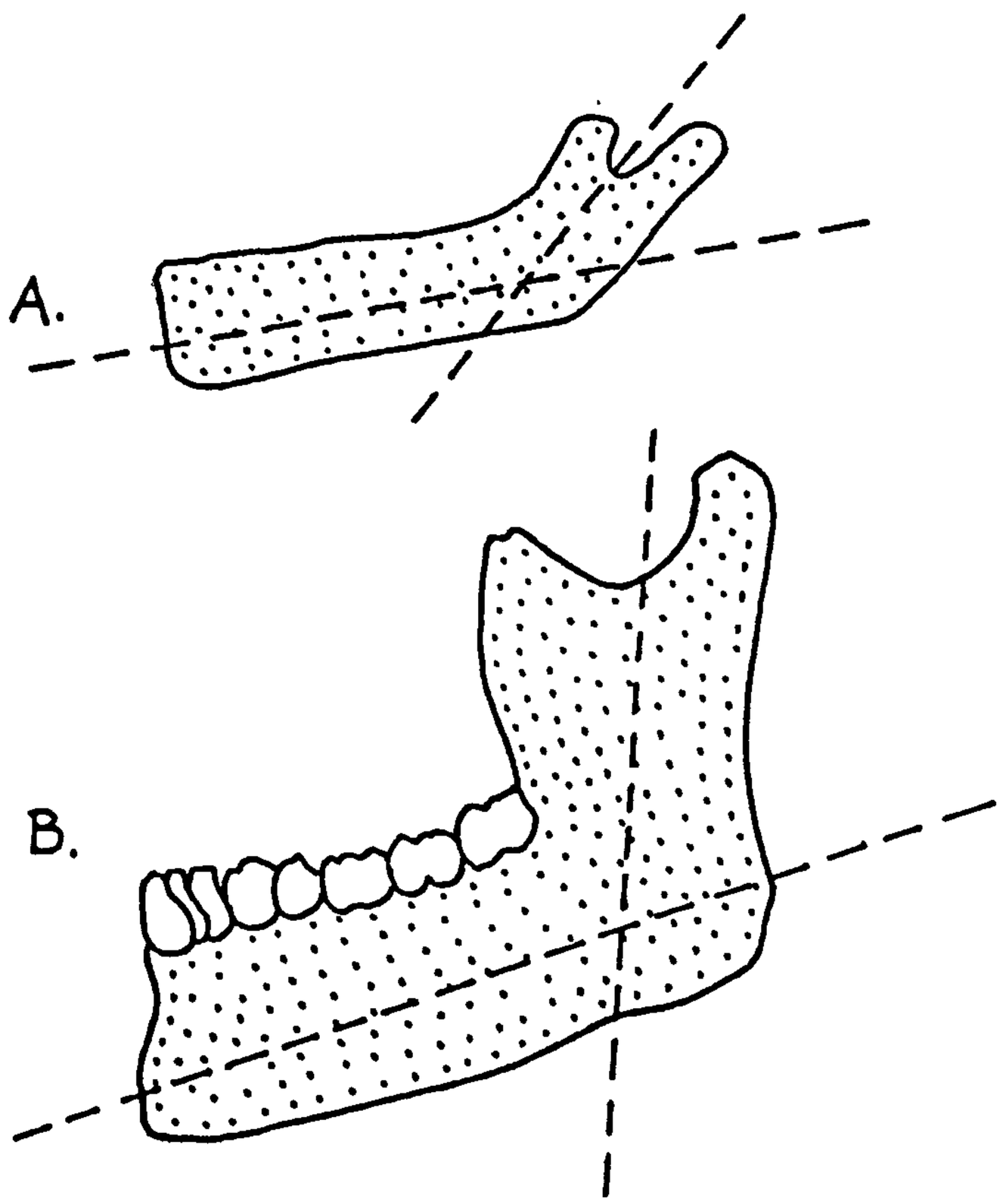
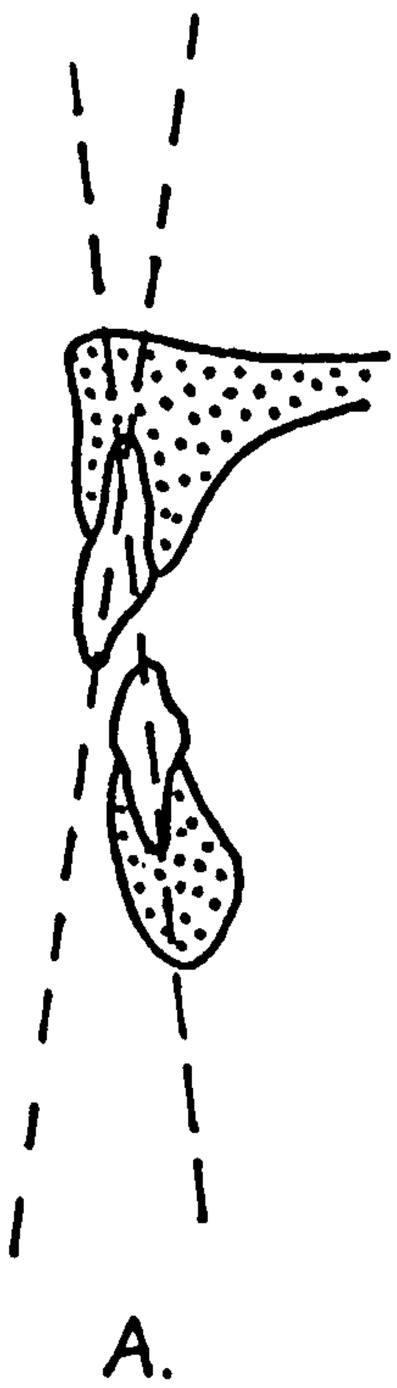
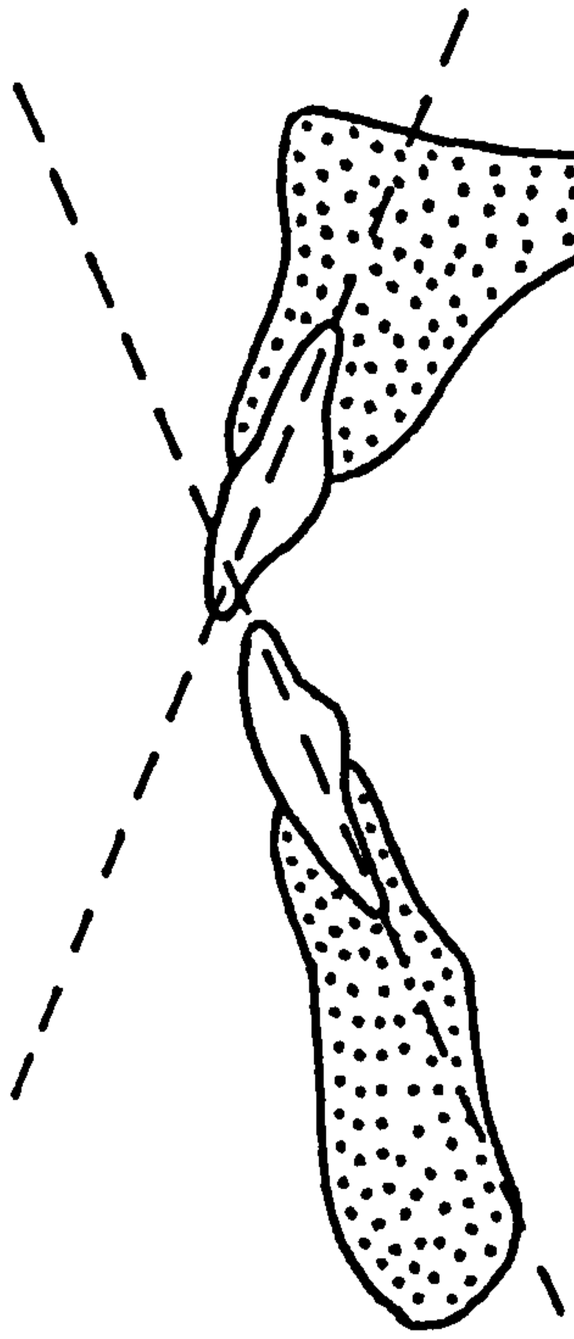


FIGURE 1.2.4/8: Changing mandibular angle (adapted from Sinclair 1978:55)
A. Infant B. Adult



A.



B.

FIGURE 1.2.4/9: Changing incisor relationships
A. Child B. Adult

As with the maxilla, loss of teeth is associated with bone resporption in the alveolar margin, so that the angle of the mandible becomes more obtuse, as in infancy, and may reach about 140° (Sinclair 1978:218).

Jaw relationships

It was mentioned earlier that growth of the mandible tends to accommodate itself to maxillary growth so that the upper and lower teeth meet in the correct relationship. This accommodation process is not foolproof, however, and minor problems of occlusion are not uncommon. These may be transient results of uncoordinated growth between the maxilla and mandible during childhood which are corrected by the later stages of mandibular growth, or they may persist into adulthood.

In normal occlusion of the teeth, the back surfaces of the maxillary teeth are in contact with the front surfaces of the mandibular teeth. Each lower tooth occludes with the corresponding upper tooth, and with the next most anterior upper tooth. The only exceptions are the lower central incisors which occlude only with the mandibular upper incisors. The vertical overlap of maxillary and mandibular incisors is as shown in Figure 1.2.4/9.

A malocclusion is most exactly defined as the abnormal relationship of one or more teeth to adjacent teeth in the same jaw, or to their normal antagonist in the opposing jaw (Hopkin 1978). The term is commonly used more loosely to describe a dento-facial anomaly, embracing any variations in morphology and relationships of the jaws and related craniofacial structures which can affect occlusion of the teeth.

Malocclusions are very common, although it is hard to suggest precise incidence figures since studies vary so

much in their standards of normality. At least 50% of individuals probably display at least a mild degree of malocclusion, but many of these will involve only the misplacement of a few teeth, and do not result from significant growth imbalances between the mandible and maxilla. The most commonly used classification of malocclusions was developed by Angle in 1899, and is based on the antero-posterior relationship of the maxillary and mandibular dental arches. The three main classes are summarised below.

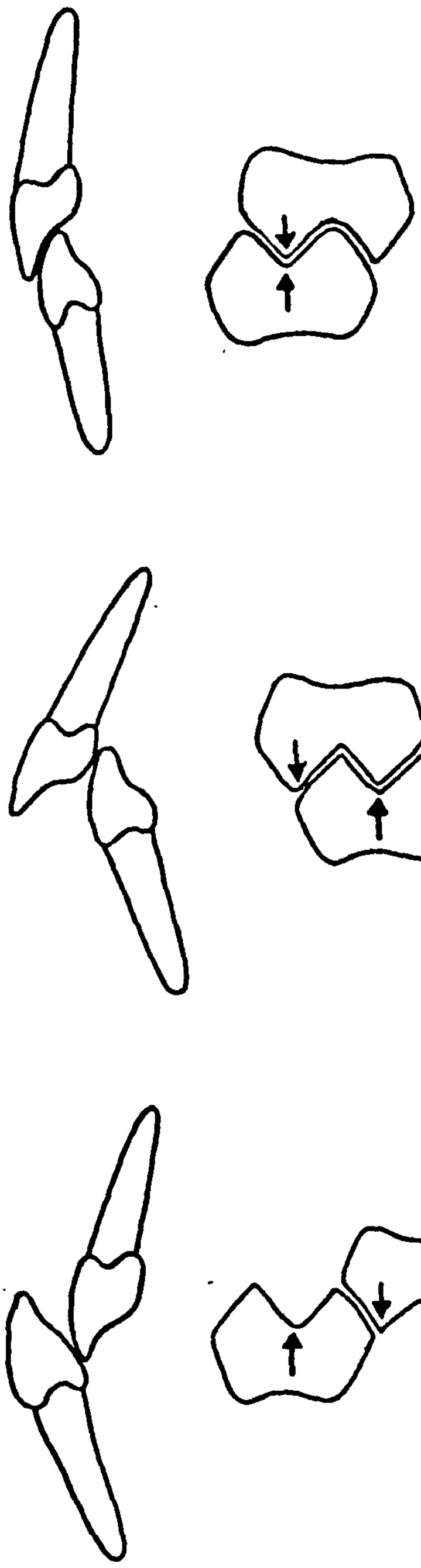
Class I: this class shows normal arch relationships, but malpositioning of one or more teeth.
Class II: in this class the mandibular arch is posterior to the maxillary arch. This class is further subdivided according to whether all the maxillary incisors protrude abnormally (= division 1) or only the lateral incisors (=division 2).
Class III: in this class the mandibular arch is anterior to the maxillary arch.

These types of malocclusion are shown schematically in Figure 1.2.4/10. Angle class I malocclusions account for about 60% of all malocclusions, Angle class II division 1 account for 25%, and the remaining malocclusions are fairly evenly spread between Angle class II division 2 and Angle class III (Hopkin 1978).

Vertebral and postural changes

Overall growth in length of the vertebral column is achieved partly by bone growth and partly by growth of the cartilage intervertebral discs. The bony vertebrae themselves grow at different rates, with a relatively larger size increase in the lumbar and sacral vertebrae than in the cervical and thoracic vertebrae. Vertebrae grow by the ossification of cartilage above and below the

A.



B.

C.

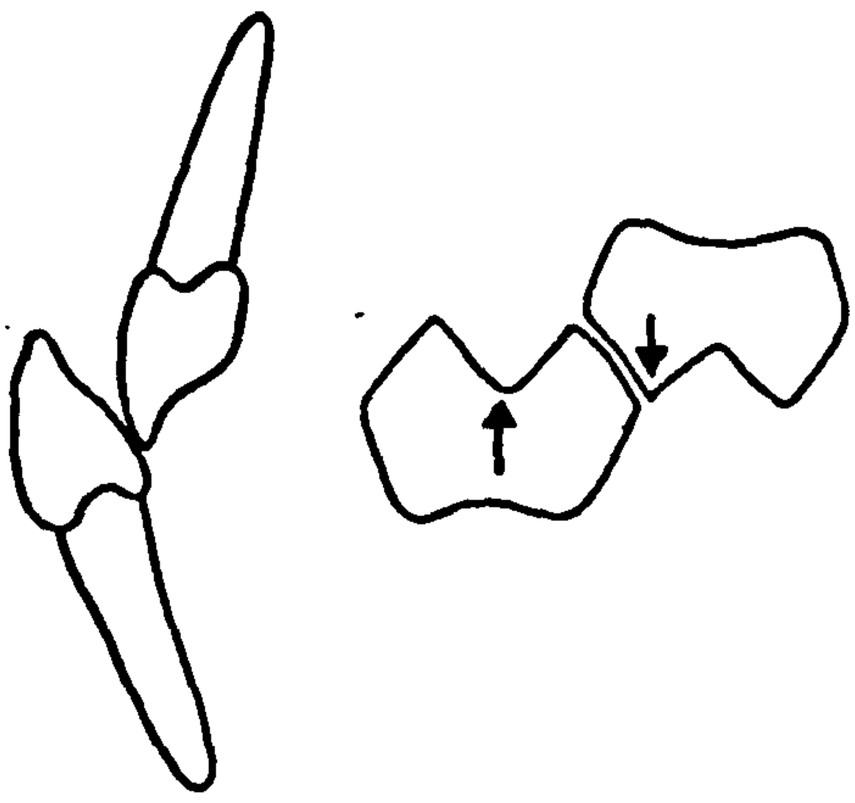


FIGURE 1.2.4/10: Angle classes of malocclusion
(adapted from Hopkin 1978)
A. Angle class I
B. Angle class II
C. Angle class III

existing bone, and ring shaped epiphyses may persist into the twenties (Sinclair 1978:79). The last epiphyses to close are those in the upper thoracic region, which may mean that the volume of the thoracic cavity can continue to increase for some time after the rest of the skeletal structure has reached its maximum size.

The contour of the vocal tract is to some extent dependent on the curvature of the upper spine and the angle of articulation between the skull and the vertebral column. At birth, the infant cannot support his or her head in an upright position, but by three months of age the head begins to be held up fairly steadily. At this stage a curvature in the cervical region of the vertebral column develops (see Figure 1.2.4/11), and this is only lost in extreme old age, as a result of degeneration of the intervertebral discs and loss of muscle tone. The centre of gravity of the head remains in front of the point of articulation with the vertebral column throughout life, so that some muscular effort is needed to keep it fully upright. Any loss of tone in the postural muscles therefore tends to allow the head to fall forwards, and this becomes increasingly common in extreme old age.

B. The lungs and thorax

At birth, the lungs are very small, both in mass and volume. During the first few weeks of life they expand greatly, and by the end of the first year the lungs have trebled in weight, and increased sixfold in volume (Sinclair 1978:89). After the first rapid period of growth, the lungs follow the general growth curve (Boyd 1952 - see Figure 1.2.4/12).

The internal structure of the lungs is vital for their efficient functioning, and this shows considerable change

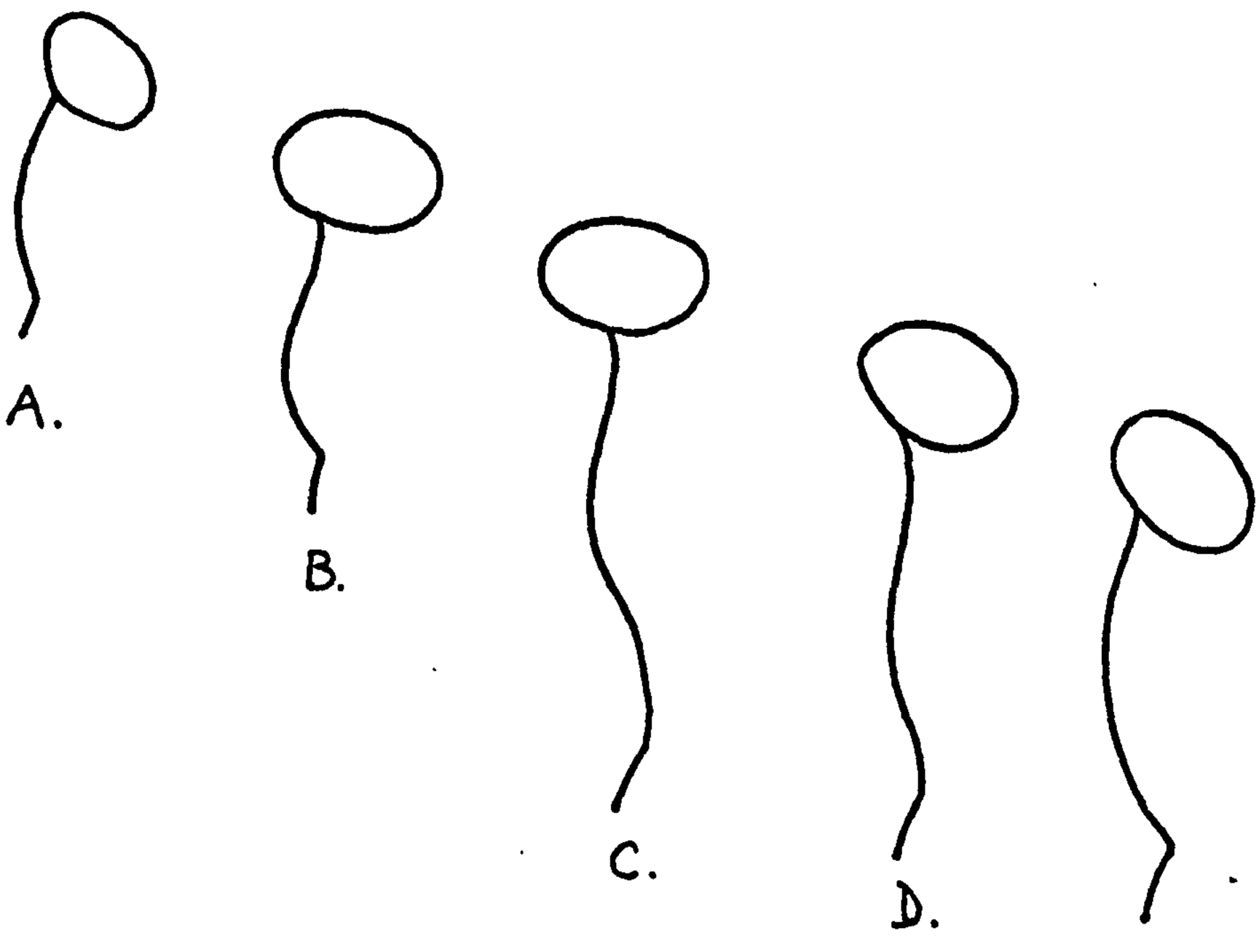


FIGURE 1.2.4/11: Spinal curvature changes from birth to old age (adapted from Sinclair 1978:101)
A. Infant, B. 6 months, C. Adult, D. Old age, E. Extreme old age

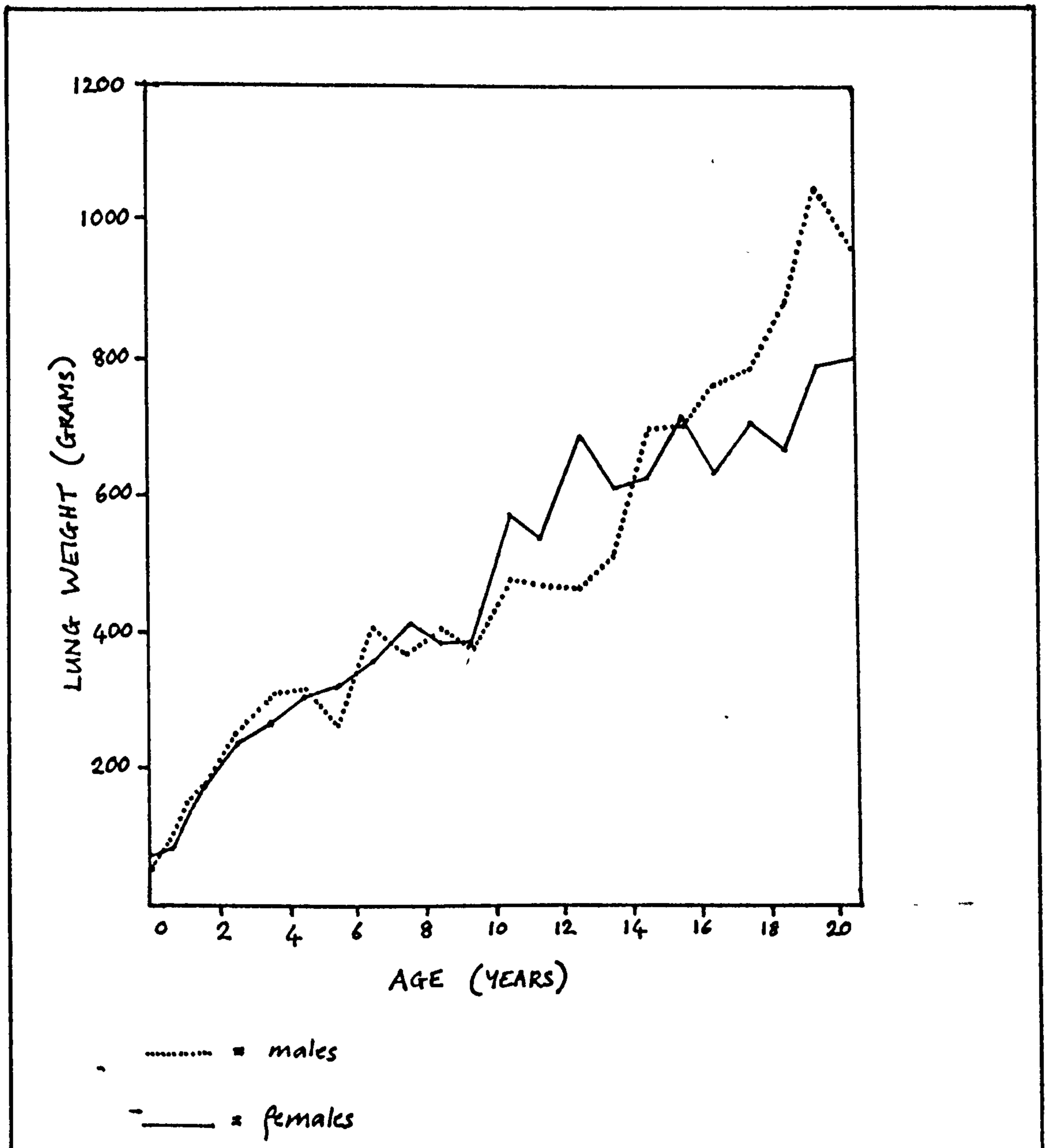


FIGURE 1.2.4/12: Lung weight growth curves for males and females (using data from Boyd 1952

following birth. The lungs do not seem, as had been thought, to be deflated at birth, but are filled with fluid. At birth, this fluid is replaced by air, and fluid is rapidly resorbed. Most of the alveoli of the lung are formed after birth, and the number of alveoli continues to increase until after puberty (Emery 1969). The development of alveoli seems to be associated with development of elastic fibres in the terminal airways. At birth, these are scarce, but they become gradually more dense, allowing the lungs to recoil more easily during expiration.

About 50% of the solid matter of the lungs is made up of collagen (Bouhuys 1970), and this probably functions to prevent overextension of the lungs during inspiration. Changes in the quality of the collagen network occur in old age, as the collagen molecules form cross links and become less flexible. This makes the whole lung structure less mobile, so that respiratory function is progressively impaired.

At birth, the whole of the thoracic skeleton and the shoulder girdle is rather high, as the pelvis is too small to accommodate the bladder and intestines, and so all the abdominal contents are compressed upwards towards the diaphragm (Sinclair 1978:119). Rapid pelvic development during the first two or three years of life allows the abdominal contents, and hence the thorax, to drop. The thoracic skeleton grows to accommodate the lungs, and follows a similar curve (Altman and Dittner 1962:334 - see Figure 1.2.4/13). The circumference of the thorax seems to be slightly larger in males than in females in chidhood, and this difference increases dramatically at puberty. The sternum is shorter in females, and in a slightly higher position relative to the vertebral column, and females also have rather more mobility of the upper ribs, allowing greater expansion of

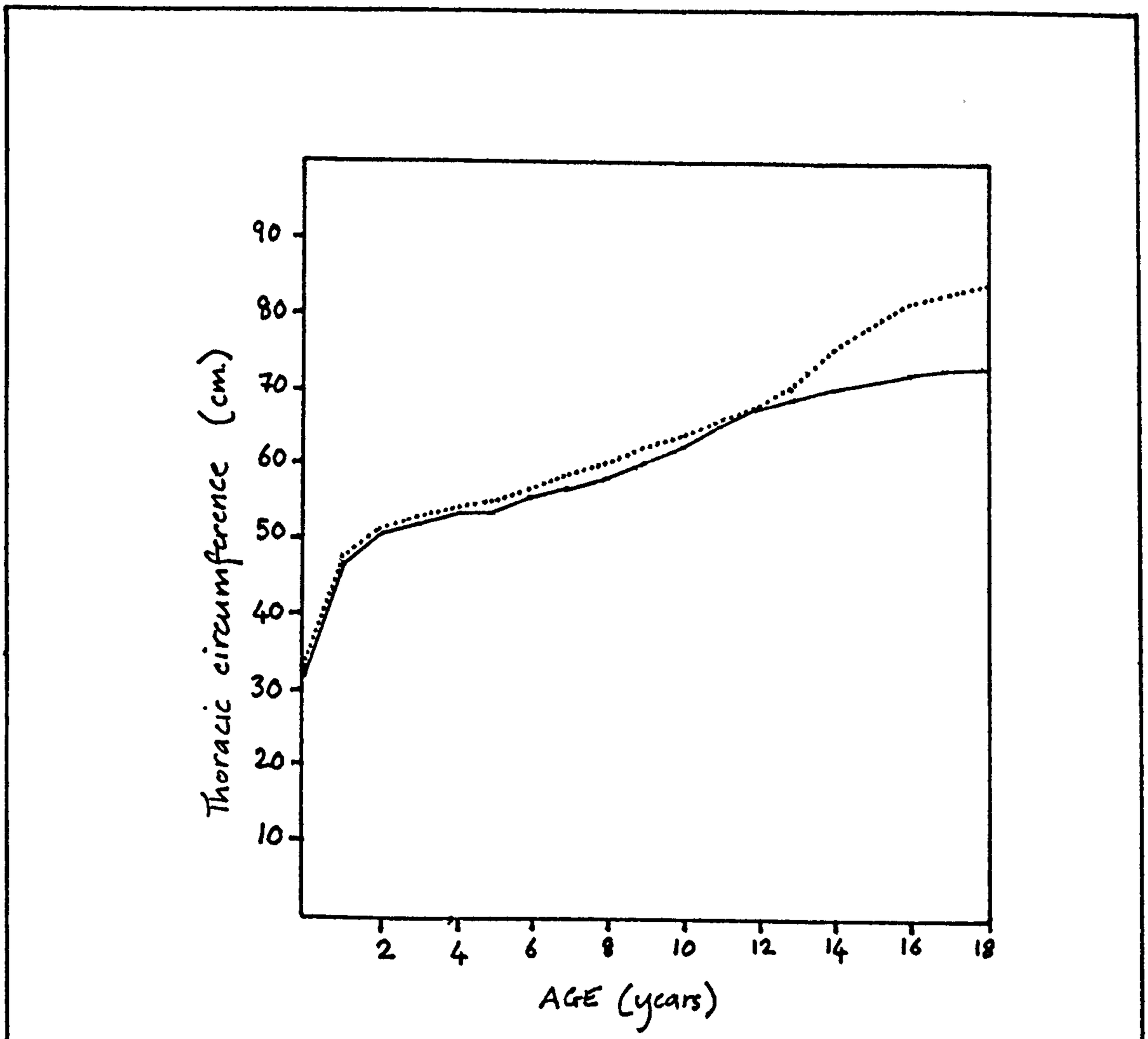


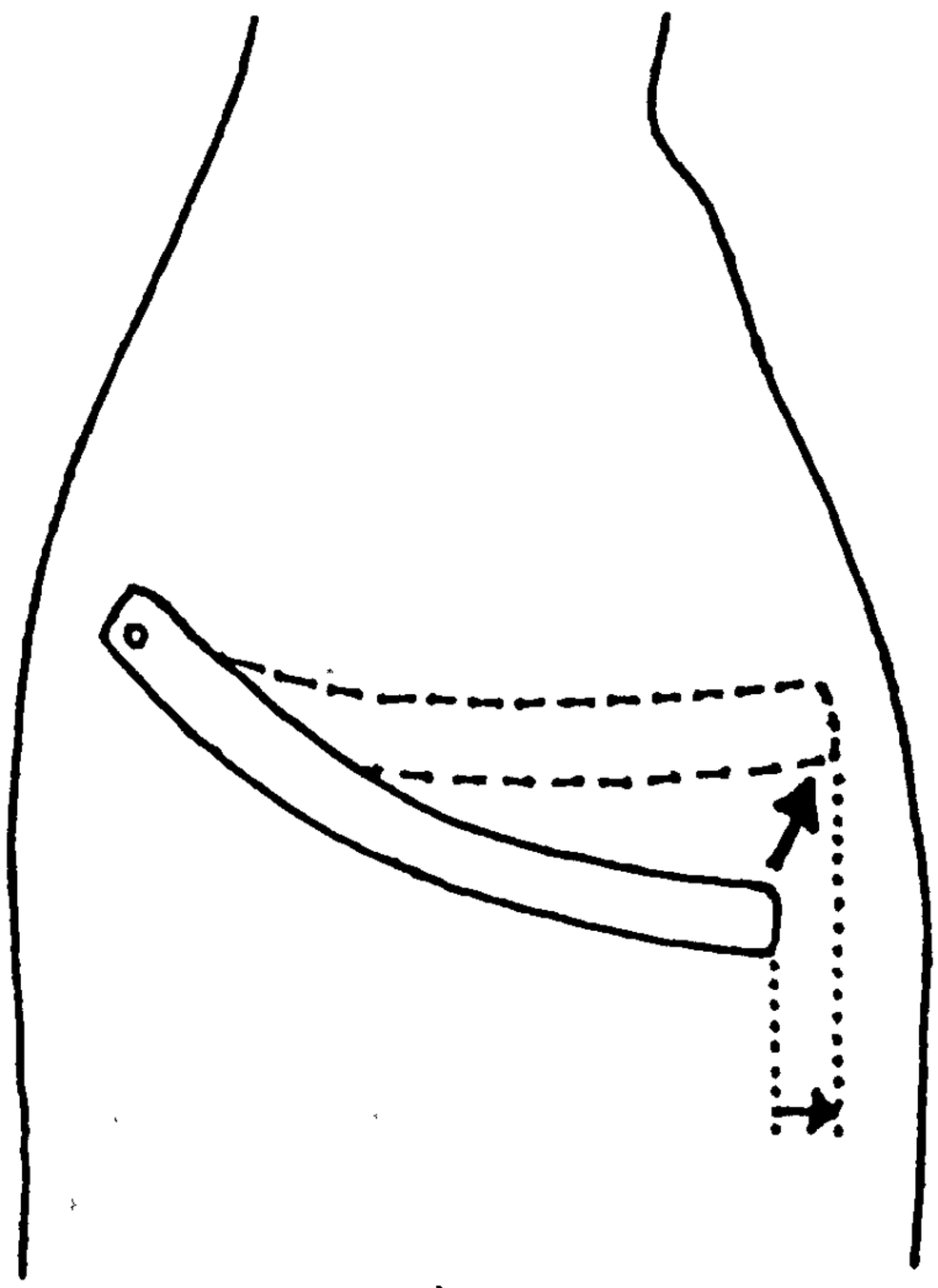
FIGURE 1.2.4/13: Thoracic circumference growth curves for males and females (using data from Altman and Dittner 1962)

the upper part of the thorax (Davies and Davies 1962:285). This is presumably an adaptation for pregnancy, when the lower thorax and diaphragm are constricted by the uterus.

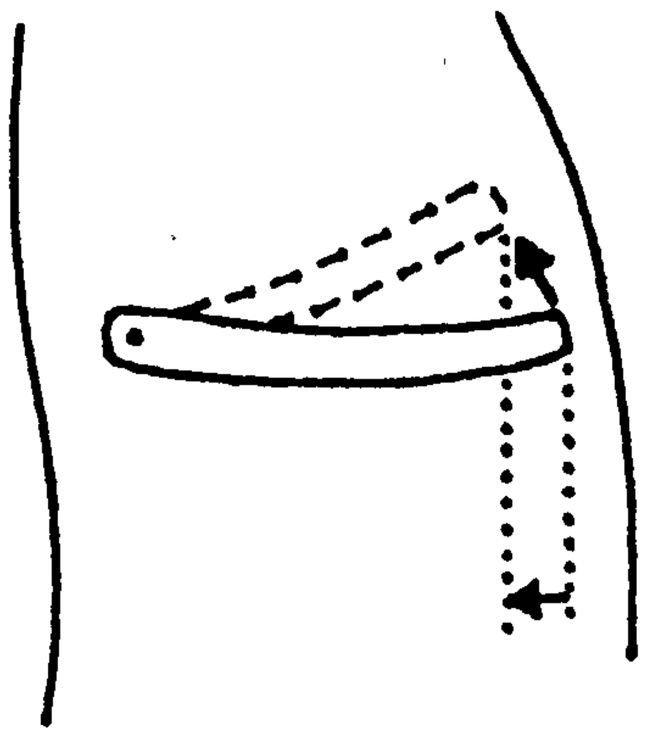
The angle of the ribs has important implications for the efficiency of respiration. In the adult, the ribs are angled downwards, and thoracic respiration increases the chest diameter by pulling the ribs to a more horizontal position. During the first two years of life the ribs lie more horizontally (Sinclair 1978:121), so that raising the ribs has little effect on chest volume (see Figure 1.2.4/14). The infant is thus much more dependent on diaphragmatic breathing. In old age, the state of the ribs again impedes respiration, as the rib cartilages become calcified, and thus lose their ability to twist and allow proper elevation of the ribs during inspiration. The vital capacity is consequently reduced, from a range of 3.5 to 5.9 litres in young adult males to a range of 2.4 to 4.7 litres after the age of 60 years (Sinclair 1978:223).

C. The larynx

Laryngeal growth during childhood has been very little studied. The position of the larynx in the new born is very high relative to other structures of the vocal tract, and the epiglottis makes contact with the soft palate. This contact is lost as there is progressive lowering of the epiglottis and larynx during the first year of life. At the age of six months the epiglottis and palate are well separated, although they make contact during swallowing, and by 12 to 18 months the contact even during swallowing is inconsistent (Sasaki et al. 1977).



A.



B.

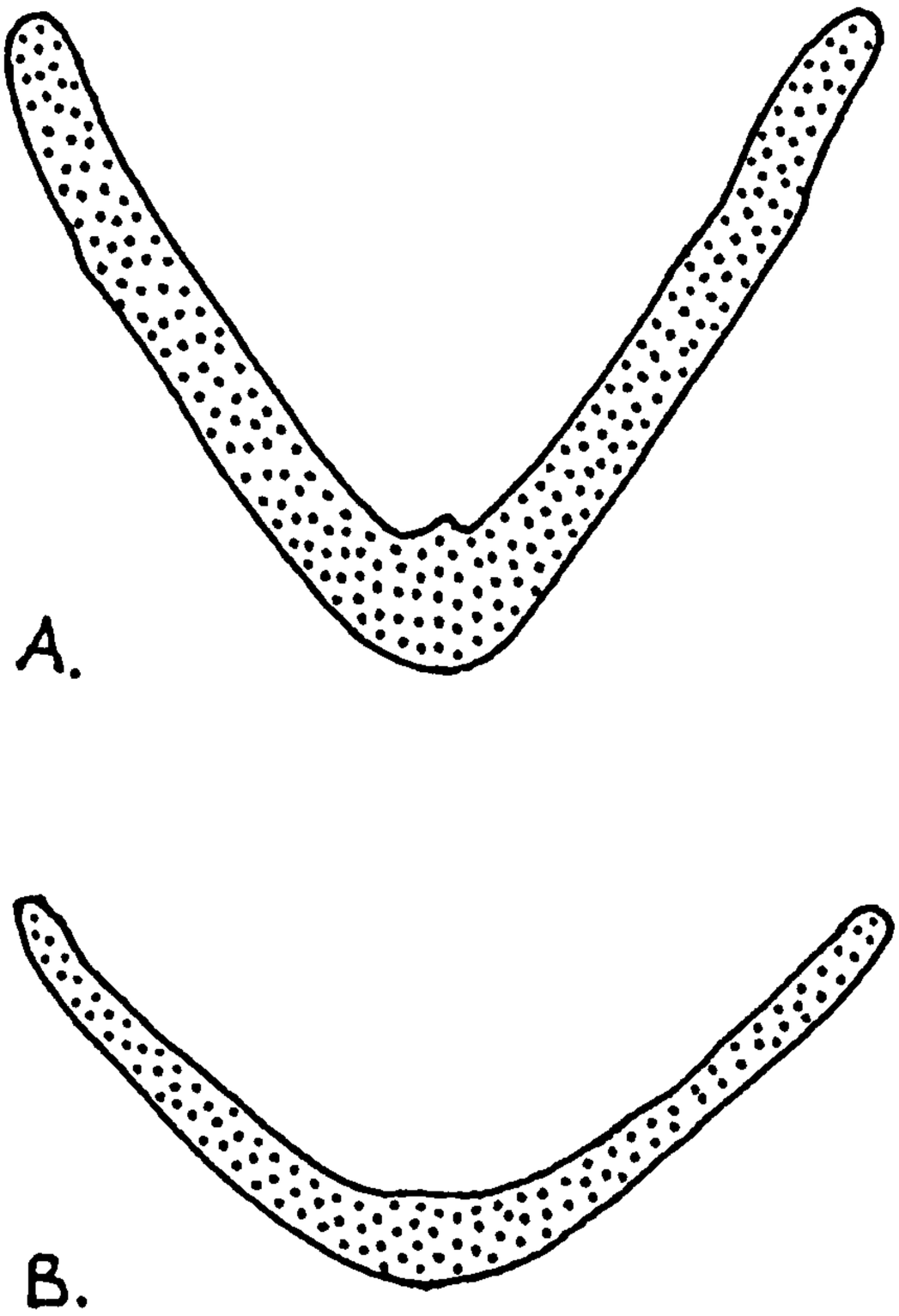
FIGURE 1.2.4/14: Rib elevation and thoracic volume in infant and adult (adapted from Sinclair 1978:120)
A. Elevating adult rib increases thoracic volume
B. In the baby, the rib is already horizontal, and elevation reduces thoracic volume

Dickson and Maue-Dickson (1982:176) report that growth of the laryngeal cartilages is linearly related to growth in height in both sexes, and that a rapid increase in size of the male cartilages at puberty results in significant adult sex differences. Maue (1970) and Maue and Dickson (1971) both cited in Dickson and Maue-Dickson (1982:142,148) give some measurements for male and female laryngeal cartilages which are summarised below.

Thyroid cartilage: this is very variable in many of its dimensions, but in all cases the male cartilage was found by Maue to be larger than the female. On average it weighs approximately twice as much in the male (8 gm) as in the female (4 gm). The average height of the thyroid cartilage from the tip of the superior horn to the tip of the inferior horn is 44 mm in males and 38 mm in females. The average anteroposterior measurement of the cartilage is 37 mm in males and 29 mm in females. The contour of the cartilage is rather different, also, with males having a more prominent angle. The laminae come together in a more rounded contour in females (see Figure 1.2.4/15).

Cricoid cartilage: this is less variable than the thyroid, and again the male measurements are consistently larger. Average laminar height is 25 mm for males and 19 mm for females. Average weight is 5.8 gm for males, about double the female weight of 2.89 gm.

Arytenoid cartilages: these show very little variability in size and shape within each sex, with an average height of 18 mm in males and 13 mm in females. Average anteroposterior length is 14 mm in males and 10 mm in females. Weight averages are 0.39 gm for males and 0.20 gm for females.



A.

B.

FIGURE 1.2.4/15: Sex differences in thyroid cartilage contour; superior view (adapted from Dickson and Maue-Dickson 1982)
A. Male, B. Female

During aging, laryngeal cartilages are subject to calcification, with consequent changes in elasticity of the cartilages. The age of onset of calcification varies considerably, and Zenker (1964, cited by Greene 1972:104) says that the thyroid cartilage may still be unaffected in some 70 year olds, although rigidity often sets in much earlier. Pantoja (1968), in a study of 100 adult thyroid cartilages, concurs with this, and reports that calcification typically begins in the inferior horns, progressing along the inferior and posterior borders, and then along the anterior border and angle.

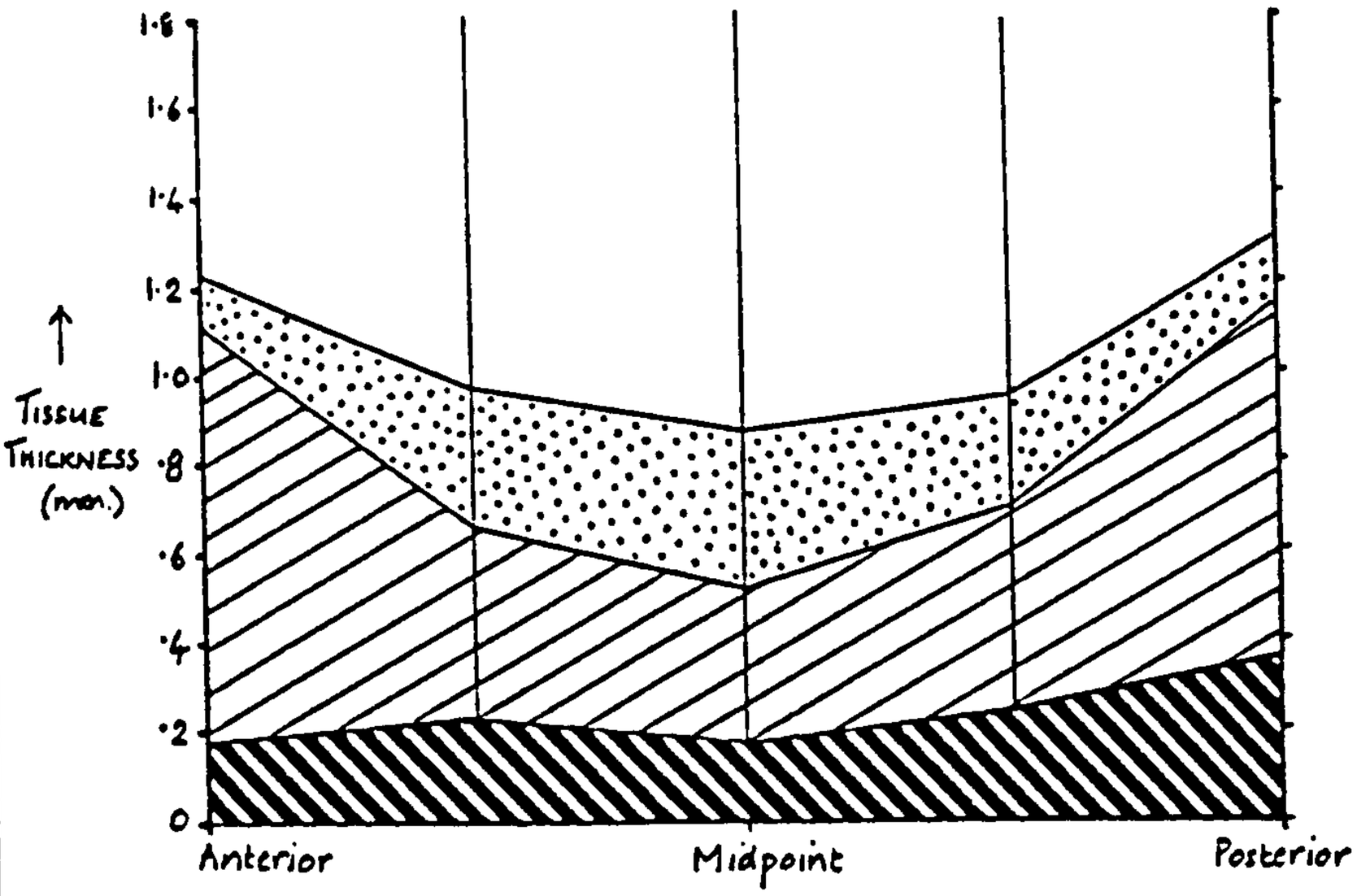
The whole larynx is extremely small at birth, but reported vocal fold length measurements are rather discrepant. Negus (1949) suggests a length of 3 mm at 14 days, growing to 5.5 mm at 1 year, 7.5 mm at 5 years, 8 mm at 6½ years, and 9.5 mm at 15 years. Terracol et al. (1956) report vocal fold lengths of 7-9 mm at 8 days, increasing to 15 mm at 5 years, after which there is little growth until the onset of puberty. There seems to be less disagreement about average adult vocal fold length, which is 23-25 mm in males, and about 17 mm in females (Morris 1953, Davies and Davies 1962, Greene 1972, Romanes 1978). The relative proportions of the ligamental and cartilaginous parts of the vocal folds are similar in both sexes, with the ligamental part constituting about two thirds of the total vocal fold length.

The structure of the vocal fold at birth is very immature. The fibres of the vocalis muscle are poorly developed, and Von Leden (1961) suggests that neuromuscular maturation of the larynx is not complete before three years. The tissue layers which make up the vocal ligament are also poorly differentiated, and adult tissue layer relationships are not seen until after puberty.

Hirano and colleagues (Hirano et al. 1981, Hirano et al. 1982) made histological examinations of 48 male and 40 female normal vocal folds taken from autopsy cases. The age range was from birth to 69 years. In new born infants there seems to be no vocal ligament, and the entire lamina propria seems to be rather uniform and pliable. The only areas of increased fibre density are at the ends of the ligamental portion of the vocal folds, and probably represent precursors of the maculae flavae. By four years of age an immature vocal ligament is present, but there is no differentiation between the elastic intermediate layer and the collagenous deep layer of the lamina propria. These two layers begin to be differentiated between the ages of 6 and 12, and by 15 years of age a clear differentiation is observed. Full maturation may not occur before 20 years of age, however, as before this the vocal ligament is sometimes thinner than in the adult, and the fibre arrangement is somewhat looser. The epithelium seems to show no significant changes during development.

After reaching maturity, too, there may be continuing changes in tissue thickness. Figures 1.2.4/16 and 1.2.4/17 represent the tissue thicknesses in two groups of subjects, in their 20s and in their 50s, using data from Hirano et al. (1982). A comparison suggests that in the older larynx there is an increase in the cover relative to the intermediate and deep layers of the lamina propria. Hirano et al. (1982:278) found no systematic age-related changes in epithelial thickness, so the increased thickness in the cover is attributed to changes in the superficial layer of the lamina propria. A decrease in fibre density in this layer is reported to be associated with oedema, which is more marked in males than in females. This pattern of generalized thickening of the cover corresponds very closely with clinical descriptions of Reinke's oedema (see Chapter 2.5), and it

i. Females



ii. Males

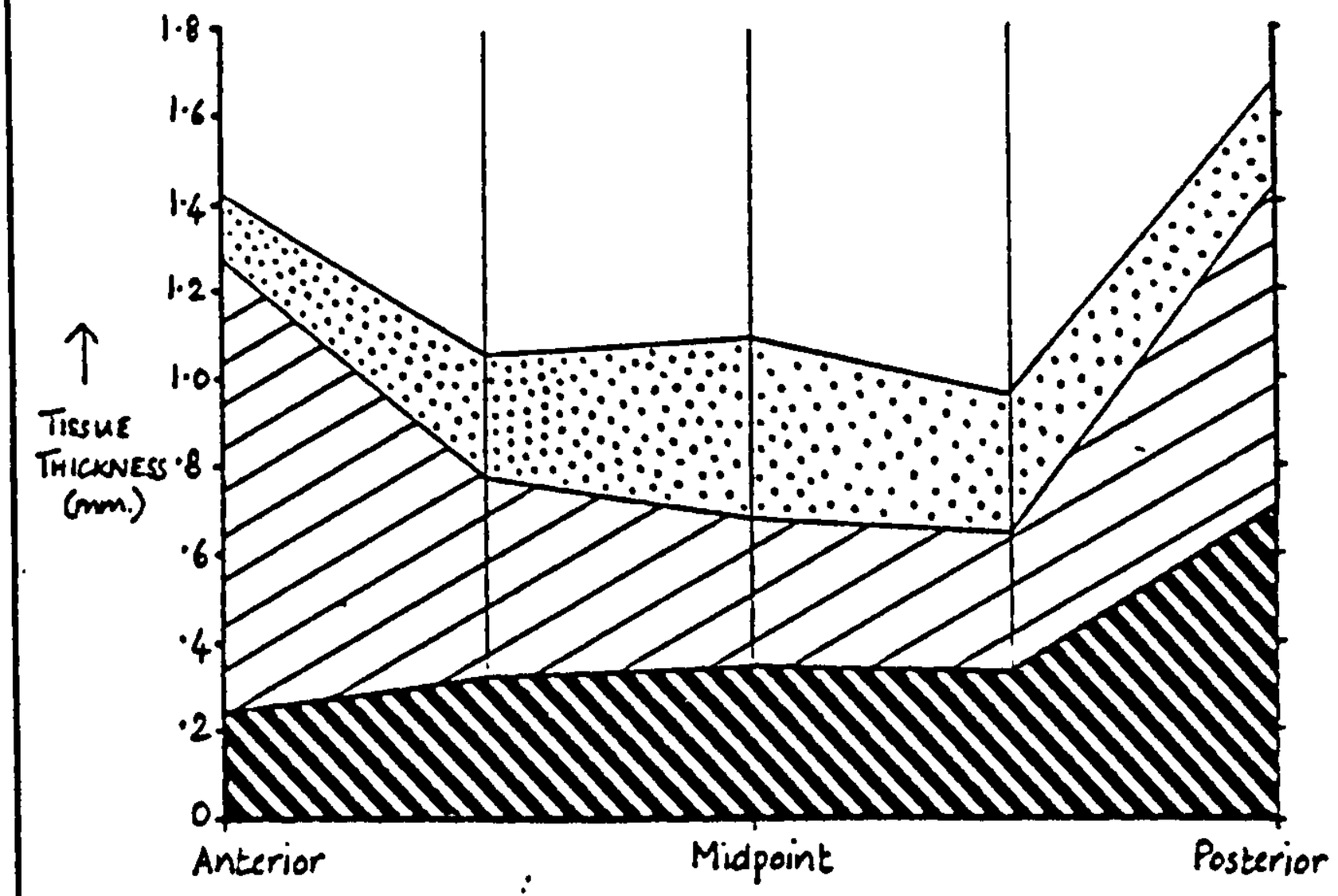
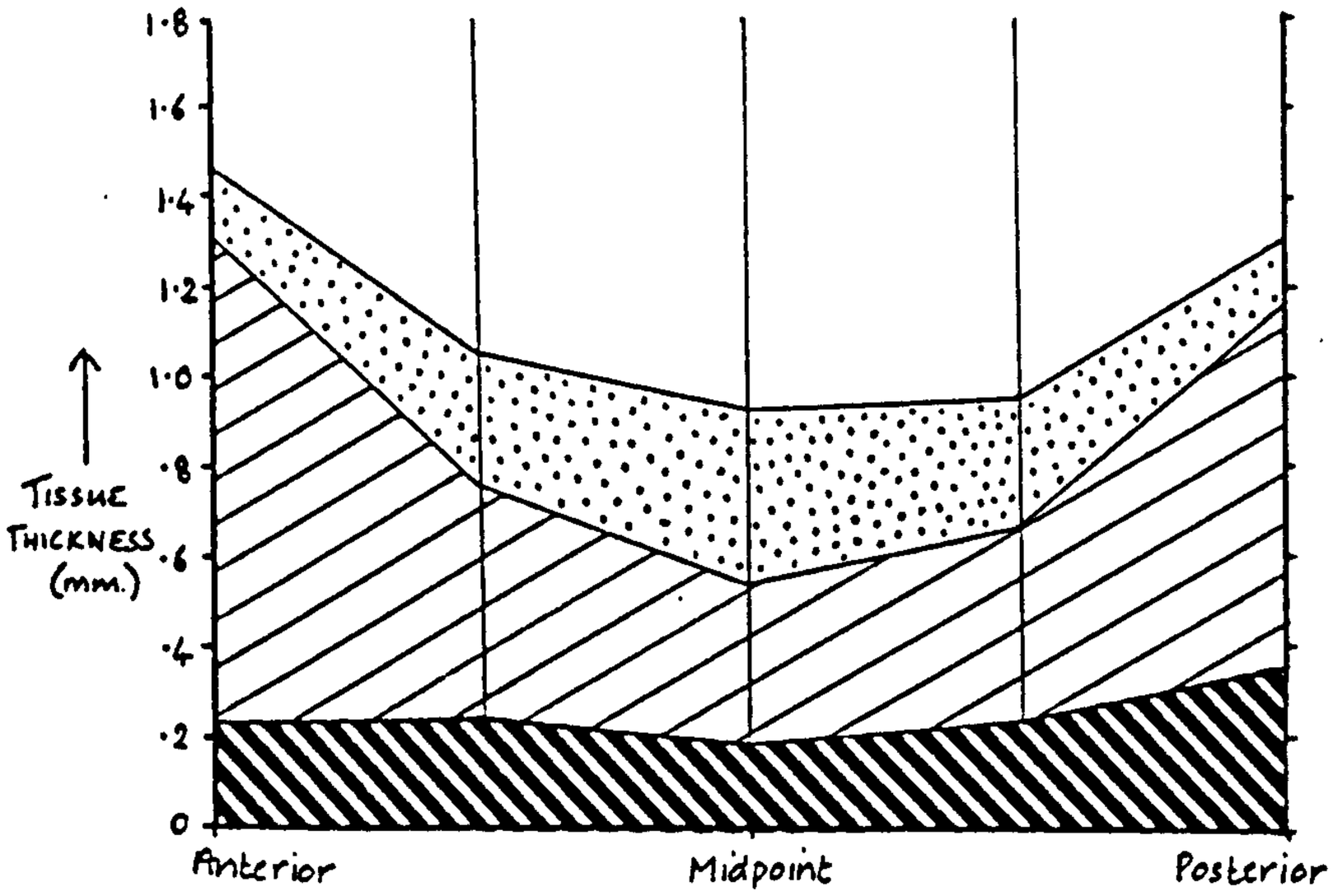
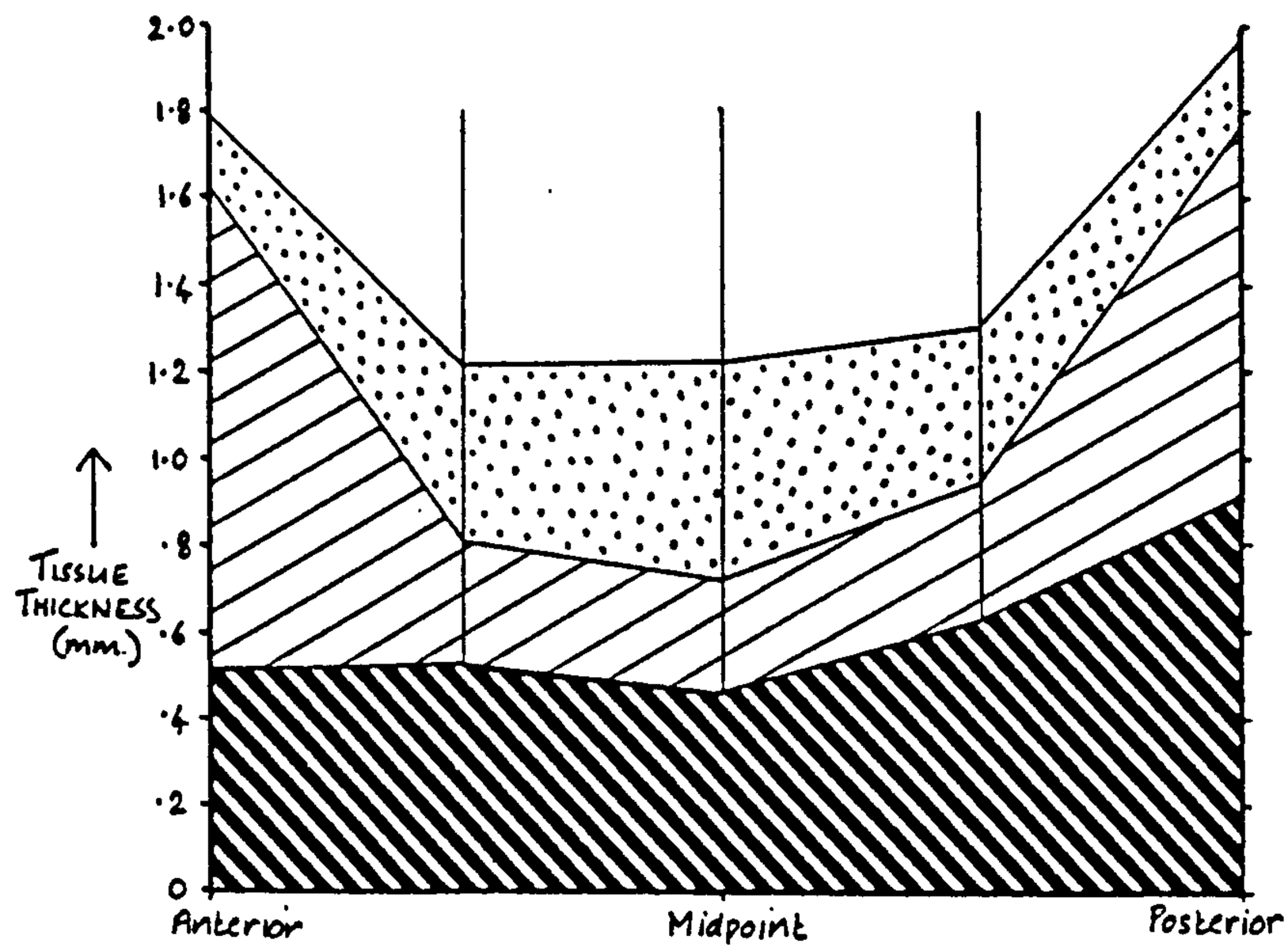


FIGURE 1.2.4/16: Graphic representation of vocal fold tissue thickness in subjects aged 20-29 years (using data from Hirano et al. 1982:274)

1. Females



ii. Males



Key as for previous figure

FIGURE 1.2.4/17: Graphic representation of vocal fold tissue thickness in subjects aged 50-59 years (using data from Hirano et al. 1982:274)

may be that some degree of Reinke's oedema is a common feature of aging.

The same study showed changes in the intermediate and deep layers of the lamina propria, which also seemed to be more common in males. The thickness of the intermediate layer decreases, and the contour of this layer may become distorted as a result of changes in the deep layer. The elastic fibres become looser and atrophied after about 40 years of age. The deep layer tends to thicken, and the collagen fibres become thicker and denser. Areas of fibrosis, where dense aggregates of collagen fibres are laid down haphazardly, develop in some males after the age of 40.

Honjo and Isshiki (1980) examined the larynges of 40 elderly subjects (20 men and 20 women), with a mean age of 75 years, and found marked oedema to be very characteristic, especially in females. It is interesting that this study found that females showed more oedema than men, in contrast to the results obtained by Hirano et al. (1982). This may be because the Honjo and Isshiki study concentrated on an older age group.

Yellowish or greyish discolouration of the vocal folds seems to occur quite often in older age groups (Luchsinger 1962, Honjo and Isshiki 1980, Mueller et al. 1985), and may indicate a degree of fatty degeneration or keratinization of the epithelium.

Atrophy of the laryngeal musculature, especially of the vocalis muscle which forms the bulk of the body of the vocal fold, is also a commonly reported feature of the aging larynx (Bach et al. 1941, Luchsinger 1962, Honjo and Isshiki 1980, Mueller et al. 1985). Honjo and Isshiki found that this was less marked in females. The result of a decrease in muscle power is an alteration in the

habitual phonatory posture of the vocal folds, so that bowing, where there is incomplete adduction at the centre of the vocal folds, becomes more common (Luchsinger 1962). Mueller et al. examined a series of 36 elderly larynges taken at autopsy from men aged 65 to 94 years (mean 78.8 years) with no history of laryngo-pulmonary disease or recent intubation. Comparing these with a younger adult control group (32-59 years, mean 49.4 years), they found a striking difference in laryngeal posture, which they explain in terms of muscular aging. The normal cadaveric vocal fold posture, which was found in all but one of the controls, is fully adducted. Only 19% of the elderly larynges had this posture. The remainder had either bowed vocal folds, vocal fold sulcus (= a longitudinal groove at the vocal fold edge in the ligamental area) or what the authors aptly describe as an "arrowhead" configuration. In this configuration there is partial approximation of the vocal processes of the arytenoids, but less complete adduction in front and behind this point. The resulting glottal opening is shaped very like an arrowhead. When the elderly group was further subdivided according to age, it was clear that there was a steady increase in the incidence of vocal fold sulcus and the arrowhead configuration with advancing age. These findings strongly support reports of decreased muscle mass and strength in the aging larynx.

D. Changes in the supralaryngeal vocal tract

Pharyngeal cavity

The pharyngeal cavity at birth is very different from its adult form, but there seems to be little normative data available to show changes in pharyngeal contour and dimensions. The tongue lies entirely within the oral cavity, and does not form the anterior wall of part of

the pharynx as in adults. This is partly to allow a direct connection between the larynx and the nasopharynx during suckling. Between the second and fourth years of life, the descent of the tongue and of the larynx means that this contact is lost (Laitman and Crelin 1975:214), and the posterior third of the tongue now forms the anterior wall of the upper part of the pharynx. By the age of 9 years, the pharynx approximates to an adult configuration. One factor which will cause some change in size of the lumen of the upper pharynx is the rapid growth and regression of lymphoid tissue which forms the palatine and pharyngeal tonsils (see Section 1.2.1.B).

By adulthood, the proportions of the male and female pharynx are rather different, with the male pharynx being longer relative to the oral cavity (Fant 1966). The ratio between oral cavity and pharynx length seems to be fairly similar in women and older children. Ingerslev and Solow (1975) found that the pharyngeal angle, relative to the cranial base and face, is similar in adult males and females, but that the antero-posterior dimension is reduced in females. Measurements of a small number of xeroradiographs reported by Berry et al. (1982) confirm this sex difference. The antero-posterior measurement of the resting pharynx at the level of the epiglottis averaged 2.1 cm. in males and 1.6 cm. in females. At the level of the soft palate, the antero-posterior dimension averaged 1.4 cm. in males and 1.2 cm. in females.

Bosma (1963) summarises some of the developmental changes which occur in pharyngeal function, and which affect the maintenance of a pharyngeal airway. At birth, the pharynx actively expands and contracts during respiration and crying, but gradually stabilises, and postural changes of the head and neck develop in such a way that the pharyngeal airway is maintained.

Degenerative changes in the pharynx in old age do not seem to be much reported, but the general tendencies of muscular atrophy and hypotonia, and changes in the mucosal covering may be expected to have some influence on pharyngeal state.

Oral cavity

The configuration of the oral cavity depends on the skeletal framework (see above), the dentition and the posture and size of the tongue. The tongue is notoriously difficult to measure, which may explain the lack of comment on tongue growth and development. As mentioned earlier, the tongue is entirely contained within the oral cavity at birth, and its permanent descent into the neck is fixed by about the fourth year of life (Laitman and Crelin 1975:214). Later in childhood, descent of the hyoid bone as the neck elongates allows the tongue to descend more, and further enlarges the oral cavity (Bosma 1963:101). At birth the tongue effectively fills the oral cavity at rest, but the facial skeleton enlarges relatively more than the tongue (Bosma 1963:101), so that the oral cavity gradually enlarges. Hopkin's (1967) study of tongue dimensions suggested that the adult tongue is only twice the size of the newborn infant's, but any two dimensional representation of tongue size must be treated with some caution. The tongue grows differentially at its tip, acquiring what Bosma describes as a "limblike mobility". Eruption of teeth, enlargement of the oral cavity and maturation of chewing and swallowing patterns are all associated with a more retracted tongue posture.

Dentition

The first primary teeth usually appear at about 6 months of age, and the primary dentition is usually complete by the age of 2½ years, but there is considerable variation

in the age of eruption. The eruption of the permanent teeth is also very variable, but usually begins between 5½ and 6 years, and is complete, with the exception of the third molars, at around 12 years. The third molars, or wisdom teeth, do not normally erupt until between 18 and 21 years. Typical ages of tooth eruption are shown in Figure 1.2.4/18. The age of eruption of the permanent dentition is slightly earlier in girls, in line with the general trend towards earlier maturity in girls.

Dental age, as judged by the number of teeth which have erupted, has been used as an index of maturity (Sinclair 1978:102), although the correlation between dental age and other indices of maturity such as bone age, as measured by closure of epiphyses, is not clear.

Tooth loss through disease is a common feature of old age. The gums begin to recede from the crowns of the teeth in early adulthood, and since the enamel covering the crown of the tooth cannot regenerate, the enamel covering becomes gradually more worn from contact with hard food stuffs.

Nasal cavity

There is little data available on growth and development of the nasal cavity, but the poor development of the nasal bone at birth, and the marked enlargement of the nasal bone at puberty, together with other changes in proportions of the facial skeleton, point to major changes in the internal structure of the nose between birth and maturity. The oral-nasal port will be influenced by the lumen of the pharynx, the size and carriage of the tongue, and the mass of lymphoid tissue which is present at any given stage in development. All these factors may have major consequences for the balance of oral and nasal resonance, since the relative sizes of

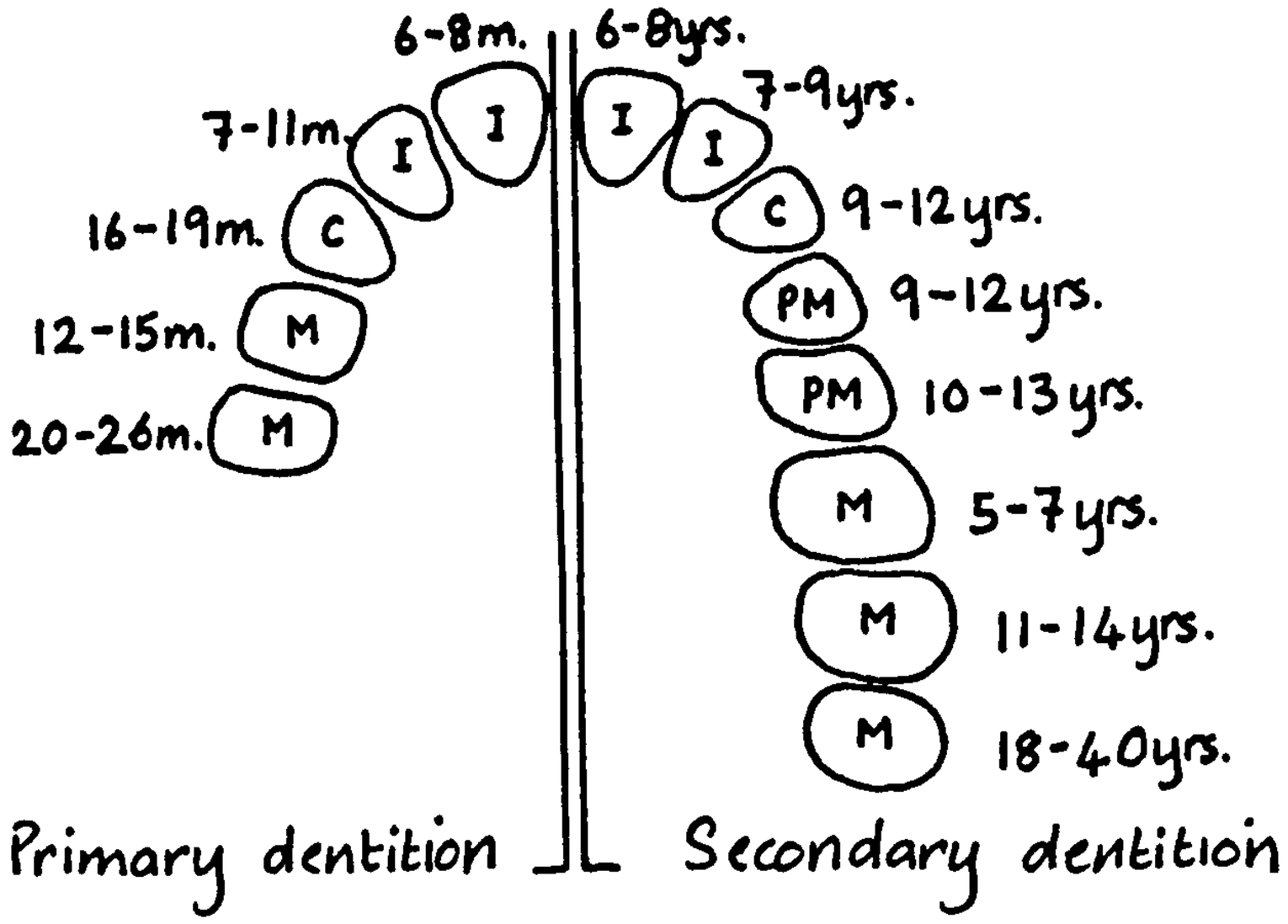
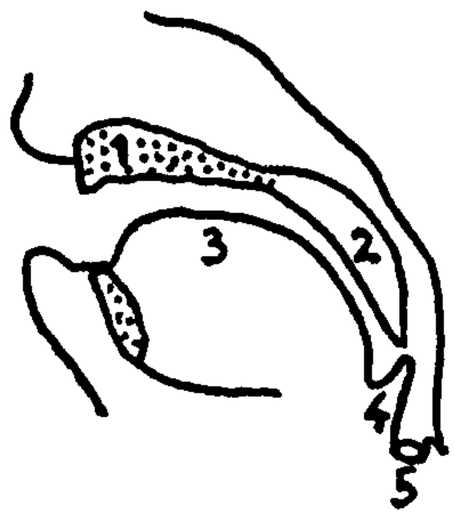
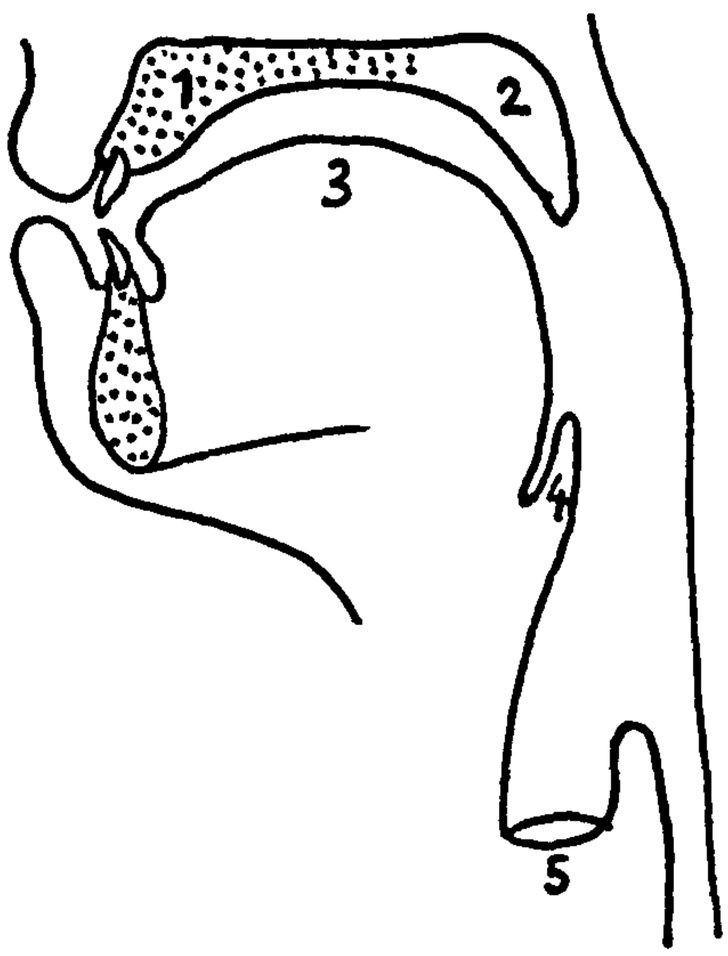


FIGURE 1.2.4/18: Typical ages of tooth eruption

A.



B.



1 = Hard Palate
2 = Soft Palate
3 = Tongue
4 = Epiglottis
5 = Larynx

Figure 1.2.4/19: A schematic diagram of the vocal tract
A. newborn
B. adult
(Adapted from Kent 1981: 163)

the posterior entrances to the oral and nasal cavities are thought to be important determinants of nasal resonance (Van Riper and Irwin 1958), but it is unfortunately hard to evaluate their effects on vocal tract configuration.

Summary of changes occurring during the three phases of life

It may be useful to summarise the overall effect which all these changes have on vocal apparatus size, shape and potential phonetic range during the childhood, adolescence and senescence.

A. Birth to puberty

It is between birth and puberty that the most obvious changes in size and configuration of the vocal tract occur. At birth, the respiratory system and the larynx are poorly developed, so that phonatory control is rather limited. The human vocal tract is similar to that of other mammals, in that the tongue is held forward within the oral cavity, the larynx lies fairly high in the neck, and the epiglottis can slide up behind the soft palate so that a continuous airway is formed between the nasopharynx and the larynx, and the infant can swallow fluid and breathe simultaneously (Laitman and Crelin 1975:206). The pharyngeal space is thus very small, and does not constitute a modifiable resonating cavity during vocalisation. The articulators in the oral region, i.e. the lips, jaw and tongue, are mobile, but immature muscular control limits their voluntary use in modifying vocalisations. The lack of teeth during the first months of life also influences articulatory potential and may have an effect on tongue posture. Figure 1.2.4/19 compares sagittal views of infant and adult vocal tracts.

The most dramatic changes occur during the first five years. After this time, the configuration of the vocal tract changes more slowly, apart from the temporary changes in dentition as permanent teeth replace the primary dentition, which may have significant, though transient, effects on front oral articulation. By the end of the first decade of life the respiratory system and the larynx are becoming more mature, and the vocal tract approximates to its adult form. Muscular development and increased neuromuscular control allow fine phonetic control of the vocal apparatus during speech.

B. Puberty to maturity

The most striking characteristic of vocal apparatus development during the adolescent years is the emergence of sexual differentiation. The most significant sex differences which are evident by early adulthood are to do with overall size of the vocal apparatus, the relative size of the larynx, and the relative proportions of the resonating cavities. Both sexes show some growth in vocal tract size during this period, and full maturation of the larynx and respiratory system will influence the range of phonation available to each individual. A rapid reduction in the mass of lymphoid tissue forming the tonsils will affect the configuration of the oropharyngeal and nasopharyngeal areas. Growth of the vocal apparatus at puberty in girls can be seen mostly as a scaling up of the pre-pubertal vocal apparatus, but in males there are significant changes in the relative proportions of the vocal apparatus. The male larynx increases rapidly and disproportionately, and the pharyngeal cavity increases its size relative to the oral cavity.

C. Maturity to senescence

General aging of the body is associated with some quite specific changes in the vocal apparatus. Respiratory function is impaired by connective tissue changes in the lungs and thoracic skeleton, and by degeneration of muscle and neuromuscular control. There are marked changes in the larynx, due to calcification of cartilages, muscular atrophy, and degenerative changes in the mucosal covering of the vocal folds. Muscular atrophy and mucosal changes will also affect the form and function of the supralaryngeal vocal tract, and the progressive loss of bone from the maxilla and mandible, together with loss of teeth, may alter the contours of the resonating cavities.

1.2.5 THE CONSEQUENCES OF GROWTH AND CHANGE FOR VOICE QUALITY

There has been relatively little objective measurement of the changes in voice features which are associated with normal developmental and regressive changes in organic structure. This is partly due to the lack, until recently, of objective assessment techniques for assessing voice features (see Section 2.1). Most reports which do exist are therefore limited either to subjective comments, which are difficult to interpret or verify, or to objective measurements of the rather limited range of voice features which are easy to measure acoustically, such as fundamental frequency (F0).

A further difficulty in the design and interpretation of studies in this field is that it is very hard to extricate the relative contributions of organic and sociolinguistic factors when comparing different age and sex groups. There is a considerable body of literature concerning stylistic and phonological differences in the verbal output of speakers of different age and sex, and these are determined by cultural factors. Reviews of such work can be found in Scherer and Giles (1979).

There have been a number of perceptual studies which suggest that listeners are able to identify both sex and age of speakers with a high level of accuracy. Taking identification of sex first, it is not surprising that the larger vocal tract and larynx in adult males makes sex identification from adult voices a relatively easy task (Schwartz and Rine 1968, Coleman 1971). Complications arise, however, when studies of sex identification from childrens' voices are considered. Some studies, at least, suggest that listeners start being able to identify the sex of children from quite an early age. Meditch (1975) found that listeners were able

to correctly identify boys as young as 3-5 years, although the sex of girls was more often guessed at. The suggested interpretation of this study is that some socially conditioned aspects of "masculine" speech are learnt very early. Certainly, there are no organic differences at this age which can explain sexual differentiation of speech. Edwards (cited by Smith 1979: 124) found very high success rates for judges identifying the sex of 10 year olds, but also found that there was an interaction between sex identification and social class of the children. Working class girls were more often incorrectly identified as boys, whilst middle class boys were more often classified as girls. Given that the organic differences between males and females are still very slight at this age, this study serves to underline the point that it is very hard to extricate organically based speech differences from sociolinguistic factors.

At least some of the studies on identification of age from adult speech have used isolated vowels or connected speech played backwards, thus aiming to eliminate most of the culturally determined stylistic differences. Ptacek and Sander (1966), for example, found that whilst correct classification of voices into young (under 35 years) or old (over 65 years) age groups was highest using connected read speech (99%), quite high levels of success were also achieved using speech played backwards (87%) or prolonged vowels (78%). It seems likely, therefore, that at least some of the features which allow sex and age identification may reflect organic differences, but the task of identifying the strands of voice quality which are influenced by specific organic changes in the vocal apparatus remains largely to be done.

Phonation characteristics

A. Subjective comments

Many authors have commented on their subjective impressions of voices at different ages, but these are very hard to interpret because of the lack of standardised descriptive systems.

At birth, vocal behaviour is very varied and poorly controlled, but during the first year or two of life the phonetic control of both the articulatory and phonatory processes become rapidly more refined (Wasz-Hockert et al. 1968, Stark et al. 1975, Stash Maskarinec et al. 1981).

Phonation changes at puberty are more obvious in boys, who are having to adjust to much greater changes in laryngeal structure. Adolescent boys are often described aa having a "husky" or "hoarse" voice quality (Curry 1949, Greene 1972:102, Aronson 1980:50), and pitch breaks and fluctuations are common. Adolescent girls may also display some "huskiness", which may be due to hormonal changes at puberty (Greene 1972:102). Huskiness is also described as a consequence of the hormonal changes which may occur during menstruation and pregnancy in adult women (Amado 1953, Tarneaud 1961, cited by Greene 1972:103).

The voice in old age has been given such labels as "weak", "tremulous", "hollow", "thin", "hoarse" and "breathy" (Greene 1972:103, Ryan and Burk 1974, Hartman and Danhauer 1976, Kahane 1978, Helfrich 1979:86), but the extent of deterioration in voice quality with age seems to be very dependent on the individual's general state of health and fitness, and on the way in which the voice has been used throughout life (Greene 1972:104).

B. Fundamental frequency mean

There have been a number of studies on F0 differences between males and females, and on F0 means at different ages (Curry 1949, Fairbanks 1942, Fairbanks et al. 1949, Hanley 1951, Linke 1953, cited in Helfrich 1979:81, Duffy 1958, Mysak 1959, McGlone and Hollien 1963, Ringel and Klungel 1964, cited in Luchsinger 1970:278, Hollien and Malcik 1962, Ostwald 1963, Hollien et al. 1965, Hollien and Copeland 1965, Michel et al. 1966, Ptacek et al. 1966, Hollien and Jackson 1967, Saxman and Burk 1967, Ostwald et al. 1968, Hollien and Paul 1969, Endres et al. 1970, Weinberg and Zlatin 1970, Hollien and Shipp 1972, Majewski et al. 1972, Montague et al. 1974, Keating and Buhr 1978, Wilcox and Horii 1980). Figure 1.2.5/1 is a graphic summary of reported average speaking F0 at different ages, showing the different sex curves. Discrepancies between studies may be due partly to different measurement procedures, and partly to sociolinguistic differences between the populations studied. Most of these studies measure mean values of F0 from samples of continuous speech, but some (e.g Keating and Buhr 1978) use median values.

At birth F0 is very high, with the mean F0 when crying being very much higher than when babbling. F0 decreases steadily up to the time of puberty, and a clear sex difference in F0 seems to emerge at some time between 7 and 10 years of age (Vuorenkoski et al. 1978, Hasek et al 1980). This difference becomes very marked at around the time of puberty, when the male voice mutates to a lower pitch, falling by as much as 100 hz over a period of a few months, whilst female voices show only a slight F0 decrease. The mutation of male voices is related to the rapid increase in size of the larynx, with a marked lengthening of the vocal folds, at this time. In adult

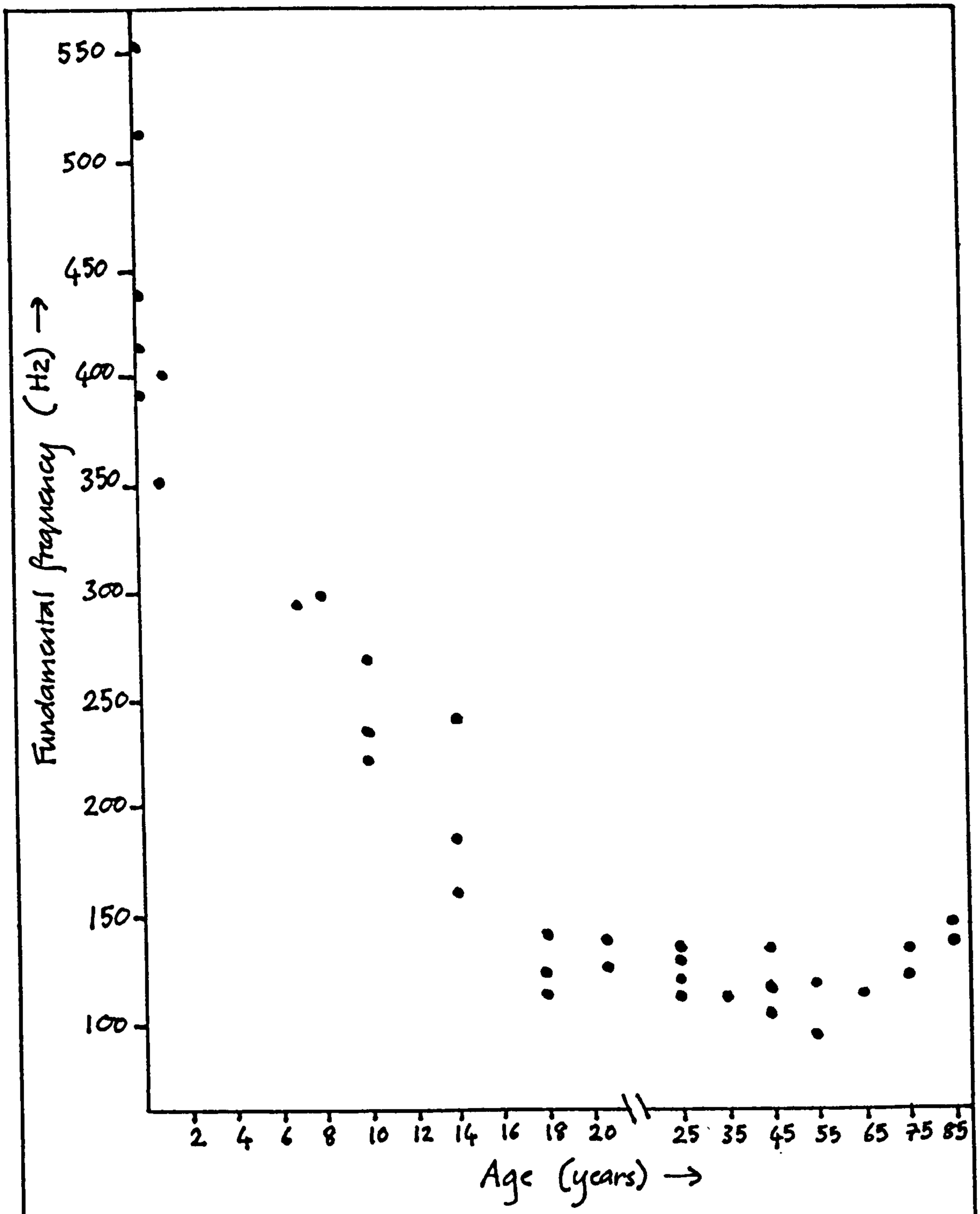


FIGURE 1.2.5/1, A: A graphic summary of variation in reported average speaking F0 of males as a function of age (adapted from Helfrich 1979:82)

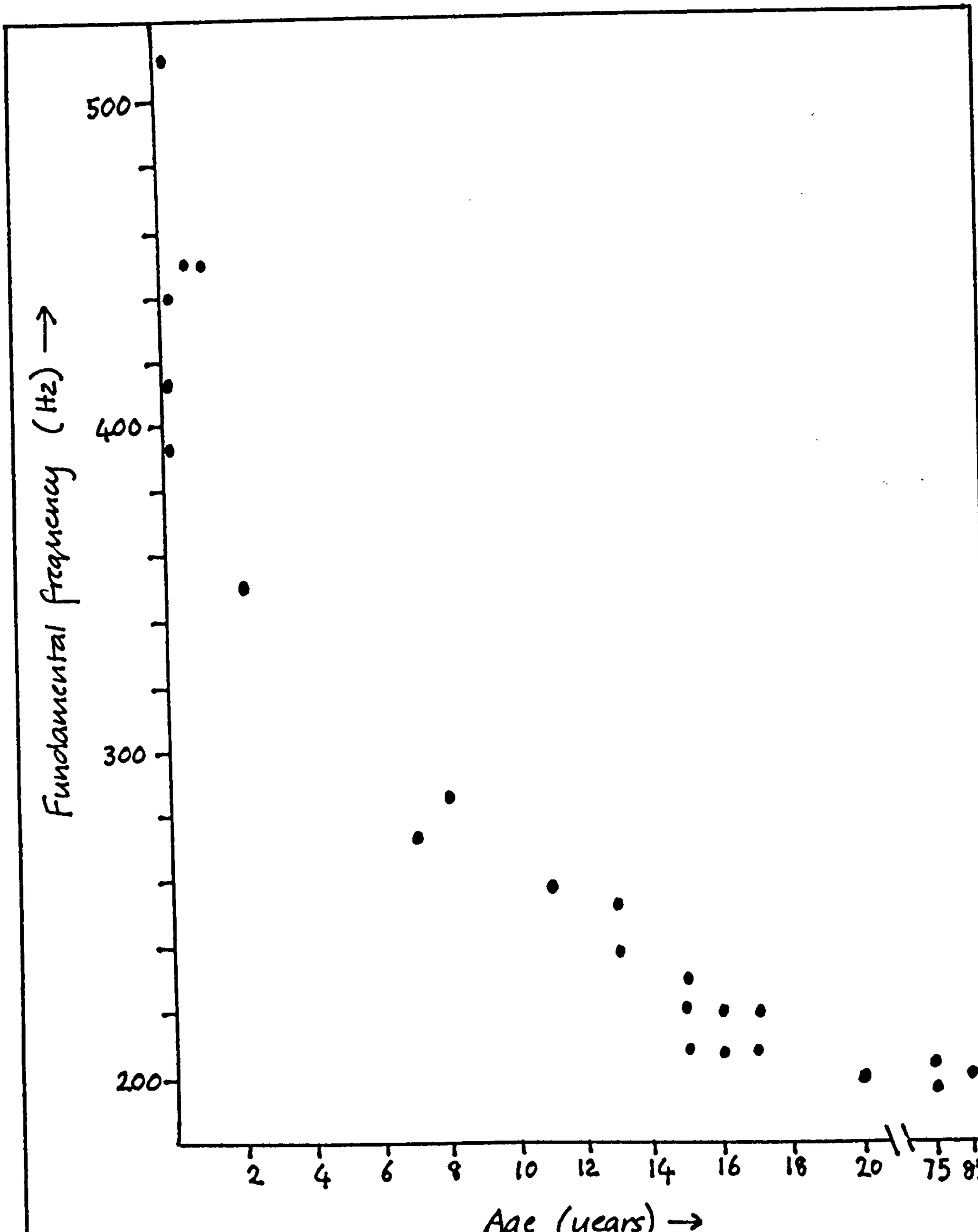


FIGURE 1.2.5/1, B: A graphic summary of reported average speaking F0 of females as a function of age (adapted from Helfrich 1979:81)

speakers, the average F0 for males is close to 100 hz, whilst in females it is closer to 200 hz.

There seems to be general agreement that old age is associated with a slight drop in F0 in females, which may be due to several factors. Mass increase of the vocal folds due to oedema, as reported by Honjo and Isshiki (1980), would certainly be expected to lower F0. A generalized loss of muscle tone, ossification of laryngeal cartilages and hormonal changes in old age may all have some effect. The relationship between F0 and age is less clear in males. The overall trend of studies reviewed in Helfrich (1979:82) was for a slight increase in F0 after the sixth decade of life. If these results reflect a real tendency for the male voice to become higher pitched in old age, then some organic or psychological reason has to be found for the fact that male voices behave so differently from females. Helfrich (1979:83) suggests that the pitch rise might be linked to a reduction in the secretion of male hormone, causing a partial reversal of the changes which result in voice mutation at puberty. The implication of such a suggestion is that the increase in laryngeal size which results from hormonal action at puberty is reversible. In fact, the observation that laryngeal changes caused in females by an excess of male hormone are not reversible, and that males who undergo a change of sex do not show any decrease in laryngeal size, makes this explanation somewhat implausible. Horii and Isshiki (1980) did find that the incidence of observable vocal fold atrophy in old age was higher in aged males than in females, and this may be a relevant factor here.

An alternative explanation for increased pitch in aged males proposed by Helfrich (1979:83) is that it is related to higher levels of psychological stress experienced by males following retirement, relative to

females whose life-style may change less if they have not been in full time employment all their lives.

There are, however, some indications that the pitch increase suggested in the literature is an artefact resulting from the collection of cross-sectional data. Helfrich (1979:81) suggests that an increase in average body size, amd hence perhaps of larygeal size may mean that the average size of the larynx is smaller in older males. Since smaller larynges tend to produce higher pitched phonation, this could explain a higher F0 in older age groups when cross sectional data is examined. Two objections to this explanation can be raised, however. Firstly, it seems surprising that the increase in mean body size, which has been observed also in females, has no equivalent effect on studies of female pitch. Secondly, there seems to be little evidence that laryngeal size is actually correlated closely with overall body size (Bristow 1980).

Further doubt about the behaviour of F0 with increasing age in males is cast by several studies which fail to show any increase in F0, or even show a decrease (Endres et al. 1971, Wilcox and Horii 1980, Benjamin 1981). Until further longitudinal studies can be carried out, it is therefore difficult to make any categorical statements about F0 in senescence.

C. Fundamental frequency range

F0 range may be measured either by measuring the distance between the highest and lowest pitches a speaker can produce, or by measuring the range habitually used during speech. Once speech is established, F0 range seems to remain constant during childhood, and to increase between adolescence and adulthood (Hartlieb 1962, Luchsinger 1970, both cited in Helfrich 1979:84). Reduced pitch

range in old age has been reported by many authors (McGlone and Hollien 1963, Saxman and Burk 1967, Endres et al. 1977), but other authors report no significant reduction (Mysak 1959, Hollien et al. 1971), or even an increase in pitch range with age (Benjamin 1981). It may be that different measurement procedures can partially explain this disagreement, but the sociolinguistic background and emotional state of speakers may also be important (Helfrich 1979:84).

D. Intensity

There seem to be few reports on speech intensity changes during childhood, although it might be expected that increased respiratory efficiency would be associated with increasing maximum intensity up until early adulthood. Similarly, intensity may be expected to fall as respiratory capacity decreases in old age, as is reported by Ptacek et al. (1966). A complicating factor affecting habitual intensity in old age may be hearing loss, which could sometimes cause speakers to use inappropriately loud voices (Helfrich 1979:86). Studies in this area should therefore be careful to draw a distinction between maximum possible intensity and habitual intensity.

E. Waveform perturbations

Pitch has been reported to be very unstable in infancy (Stark et al. 1975) and at puberty (Depons and Pommez, cited by Helfrich 1979:85), with rapidly varying F0.

Several studies have found increased pitch perturbation (jitter) in aged voices, (Sedlackova et al. 1966, Wilcox and Horii 1980, Benjamin 1981, Linville and Fisher 1985) but differences in measurement techniques do not allow easy comparison of results, and Ramig and Ringel (1983) found no significant relationship between age and jitter

measurements. They did, however, find that jitter was greater in individuals in poorer physical condition (as assessed by heart rate, blood pressure, vital capacity and percentage fat) in all age groups. This study did find a significant correlation between intensity perturbation (shimmer) and age.

Helfrich (1979:85) attributes the pitch perturbations at all ages to different varieties of lack of cortical control, but variations in the tissue layer structure of the vocal fold may also be important, since these can effect the efficient functioning of the vocal fold as a vibrating body (see Section 2.5). This may be especially important in the elderly age groups, where the histology of the vocal fold may be markedly degenerate (see Section 1.2.4).

Resonance characteristics

The dramatic changes in vocal tract size and configuration which occur in early childhood must have direct consequences for the potential range of phonetic production, but it is extremely difficult to extricate the contributions of neuromuscular maturation, language development and organic change to overall phonetic output of young children.

Some studies have attempted to relate measurement of formants during vowel production to organic differences in the vocal tract of men, women, and older children (Fant 1960, 1966). Prior to Fant's 1966 paper it was generally accepted that the formant patterns of children, women and men were related by a simple scale factor, inversely related to vocal tract length. Fant (1960) suggests that womens' formants are approximately 17% higher than mens', and childrens' are about 25% higher. In 1966 he points out that whilst this may be true if an

average is taken over all vowels, the relationship between male and female formant patterns is rather different for close front vowels, rounded back vowels, and open unrounded vowels. Children and women are still related by a simple scale factor. The reason suggested for this is that males have a longer pharynx relative to oral cavity length than do women and children.

There has been little research into changes in resonance characteristics of the vocal tract in aging, but Linville and Fisher (1985) studied women in three age bands (25-35 years, 45-55 years and 70-80 years) and report that the production of /a/ changes with age. When this vowel is fully voiced, F1 and F2 were both lower in older age groups. When the vowel is whispered, only F1 shows a lowering effect with age. The authors suggest that these differences may be explained by continuing growth of the craniofacial skeleton in adulthood, and by a lowering of the larynx in old age (Wilder 1978). This is a very limited study, but it does indicate that age related changes in the resonating cavities of the vocal tract may have detectable acoustic effects.

1.3. CONCLUSION TO PART ONE

It should be clear from this part of the thesis that the human vocal apparatus is subject to variability from many sources. During our life-span, each one of us will undergo a series of gradual changes in vocal anatomy and physiology which are the inevitable result of development and degeneration. Many processes are involved in the creation of such changes, and these processes will interact in subtly different ways so that each one of us is endowed with a unique vocal apparatus. In addition, the consequences of illness or trauma of various kinds may include alterations in the organic state of the vocal apparatus. These alterations may be transient, lasting

for a few hours or days, as in inflammation following sudden vocal misuse at a rugby match, for example, or they may be long-term or even permanent. In other words, day to day variations in vocal anatomy, in response to environmental factors and state of health, may be superimposed upon the sorts of interspeaker variation which arise from normal variability in the cycle of development and dissolution.

Since the output of the vocal instrument at any given time depends upon its form and upon its potential for phonetic adjustment, anyone concerned with speech should be aware of the kinds of inter- and intra-personal variation in the vocal apparatus which may occur. This part of the thesis should have given some indication of the range of factors which sustain variation. The second part will address the problem of determining to what extent variation in vocal features may be directly related to organic variation.

PART TWO: ORGANIC AND PHONETIC RELATIONSHIPS IN SPEECH

2.1 VOICE QUALITY ANALYSIS TECHNIQUES

This section will introduce two systems for voice quality analysis, which are complementary to one another, and which were used in the investigations described in Sections 2.2, 2.3, 2.4 and 2.5. The author was involved in the development of both systems.

2.1.1 INTRODUCTION

The first part of this thesis has, it is hoped, shown how normal and abnormal growth processes may result in individuals with widely different vocal apparatuses, and indicated that such organic variations may be reflected in speakers' habitual voice qualities. The major obstacle inhibiting adequate research in this area has been the lack of objective techniques for voice quality analysis. Voice quality analysis can take three basic approaches; it can concentrate on physiological aspects of speech production, it can concentrate on the auditory perception of speech output, or it can measure acoustic parameters of the speech wave form. All of these approaches have advantages and disadvantages, and they should be seen as complementary rather than competing strands of voice quality research.

Physiological techniques include such things as xeroradiography, cinefluoroscopy, airflow measurements, laryngography, myography and fibroptic examination of the larynx or velopharyngeal mechanism. All these techniques have yielded valuable information about the speech mechanism, and provide objective measures of speech activity, but they share certain disadvantages. The first problem stems from the intimidating effect upon the speaker of the necessary technical apparatus. This is

exacerbated in some cases by the invasiveness or discomfort of the apparatus (e.g. myography, fibroptic examination, airflow measurements), or by the need for the patient to remain in an unnatural or static position (cinefluoroscopy, xeroradiography). Most of these physiological techniques are able to give information about only a small portion of the vocal apparatus at any one time, the only exceptions being radiological techniques. Although it has been argued that the radiation dose during xeroradiography is within acceptable limits (Berry et al. 1982a), it does not seem justifiable to expose individuals to any X-ray dosage unless there are pressing medical indications. This makes it ethically unacceptable to use xeroradiography or any other radiographic measures as research tools for the investigation of normal populations. A further general difficulty in the use of physiological techniques for the assessment of voice quality is that there is not always a clear correlation between physiological activity and the auditory characteristics of speech. This is especially true of velopharyngeal activity and nasality (Laver 1980:77), and may be a problem wherever the organic state of the vocal apparatus is abnormal. A final economic disadvantage is that the cost of physiological measurement is typically high, needing both skilled personnel and expensive equipment.

Auditory perceptual techniques for the evaluation of voice quality have many potential advantages, as the ear is a highly sensitive organ for the evaluation of speech sounds. Most people are constantly alert to a wealth of information carried by the voice about social, psychological and physical factors (Laver and Trudgill 1979), although they may not normally be conscious of their skill in receiving and interpreting this information. The main problem in utilising auditory perception of voice quality in research is the lack of

objectivity and the difficulty in providing a readily understood and clearly defined terminology. These difficulties will be discussed further in the following section (Section 2.1.2), which introduces the Vocal Profile Analysis Scheme.

Acoustic measurement techniques share the advantage of objectivity with their physiological counterparts, but they also share some of the disadvantages. These include the need for expensive and often daunting equipment, which may require highly trained operators. In addition, the ability of acoustic measurement to differentiate between subtle habitual adjustments of different parts of the vocal apparatus is still in its infancy. Advances in speech technology and computing are gradually making acoustic techniques more widely useful and accessible, and the possibility of basing acoustic measurement on tape recorded samples of speech minimises disturbance of the speaker. The acoustic measurement of phonatory features will be discussed further in Section 2.1.3.

2.1.2 PERCEPTUAL ANALYSIS OF VOICE QUALITY: THE VOCAL PROFILE ANALYSIS SCHEME.

The Vocal Profile Analysis Scheme (VPAS) has its roots in a framework for the phonetic description of normal voice quality which was developed by Laver (1968, 1974, 1975) and is described in detail in Laver (1980). From 1979 to 1982 a project funded by the Medical Research Council (M.R.C. Grant No. G978/1192) was set up to further develop the scheme into a clinical assessment tool which could be used to describe the voice quality of both normal and pathological speakers. The author was employed as a full time research associate on this project, and played a substantial role in the development of the VPAS. The scheme as it is described here is largely as it existed at the end of the project.

The VPAS possesses several features which make it rather different from most other schemes for voice quality description. Firstly, the behaviour of the whole of the vocal apparatus is seen as contributing to a speaker's characteristic voice quality. The traditional approach, taken by both phonetics and speech therapy, is to consider only phonation (and, sometimes, velopharyngeal features) as "voice" features. The term "voice quality" as used in this thesis refers to this more global quality: the more traditional sense of "voice quality", meaning the quality of sound due to phonatory action, will be called "phonatory quality". Nasal features will be specifically identified as such. The VPAS highlights the interrelationships between the various parts of the vocal tract, and the ways in which habitual adjustments of any part of the vocal apparatus may colour an individual's voice.

The second important feature of the VPAS is that it analyses voice quality in terms of various potentially

independent strands, or components, which can be combined in various ways. This allows a much larger range of voices to be differentiated than is possible using holistic schemes, where a single label is used to describe overall voice quality. Holistic schemes for voice quality assessment are limited by the small number of global voice qualities which can be easily memorised. This was illustrated by an extensive study conducted by Wynter and Martin (1981). They tested the ability of judges to memorise 15 voice types, and to use these as a basis for classifying other voices. The results suggest that this is a difficult task, even with a relatively small number of voice types.

The components which the VPAS deals with are known as settings (Honikman 1964). A setting results from a long-term tendency for a speaker to impose a particular type of muscular adjustment upon the vocal apparatus. This will contribute to the characteristic voice quality of that individual. Settings can be thought of as long-term-average configurations of the vocal tract around which the short-term changes needed for articulation of phonetic segments are made. This will be discussed further below.

Settings affecting different parts of the vocal apparatus may be combined in various ways, which are characteristic of each speaker's habitual voice. It is therefore possible, in analytical terms, to build up an overall vocal profile for any speaker, which shows which settings are present, and which quantifies any deviations from a neutral baseline.

This brings us to the third crucial feature of the VPAS, which is that every setting, in every voice, is compared with a neutral baseline setting. The neutral setting is a perceptual quality that can be defined in terms of its

acoustic and physiological correlates. This gives the scheme an objective base, and allows the judge to make both qualitative and quantitative judgements about deviations from the neutral setting.

One other valuable feature of the scheme is that it is firmly rooted in general phonetic theory. The evolution of phonetic theory has led to a system for the analysis and transcription of speech sounds which can be accurately used and clearly understood by any trained phonetician. There is a large body of information about the relationships between perceptual judgements, acoustic correlates, and the physiological bases of speech production. Laver sensibly built his scheme for voice description around this framework, and one of the aims during further development of the scheme was to maintain the theoretical rigour of a phonetic approach. This gives the scheme several advantages over other perceptual voice analysis schemes. Firstly, it utilises perceptual skills which are already possessed by anyone trained in phonetics. Secondly, it allows the relationship between segmental and voice quality aspects of speech to be made explicit. Thirdly, the specification of acoustic and physiological correlates for any voice quality feature gives the VPAS a sound scientific base.

Laver's scheme for the phonetic description of normal voice quality is described in detail in Laver (1980). The following outline of the VPAS will not, therefore, describe in detail the acoustic and physiological correlates of each setting. Full details of these, together with a survey of the relevant literature, can be found in Laver's book. The aim of this section is to present and explain the main features of the VPAS in such a way as to allow an easy understanding of the following studies (see Chapters 2.2 and 2.3) which apply the technique to two populations. A description of the VPAS

will therefore be followed by a discussion of inter- and intra-judge agreement levels, and a description of the basic procedure followed when the scheme is used for an investigation of the voice quality characteristics of a group of subjects.

A "User's Manual", which is intended to act as a summary perceptual guide to users of the scheme, is also included as Appendix 1.

Basic concepts of the VPAS

There are a few basic concepts which need to be clearly understood by anyone using the VPAS. Some of these have already been mentioned, but they need some expansion. Again, the reader is referred to Laver (1980) for a fuller discussion.

Settings and segments

The scheme rests on the proposal that it is possible to perceive long term tendencies (the "settings") when listening to a stream of speech. Examples include the habitual tendency to keep the lips in a slightly spread posture throughout speech, or to keep the tongue body slightly fronted and raised towards the hard palate. Since speech is a dynamic process, involving constant movement of the vocal tract for the production of segments, the relationship between settings and segments has to be very clearly spelt out.

Settings can be seen as a second order strand of analysis, abstracted from the segmental level of analysis. For any given speaker, it may be true that so many of the segments share some common phonetic feature that it is reasonable to abstract that common feature as a long term tendency, and class it as a voice quality

feature, or setting. An example may help to make this clear. A narrow phonetic transcription of the utterance "Jane walked to the zoo ", produced by one speaker, might be:

[dj$\tilde{e}$n $\tilde{w}\tilde{ɔ}$kt t$\tilde{ə}$ $\tilde{ð}\tilde{ə}$ $\tilde{z}\tilde{u}$:]

It is clear that the nasalization diacritic recurs frequently throughout the utterance, and not only in relation to segments adjacent to the nasal consonant. It is therefore possible to abstract this tendency towards nasality and to class it as a voice quality setting.

The perceptual identification of settings is thus dependent on their relationship with segments, and the idea of "susceptibility" is very useful when using the VPAS. Most settings influence only a proportion of the segments in a speech sample. In other words, only a proportion of segments are susceptible to the effects of a given setting. Only voiced segments, for example, will be susceptible to the effects of phonation type settings. Laver (1980:20) gives two main reasons for individual segments not being susceptible to the effects of a setting. The first is that the phonetic tendency imposed by a setting is redundant. An example might be the effect of a nasal setting on nasal stops. The second reason is that the requirements of a segment may over-ride the setting. An example of this is the production of oral stops by a speaker with a nasal setting.

In fact, the second reason for lack of susceptibility may not hold in pathological speech. To quote Laver (1980:20), "Susceptibility is a scalar concept, rather than a binary one". Even segments which are phonologically required to be oral stops may be produced as nasal stops when a pathological degree of the nasal setting is present.

For most settings it is possible to specify a few of the susceptible segments which display the articulatory or perceptual effects of the setting most clearly. These can be called "key segments", and they play a useful role in the practical application of the VPAS. Examples of these will be given where appropriate, and a summary can be found in the "User's Manual" in the Appendix.

The neutral baseline

The neutral setting (henceforward referred to as neutral) is a reference setting, against which any other setting can be judged. Neutral is a convenient baseline, chosen because its acoustic and physiological correlates can be clearly specified, at least for adult males with standard vocal tracts (Laver 1980:15). It is most definitely not intended to reflect any idea of a "normal" setting, and it will, indeed, become clear in Section 2.2 that the settings used by normal speakers are usually markedly different from the specified neutral setting. Neither should it be confused with any idea of the "rest" position of the vocal tract.

A distinction actually needs to be drawn between a neutral vocal profile, in which the value of every single setting is neutral, and the neutral value of a given setting. A speaker with a neutral vocal profile is a rarity. In examining the profiles of over 200 speakers in the MRC project, we failed to register a single speaker who showed a neutral value on every single setting analysed. Context will normally make it clear whether the term "neutral" is being used to refer to the complete profile of a speaker's voice, or to the neutral value of a single setting.

"Neutral", therefore, may be used to describe a composite setting, for which the situation at different points

along the vocal tract can be specified. The following description assumes standard vocal tract anatomy:

- the lips are not protruded
- the larynx is neither raised nor lowered
- the supralaryngeal vocal tract is as nearly as possible of equal cross-section along its whole length
- Segments which are phonologically described as alveolars are produced at an alveolar place of articulation
- the body of the tongue is neither raised nor lowered, and neither fronted nor backed
- the faucal pillars do not constrict the vocal tract
- the pharyngeal constrictor muscles do not constrict the vocal tract
- the jaw is neither unduly close nor unduly open
- there is audible nasality only when necessary for linguistic purposes, and there is no audible nasal escape of air
- the vibration of the true vocal folds is regularily periodic, efficient in air use and without audible friction. The full length of the vocal folds are involved in vibration, and there is balanced, moderate muscular tension (Laver and Hanson 1981).

The configuration of the vocal tract in the neutral setting corresponds to the position assumed for the production of the central vowel [ə]. In this posture, the vocal tract is as close as is humanly possible to a bent tube of equal cross section along its entire length. It is this simple physical shape which allows a straightforward prediction of the acoustic correlates of the neutral setting (see below). Figure 2.1.2/1 shows a sagittal section of a speaker with his vocal tract in the neutral configuration.

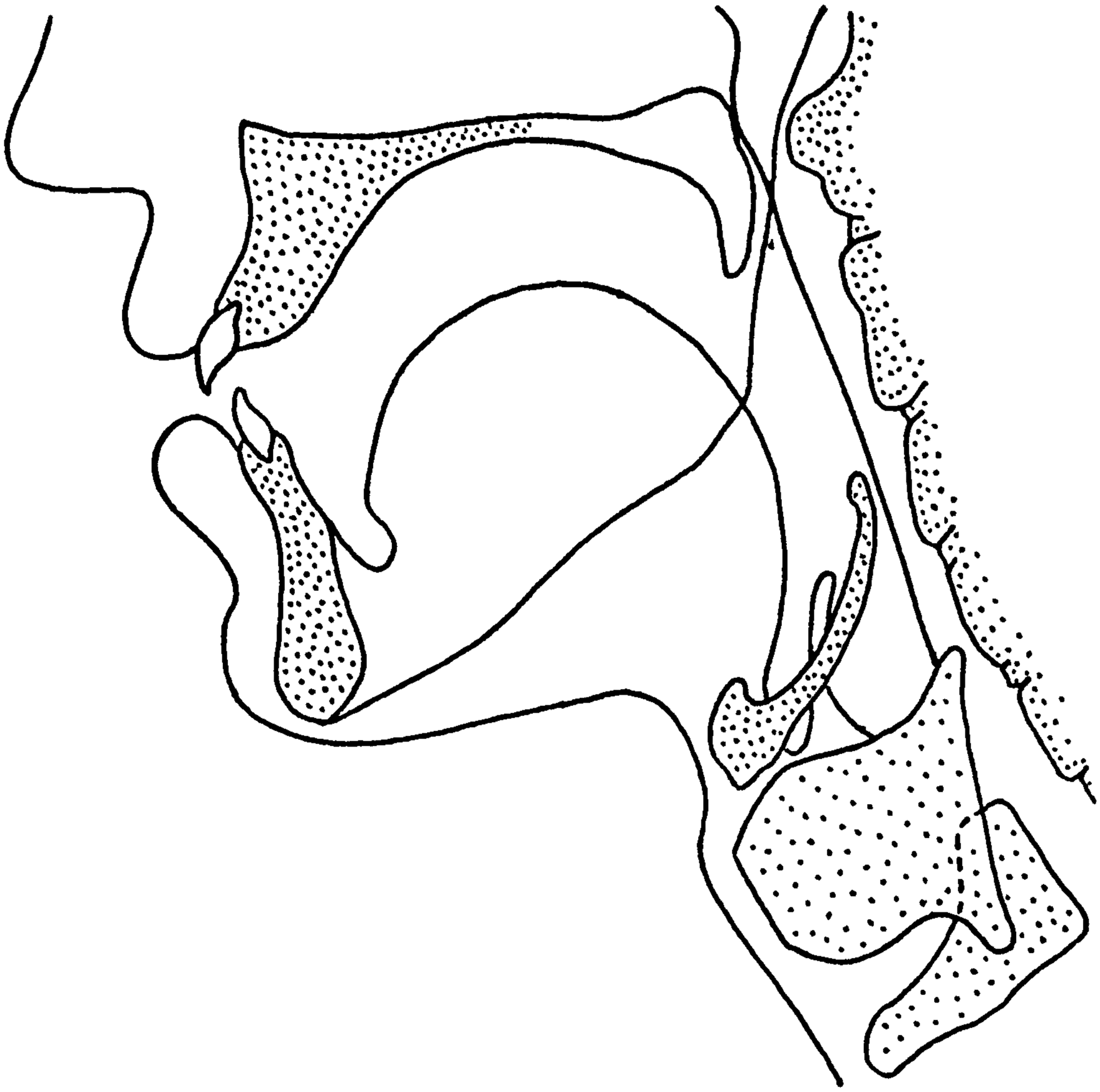


FIGURE 2.1.2/1: Radiographic diagram of the vocal tract in a neutral setting (redrawn from Laver 1980:24)

It should be noted that in some respects this definition differs slightly from the definition given in Laver 1980:14. This is because some adjustments of the original scheme were necessary when it was more widely applied to pathological populations.

The acoustic correlates of this neutral setting are easy to specify for normal adult male vocal tracts, since it is such individuals who have provided the data base for most of the earlier studies in acoustic phonetics. The following summary of the acoustic characteristics of neutral assumes an adult male speaker with a vocal tract which is 17 cm. long and of normal proportions:

- the average value of the first formant is about 500 Hz, and higher formants are odd multiples of this, giving a ratio of 1:3:5 etc., i.e. the second formant is 1500 Hz, the third is 2500 Hz and so on. The first three formants have bandwidths of 100 Hz.
- fundamental frequency is in the 60-240 Hz range.
- larynx pulses* show an approximately triangular waveform, which is regular in frequency and amplitude, with maximum excitation during the closing phase of the glottal cycle. The closing phase occupies about one third of the glottal cycle.
- spectral slope of the glottal waveform is between -10 dB and -12 dB per octave (Laver and Hanson 1981).

It is obvious that details of this acoustic description must be adjusted for females, for children and for males with vocal tracts which are of greater or smaller length. At the perceptual level, however, there seems to be no difficulty in applying the neutral reference quality, except perhaps in the case of very young children where both anatomical and phonological systems are very immature.

(* As seen on laryngographic recordings)

Non-neutral settings

Settings may deviate from neutral in four main ways. They may affect the length of the vocal tract, they may affect the cross section of the vocal tract, they may affect the frequency of occurrence of audible nasal resonance, or they may affect the mode of phonation used. Laver describes these classes of non-neutral settings as longitudinal, latitudinal, velopharyngeal and phonatory settings, respectively. Examples of longitudinal settings include lip protrusion and lowered larynx, both of which will increase the length of the vocal tract. Latitudinal settings include raised tongue body, which will constrict the oral cavity, and pharyngeal constriction, which reduces the cross-sectional area of the pharynx. Velopharyngeal settings may differ from neutral either by having audible nasal resonance on more segments than those which are phonologically described as nasals, or by having a reduction in nasal resonance heard on these "nasals". Phonation may differ from neutral either in the mode of the laryngeal vibration, or by the addition of audible, fricative airflow through the glottis.

A voice may also differ from neutral as a result of the overall levels of muscular tension which exist throughout the vocal apparatus. Altering the overall tension level tends to produce a constellation of changes, which could be described at the local level, but it is often useful to abstract the common underlying tension feature and to describe it as a vocal quality setting in its own right.

One final class of setting appears in the VPAS, but was not considered in Laver's original work. This covers the range of articulatory movement habitually used by the lips, jaw and tongue. When pathological speech is the focus of attention, it rapidly becomes obvious that the extent of articulatory movement is just as much a

characteristic vocal feature as is the long term average position of the vocal organs. Having once introduced this new class of settings, it became clear that differences in habitual articulatory range may also be important in characterising normal speakers.

The presence of a non-neutral setting may be due to either of two reasons. The speaker may be making a phonetic adjustment of the vocal apparatus, which is potentially under voluntary control, or she may be blessed with a vocal anatomy which makes the use of a non-neutral setting unavoidable. Two simple examples may serve to illustrate this. The habitual use of a whispery phonation, where the vocal folds do not adduct fully and there is fricative airflow through the glottis throughout phonation, is a frequently occurring feature in British English speakers. Exaggerated levels of whisperiness may occur either as a result of personal phonatory style, due to phonetic adjustment of the larynx, or as the inevitable consequence of organic abnormalities such as polyps, which protrude into the glottis and prevent full vocal fold adduction. Similarly, the habitual use of a protruded jaw setting may be due either to a muscular adjustment, pushing the mandible forward relative to the maxilla, or it may be due to the possession of a mandible which is disproportionately large so that the speaker cannot do other than hold the mandible in a protruded position.

Combination of settings

A recurring theme in any description of the VPAS must be the high level of physiological interdependence between settings. The complex interlinkage of the muscles which control the vocal apparatus makes it highly probable that the presence of a setting affecting one part of the vocal tract will influence the settings which may be used

elsewhere in the vocal tract. For a speaker with a normal vocal apparatus, it is possible to specify some pairs of settings which are mutually incompatible. For example, harsh phonation normally requires high levels of muscular tension, and will not therefore co-exist with a lax laryngeal tension setting. Falsetto and neutral phonation demand entirely different physiological actions of the larynx, and so cannot occur in combination.

There are other instances where the presence of one setting has an "enabling" effect on other settings (Laver 1980:18). For example, protruding the jaw may facilitate the use of a fronted tongue tip/blade setting, by carrying the tongue forward relative to the maxillary dental arch.

Exceptions to these guidelines must, however, be expected in cases where organic abnormality is suspected. For example, a lax laryngeal tension setting may be found in combination with harsh phonation in individuals with laryngeal pathology. This will be discussed in more detail in Section 2.5.

Acoustic and auditory interaction also have to be taken into account when constructing and interpreting vocal profiles. There is a certain amount of acoustic interaction between the larynx and the supralaryngeal vocal tract, so that laryngeal vibration may be affected by the configuration of the pharyngeal, oral and nasal cavities. Such effects are probably negligible, except to the degree that there is coupling between the nasal tract and the rest of the vocal tract (Stevens and House 1961, Laver 1980:18). More important for the functioning of the VPAS is the problem of auditory interaction between settings. In some cases, settings may share very similar auditory characteristices, leading to the possibility of confusion. For example, the fricative airflow from the

nose which is the principal feature of audible nasal escape is easily confused with the fricative airflow through the glottis which is found in whispery phonation. Auditory masking may also be a problem, with some settings being perceptually less prominent when other settings are present. Nasality, for example, is much harder to hear in the presence of whisperiness and some other settings.

The principle of auditory equivalence

In spite of earlier suggestions that non-neutral settings may derive either from phonetic or organic causes, it is very easy to assume that there is an invariable link between the perceived voice quality setting and the physiological or phonetic adjustments which are used to produce it. This impression is strengthened by a survey of the labels used to describe the settings (see Figure 2.1.2/2), which are mostly a direct reflection of the phonetic adjustments which a normal speaker uses to produce each voice quality type. This fits with the tendency within general phonetic theory to assume that all speakers perform more or less the same muscular actions to produce sounds which are perceived as being the same. Whilst such an assumption may stand up reasonably well for speakers with more or less standard vocal tract anatomy and physiology, it has to be questioned when organic quirks of the vocal tract are found. Since one of the aims during development of the VPAS was to widen its applicability to pathological populations, within which the incidence of organic abnormality is relatively high, it rapidly became neccessary to consider the strength of the link between perceived voice quality and phonetic adjustments in cases with organic anomalies. If the presence of organic abnormality weakens or dissolves the link, then further

Vocal Profile Analysis Protocol

Speaker: ..... **Sex:** ..... **Age:** ..... **Date of Analysis:** ..... **Tape:** ..... **Judge:** .....

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS: Neutral	FIRST PASS: Non-neutral Normal	FIRST PASS: Non-neutral Abnormal	SECOND PASS: SETTING	Scalar Degrees: Normal 1	2	3	Abnormal 4	5	6
A. Supralaryngeal Features											
1. Labial				Lip Rounding/Protrusion							
				Lip Spreading							
				Labiodentalization							
				Extensive Range							
				Minimised Range							
2. Mandibular				Close Jaw							
				Open Jaw							
				Protruded Jaw							
				Extensive Range							
				Minimised Range							
3. Lingual Tip/Blade				Advanced							
				Retracted							
4. Lingual Body				Fronted Body							
				Backed Body							
				Raised Body							
				Lowered Body							
				Extensive Range							
				Minimised Range							
5. Velopharyngeal				Nasal							
				Audible Nasal Escape							
				Denasal							
6. Pharyngeal				Pharyngeal Constriction							
7. Supralaryngeal Tension				Tense							
				Lax							
B. Laryngeal Features											
8. Laryngeal Tension				Tense							
				Lax							
9. Larynx Position				Raised							
				Lowered							
10. Phonation Type				Harshness							
				Whisper(y)							
				Breathiness							
				Creak(y)							
				Falsetto							
				Modal Voice							

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

II PROSODIC FEATURES

CATEGORY	FIRST PASS: Neutral	FIRST PASS: Non-neutral Normal	FIRST PASS: Non-neutral Abnormal	SECOND PASS: SETTING	Scalar Degrees: Normal 1	2	3	Abnormal 4	5	6
1. Pitch				High Mean							
				Low Mean							
				Wide Range							
				Narrow Range							
				High Variability							
				Low Variability							
2. Consistency				Tremor							
3. Loudness				High Mean							
				Low Mean							
				Wide Range							
				Narrow Range							
				High Variability							
				Low Variability							

III TEMPORAL ORGANIZATION FEATURES

CATEGORY	FIRST PASS: Adequate	FIRST PASS: Inadequate	SECOND PASS: Scalar Degrees Inadequate 1	2	3	
1. Continuity						Interrupted
2. Rate						Fast
						Slow

IV COMMENTS

	Adequate	Inadequate	1	2	3
Breath Support					
Rhythmicality					

	Present	Absent
Diplophonia		

Other Comments:

FIGURE 2.1.2/2: Vocal Profile Analysis Protocol

questions may be asked about the validity of using the VPAS in such speakers.

The task of empirically testing the strength of the link, and of discovering the extent and type of organic anomaly required to disturb it, would need a major project in its own right. There are, however, logical reasons for stating that in many, if not most, cases the output of abnormal vocal tracts can perfectly well be analysed using the VPAS, but that the usual assumptions about phonetic bases do indeed need to be treated with great caution.

The acoustic and auditory characteristics of a speaker's voice depend on the configuration and movement of the vocal organs rather than on the underlying muscular activities. As long as the configuration of the vocal apparatus in an organically abnormal speaker is similar to one which is possible for a normal speaker to produce, then the output can be judged using the existing scheme. In other words, the principle of auditory equivalence can be summarised by saying that as long as the settings produced by an organically abnormal vocal apparatus sound the same as settings which a normal speaker can produce, then the VPAS can be applied as a tool of practical and convenient description.

Since the VPAS analyses the various components of voice quality separately, it does not even matter if the combination of settings used by a speaker is one which a normal vocal apparatus could not produce.

Some examples of cases where the VPAS has been successfully used to describe the output of speakers with abnormal vocal anatomy will be found in Section 2.3, but a simple example here may help to clarify the principle. For a normal speaker to produce a raised and fronted

tongue body setting, which constricts the vocal tract in the palatal area, there must be a muscular adjustment which actually pushes the body of the tongue forwards and upwards. It is the constriction of the vocal tract which is actually responsible for the auditory quality, however, and a speaker who has an abnormally small palatal volume because of a narrow, low palatal arch may produce an auditorily identical effect. This is attributable for practical purposes to a palatalised tongue body setting, but such a speaker may produce it without making an equivalent muscular adjustment.

Procedure for completing a Vocal Profile Analysis

When faced with the task of analysing an individual's vocal profile, the judge must first decide what speech material to base the analysis upon. Ideally, the analysis should be based on both a face-to-face interview and on a tape recorded sample of speech. As with segmental phonetic analysis, visual cues may be valuable in confirming auditory impressions, but it is possible to complete the analysis without seeing the speaker. Tape-recording is, however, essential, as it is not often feasible to attempt full Vocal Profile analysis in a live interview. Recordings must be of reasonable quality, since some settings are particularily prone to distortion by common recording faults. Tape hiss, for example, may mask or mimic whisperiness, and loss of high frequency energy mimics one acoustic effect of increased nasality.

Choice of speech sample (reading, spontaneous speech etc.) will vary according to the aims of the analysis, but in all cases the sample should be of reasonable length. It is not practicable to abstract long-term-average supralaryngeal tendencies from a sample of much less than 40 seconds, although some features, such as phonation type, may be analysed from shorter samples.

The VPA protocol form

The VPA protocol form shown in Figure 2.1.2/2 is the end point of a long evolutionary chain, shaped by the combined forces of theoretical requirements, clinical needs and graphical constraints. A kind of natural selection has resulted in the exclusion of certain settings which have proved to be of little adaptive value for the function of the VPAS as a clinical tool, and the inclusion of others which were not included in Laver's original descriptive scheme. Development of the protocol form was seen as a crucial factor in making the scheme maximally efficient in both clinical and research contexts. The form needed to reflect the underlying phonetic theory in as much detail as possible, without being so unwieldy as to be unusable.

A general description of the form and of the procedure for completing a vocal profile analysis will be followed by a more detailed explanation of the individual settings.

The form is divided into three main sections, allowing comment on vocal quality features, prosodic features and temporal organization features. The theoretical basis for the first section, on vocal quality, is by far the firmest, so this part will be used to illustrate the way in which the protocol form works.

The vocal quality section is itself subdivided into two parts; a supralaryngeal section, which is concerned with the state of the vocal tract above the larynx, and a laryngeal section which is to do with the configuration, position and performance of the phonatory system. It must be stressed that this division is merely a matter of convenience, and that it is somewhat artificial. It would be misleading to suggest that there is any real

physiological or phonetic separation between the larynx and the supralaryngeal vocal tract. The interlinking of the muscle systems throughout the vocal tract, which causes a high degree of interdependence between the muscles controlling the larynx and those affecting the rest of the vocal tract, has already been mentioned. This means that phonetic adjustments of the larynx are likely to have repercussions elsewhere in the vocal tract, and vice versa. For example, raising the larynx is often associated with pharyngeal constriction because of the way in which the larynx is suspended from the hyoid system (Laver 1980:24-27).

The graphical separation between laryngeal and supralaryngeal settings is retained largely as a result of pressure from speech therapists who were involved in development of the scheme. The consensus was that it is useful to retain the distinction because, in spite of the interdependence between the two sections, it is true that many pathological speakers show deviations from normal which cluster mainly in one section or the other. The separation of the two sections on the form therefore allows an instant evaluation of the extent to which any patient presents as a laryngeal or a supralaryngeal disorder. Too casual an acceptance of the separation is, however, dangerous, because it can easily perpetuate the tendency for clinicians to forget the links between different parts of the vocal apparatus and fall back into the way of treating 'laryngeal' disorders as totally apart from supralaryngeal or 'articulatory' disorders.

The basic layout of the form allows a two stage process of evaluation, at two levels of subtlety. On the left hand side of the form are listed the major categories within which muscular adjustments away from the neutral position may occur (labial, mandibular, lingual etc.). To the right of the category labels the form is divided

vertically into two sections headed 'First Pass' and 'Second Pass'. This division is a response to the experience that it is often an easier perceptual task to judge that a given voice deviates from neutral than it is to specify the exact nature of that deviation. It seems to be true, for example, that people learning the scheme find it relatively easy to discern an adjustment of the larynx away from neutral, but find it considerably more difficult to differentiate between the qualities associated with raising and lowering of the larynx. This is in spite of the fact that the acoustic correlates of raised and lowered larynx are markedly different. The 'First Pass', or first listening, therefore requires only a rather crude decision between neutral and non-neutral for each category.

Under 'Second Pass' are listed the specific settings within each category, and the judge is here required to specify not only the precise nature of the deviation away from neutral, but also the degree of each deviation.

There are six scalar degrees of deviation from neutral for most settings, and the form designates scalar degrees 1-3 as normal and 4-5 as abnormal. Scalar degree 1 for any setting is the minimum deviation from neutral which can be auditorily identified. Scalar degree 6 corresponds to the maximum deviation which a normal vocal apparatus is capable of producing. The remaining scalar degrees are intended to reflect, as far as possible, equal auditory steps between these extremes. The meaning of the terms 'normal' and 'abnormal', and the reasons for using them on the protocol form, need some expansion. Firstly, there are some things which the labels do not mean. There is certainly no information to suggest that 'normal' relates to statistical norms, and any suggestion that the presence or absence of settings at scalar degree 4 or above is indicative of overall vocal abnormality is

highly dubious. It is not true that a speaker whose vocal profile shows one or two settings within the 'abnormal' range is necessarily pathological, or even dramatically unusual. Similarly, it is not true that the vocal profile of a speaker with a grossly pathological voice will inevitably have many settings within the abnormal range. The interpretation of a vocal profile as normal or abnormal will depend on an examination of the co-occurrence of settings within the whole profile, and on a knowledge of what non-neutral settings are characteristic of a given speech community. In some speech communities it is not uncommon to find one or two settings within the 'abnormal' range. Many American accents, for example, are typically nasal at scalar degree 4. This underlines the point that neutral is definitely not synonymous with normality.

Having said all that, there are some points which favour the retention of the normal/abnormal labels. It is true to say that, for most settings, scalar degree 3 is the maximum deviation which is frequently characteristic of specific accents. Exceptions to this rule, like the case of nasality in American accents mentioned above, are relatively uncommon. As a result, non-clinical phoneticians, who are unfamiliar with the wide range of voice types which pass through speech therapy clinics, may be tempted to let their judgements drift towards higher scalar degrees than is appropriate. The dividing line emphasised on the form between scalar degrees 3 and 4 may help to check this tendency.

Scalar degree conventions

For most vocal quality settings it is possible to specify the phonetic characteristics which determine precisely the choice of scalar degree (see later in this section),

but it may be useful to offer some general guidelines for understanding the meaning of the scalar degrees.

- Scalar degree 1 is used when the presence of a setting is just noticeable.
- Scalar degree 2 suggests that the judge is fairly confident about the presence of a setting, but that there is only moderate deviation from neutral.
- Scalar degree 3 can be taken as the strongest degree of a setting which could reasonably be expected to act as a regional or sociolinguistic marker for a hypothetical community, although there are exceptions to this rule.
- Scalar degree 4 indicates that there is no doubt at all about the presence of a setting, and that it is beyond the limits of widespread use amongst accents marking membership of a sociolinguistic community.
- Scalar degree 5 represents almost the maximum strength of deviation of which the normal vocal apparatus is capable.
- Scalar degree 6 is reserved for the auditory effect which corresponds to the most extreme adjustment of which the normal, non-pathological vocal apparatus is capable.

Since this definition of scalar degree 6 is limited by the potential of a normal vocal apparatus, the possibility exists that a speaker with organic abnormality may produce a higher degree of some setting. Scalar degree 6 may therefore be seen as an open-ended category which includes any level of a setting which exceeds that which a normal vocal apparatus could produce. In practice, it has been found that the auditory qualities associated with even grossly abnormal vocal anatomy seldom exceed the potential output of organically normal speakers.

Intermittency

It is normally adequate simply to tick the appropriate scalar degree box to indicate that a setting is more or less continuously present throughout a speech sample. Many speakers, however, are characterized by the regular, but intermittent, adoption of a setting. A useful scoring convention in these cases is to use the letter 'i' in the appropriate scalar degree box, to indicate intermittent presence of the setting. The scalar degree used should reflect the strength of the setting when it is present, rather than the frequency of occurrence. As a general rule, 'i' is used whenever a setting is heard on less than about 90% but more than 10% of the susceptible segments. Where a judge feels that it is important to indicate the proportion of susceptible segments which are affected, a percentage may be written alongside the scalar degree judgement. In a clinical context this is often useful in monitoring the progress of some dysphonic cases, for example, where the aim of therapy is to reduce the incidence of intermittent harshness associated with peaks of laryngeal tension.

A perceptual guide to voice quality settings

It may be useful to preface this section with a few general guidelines about approaches to listening. The skills required are similar to those used in segmental phonetics, but the emphasis is somewhat different. In segmental analysis much of the emphasis is placed on isolation of features which distinguish each segment from its neighbours. In Vocal Profile analysis the task is instead to identify those features which are common to all, or to some sizeable subset, of the segments in a sample of speech. The analysis of a particular setting is often a two stage process, using two rather different perceptual strategies. The first involves the abstraction of any long-term-average biasing which underlies the rapid movements required for segmental production. This means cultivating the ability to ignore the linguistic message, and to concentrate on the overall phonetic impression. This strategy is most useful in the initial identification of a setting.

Confirmation of the presence of a setting, and assignment of a scalar degree often demands the detailed analysis of classes of segments. This requires the auditory ability to isolate segments from the stream of continuous speech, and hold them in memory long enough to analyse their perceptual characteristics.

The concept of susceptibility, which has already been outlined, is very pertinent here. The second stage of analysis is obviously much simpler if it is known that only a small subset of segments are susceptible to the effects of the setting in question. Phonation type settings, for example, will affect only those segments which are phonologically voiced. Voiceless segments will not be susceptible, and can therefore be ignored. Similarly, a spread lip setting will have a major effect

on segments which are normally expected to be rounded, such as /u/, whilst segments such as /i/, which are normally spread anyway, will be much less susceptible to its effects.

Within the group of susceptible segments for a given setting, it is often possible to identify a smaller set of segments on which the auditory effect of the setting is especially prominent. These "key segments" allow an economical listening strategy, since once the presence of a particular setting is suspected, the judge can test her initial impressions by concentrating primarily on the key segments.

The following descriptions of individual settings will include comments on susceptibility and key segments wherever appropriate, since an analysis of the precise phonetic identity of key segments is often crucial in assigning a scalar degree to a setting.

An underlying assumption of the following descriptions is that the native language of the speakers whose voices are being analysed is English. The general principles of the scheme are universal, and apply to all languages, but the phonological details discussed below are specific to English.

I. Vocal Quality Features

Although almost all of the vocal quality settings have six scalar degrees, it is useful to distinguish between settings which can actually be thought of as seven point scales, where neutral acts as the first point on the scale, and those which can be thought of as making up thirteen point scales, where neutral forms the mid-point between two diametrically opposed setting types. Examples of seven point scales include labiodentalization and

protruded jaw, whilst examples of thirteen point scales include fronted / backed tongue body and nasal / denasal resonance.

I.A.1) Supralaryngeal features - configurational settings

1. Labial features

The neutral setting for the labial category is where the long-term-average lip posture is as it would be for the production of the vowel [ə]. The lips are neither spread, nor rounded, nor protruded.

Lip posture may differ from neutral in various ways, and in three dimensions. Laver differentiates 17 types of non-neutral labial setting (1980:31,33,35), by specifying the following features: vertical expansion or constriction of the labial aperture, horizontal expansion or constriction and labial protrusion. The range of possible lip settings is further extended when combinations of these parameters with labiodentalization are considered. In practice, such detail makes the analysis procedure very unwieldy, so that only the commonest non-neutral settings are included in the VPAS. These are lip rounding with protrusion, lip spreading and labiodentalization.

When judging labial settings, a useful first step is simply to visualise the "set" of the speaker's face, copy it, and then imitate a few phrases for auditory comparison. This kind of non-analytical approach is often surprisingly accurate, and these first impressions can then be checked using the information below.

Lip rounding/protrusion

Lip rounding and protrusion are physiologically separable, but since lip rounding most commonly occurs with a comparable amount of protrusion, and vice versa, they have been collapsed into a single setting. In the rare instances where there is a major discrepancy between the degree of rounding and protrusion, then it is a simple matter to delete the part of the setting label which does not apply from the protocol form.

Key segments

Lip rounding/protrusion is most prominent on the following segments:

- front oral segments [s] and [θ] have a lower apparent "pitch" than when not rounded/protruded.
- /i/ and other vowels which are conventionally spread or unrounded will tend to become more rounded. The actual phonetic realization of /i/ in a word like 'heed', in a speaker with a lip rounded and protruded setting, will tend to be rather rounded, and closer to [y] than to [i].
- /r/, /ʃ/, /t͡ʃ/ and /d͡ʒ/, where lip rounding is optional in English, will tend to be produced with lip rounding.

Scalar degrees

Scalar degrees 1-3 are used for long-term-average (LTA) lip positions of open rounding, and scalar degrees 4-6 are used for close rounding. Scalar degree 3 is where the LTA lip position is equivalent to that used for Cardinal Vowel 6 [ɔ]. In scalar degrees 4-6 the labial aperture becomes progressively smaller, until scalar degree 6 has a LTA lip position comparable to that used to produce Cardinal Vowel 8 [u].

Lip spreading

Lip spreading involves horizontal expansion of the labial aperture, as in a smiling expression. Most judges seem to find this very easy to perceive, which may reflect the emphasis our culture places on smiling.

Key segments
- Front oral fricatives [s] and [θ] have a higher apparent "pitch" in lip spreading.
- /r/, /ʃ/, /tʃ/, and /dʒ/ tend to be pronounced without lip rounding. This is most easily heard in the transitions to and from these segments.
- /w/ and vowels which are normally produced with lip rounding, such as /u/ and /ɔ/, will tend to lose their lip rounding and even become spread.

Scalar degrees
Scalar degree 4 is used to mark the point where the LTA lip position is as spread as it would be for Cardinal Vowel 2 [e]. Scalar degree 6 corresponds to the position for an overspread [i].

Lip rounding/protrusion and lip spreading can be thought of as diametrically opposed deviations from neutral. Together they form a 13 point scale, with neutral forming the central point. Although lip protrusion affects the length of the vocal tract, the focus of attention is on the cross sectional area of the labial opening. The next setting is rather different.

Labiodentalization

This setting is produced by bringing the lower lip closer to the upper teeth, thus shortening the vocal tract. Labiodentalization may co-exist with either lip rounding or lip spreading, and indeed many people produce some

degree of labiodentalization with the kind of short term use of a spread lip setting that results from talking whilst smiling or laughing. Approximation of the upper lip and the lower teeth may produce a similar auditory quality.

Key segments

- Bilabial stops /p/, /b/ and /m/ are most susceptible to the effects of labiodentalization. There may be audible labiodentalization at onset and offset of these segments, or they may actually be produced as labiodental stops.
- Front oral fricatives, especially [s], may have a lower apparent "pitch". This is a rather variable feature, however, because of the possible interaction with lip rounding or spreading.
- /r/, /w/ and /u/ often have audible labiodentalization.

Scalar degrees

Scalar degrees 1-3 add an audible labiodental factor to the onset and offset of /p/, /b/ and /m/. In scalar degrees 4-6 there is a progressive increase in the realization of these segments as labiodental stops, so that by scalar degree 6 they are all produced as fully labiodental stops.

2. Mandibular Features

In the neutral setting for the mandibular category there is a very small vertical gap between the upper and lower incisors for most speakers. In the horizontal plane, the lower incisors lie just inside the upper ones.

Open and close jaw

The long term average position of the jaw may be more open or more close than the neutral position. In the VPA protocol used for the studies in this thesis, open and close jaw were treated as a 13 point scale, with neutral as the mid point. It can, however, be argued that the

physical and auditory distance between neutral and scalar degree 6 of open jaw is much greater than the distance between neutral and scalar degree 6 of close jaw. For this reason, future adaptations of the protocol form might have only 3 scalar degrees for close jaw. These would correspond to a collapsing of scalar degrees 1 and 2, scalar degrees 3 and 4, and scalar degrees 5 and 6.

Key segments

The amount of jaw opening used by a speaker may have rather general effects, since in the absence of any compensatory adjustments it will have consequences for labial opening and for the carriage of the tongue relative to the roof of the mouth. The amount of "travel" heard during the articulation of front consonants and close front vowels is often a useful clue.

Scalar degrees

Scalar degree 1 of close jaw corresponds to the position in which there is no longer any vertical gap between the upper and lower incisors. Scalar degree 6 corresponds to totally clenched teeth. For open jaw, scalar degree 4 marks the jaw position which just allows the upper surface of the tongue to be clearly visible. Scalar degree 6 is the maximum possible opening achievable with normal anatomy.

Protruded jaw

Protruded jaw is associated with a change in the horizontal relationship between the upper and lower incisors, and between the tongue and the roof of the mouth.

Key segments

- /s/ and /ʃ/ have a 'darker', low-pitched quality, which becomes very obvious at scalar degrees of 4 or more.

- Since the protruded jaw carries the tongue forward relative to the upper teeth and the palate, all lingual articulations will tend to be fronted unless compensatory adjustments of the tongue are made. Where compensatory adjustments are made, a slightly retroflex quality is often heard on front oral consonants.

Scalar Degrees

In scalar degree 4 the lower incisors are held just in front of the upper incisors. In scalar degree 6, the lower teeth are level with the upper lip, as long as the lip itself is not protruded.

3. Lingual Tip/blade settings

The first category of lingual settings is specifically concerned with the actual place of articulation of the set of segments which are conventionally described as 'alveolars', i.e. /t, d, s, z, n, l/. The articulatory activity of the tip/blade area of the tongue is to some extent independent of the body of the tongue, which will be dealt with in the next category. This is shown by the fact that it is perfectly possible to produce dental or interdental stops whilst keeping the tongue body back so as to produce a secondary velarized or even pharyngealized articulation. There is, however, a strong tendency for lingual tip/blade and lingual body settings to be closely associated, and it is obviously more common to find an advanced tip/blade setting combined with a fronted tongue body setting than with a backed tongue body setting.

In a neutral tip/blade setting all the so-called 'alveolar' segments are produced with a truly alveolar place of articulation. The active articulator may be either the tip or the blade of the tongue. This is slightly different from the definition of neutral in

Laver (1980:48), which specifies that the blade must be the active articulator for alveolar consonants, and then contrasts this with two possible non-neutral settings; tip and retroflex articulation. It was thought that the tip/blade distinction was not really relevant for clinical assessment, but that the precise place of articulation was important.

Advanced and retracted tip/blade

It is possible to produce the above set of 'alveolar' segments with a place of articulation which is either in front of the alveolar ridge (advanced) or behind the alveolar ridge (retracted). It is usual in speakers of English for retraction to be associated with increasing degrees of retroflection, so that extreme degrees of retraction involve retroflex articulation of the so-called alveolar segments. This extreme degree of retracted tip/blade setting is almost invariably associated with rotation of the tongue body towards a backed and lowered setting, which enables the retroflection of the tongue tip.

Key segments

All the susceptible segments, i.e. /t,d,s,z,n,l/, should be used as key segments. The effect of advanced or retracted tongue tip/blade is often most prominent on /s/, but the judge must check that any deviation from the alveolar position in /s/ production is generalized throughout the whole set of segments. It is not uncommon for an accent, or an individual, to be characterized by non-alveolar pronounciation of only one of the set, often /s/. In this case it is more appropriate to view this deviation from neutral as a segmental characteristic than as a vocal quality characteristic.

Scalar degrees

For an advanced tongue setting, scalar degree 1 is the point where the tongue tip or blade begins to make contact with the back surface of the teeth as well as with the front of the alveolar ridge. Scalar degree 4 corresponds to fully dental articulation, with no alveolar contact. Scalar degree 6, being the maximum possible for normal speakers, corresponds to extreme interdentalization.

In retracted settings, the place of articulation moves progressively back, so that scalar degree 3 involves a post-alveolar place of articulation. In scalar degree 4 the tongue tip is beginning to move towards a retroflex position, with the tongue tip pointing directly up just behind the post-alveolar place of articulation. In this degree of retraction /s/ often has a very distinctive 'whistling' quality. Scalar degree 6 has the underside of the tongue tip making contact with the roof of the mouth in fully retroflex articulation.

4. Lingual body settings

The second category of lingual settings is concerned with the LTA position of the central mass of the tongue. In the neutral setting, the tongue body lies fairly centrally, vertically below the junction of the hard and soft palates (see Figure 2.1.2/1). From the neutral position, the long term articulatory tendency of the tongue body may move up or down, and backwards or forwards. Several listening strategies may be employed, of which two are most useful. The first is to try to abstract a LTA vowel quality from the continuous stream of speech. If this can be done, it follows that the LTA tongue position must correspond to the position needed to produce the abstracted vowel. For example, in the neutral setting, the abstracted vowel should be [ə]-like. If the

abstracted vowel quality is [i]-like, then the LTA tongue body setting must be fronted and raised. If it is [ɑ]-like, then the tongue body setting must be backed and lowered. A second technique is to concentrate on specific vowel segments, and to judge where they fall in a traditional vowel area diagram. In a neutral setting the vowels will be evenly distributed around the centre of the vowel area, but in non-neutral settings the distribution will be skewed away from the centre (see Figure 2.1.2/3). On the protocol form there are two pairs of diametrically opposed setting scales; fronted/backed and raised/lowered, but in practice tongue body settings are often combinations of these, such as fronted + raised, or backed + lowered.

Fronted/backed tongue body

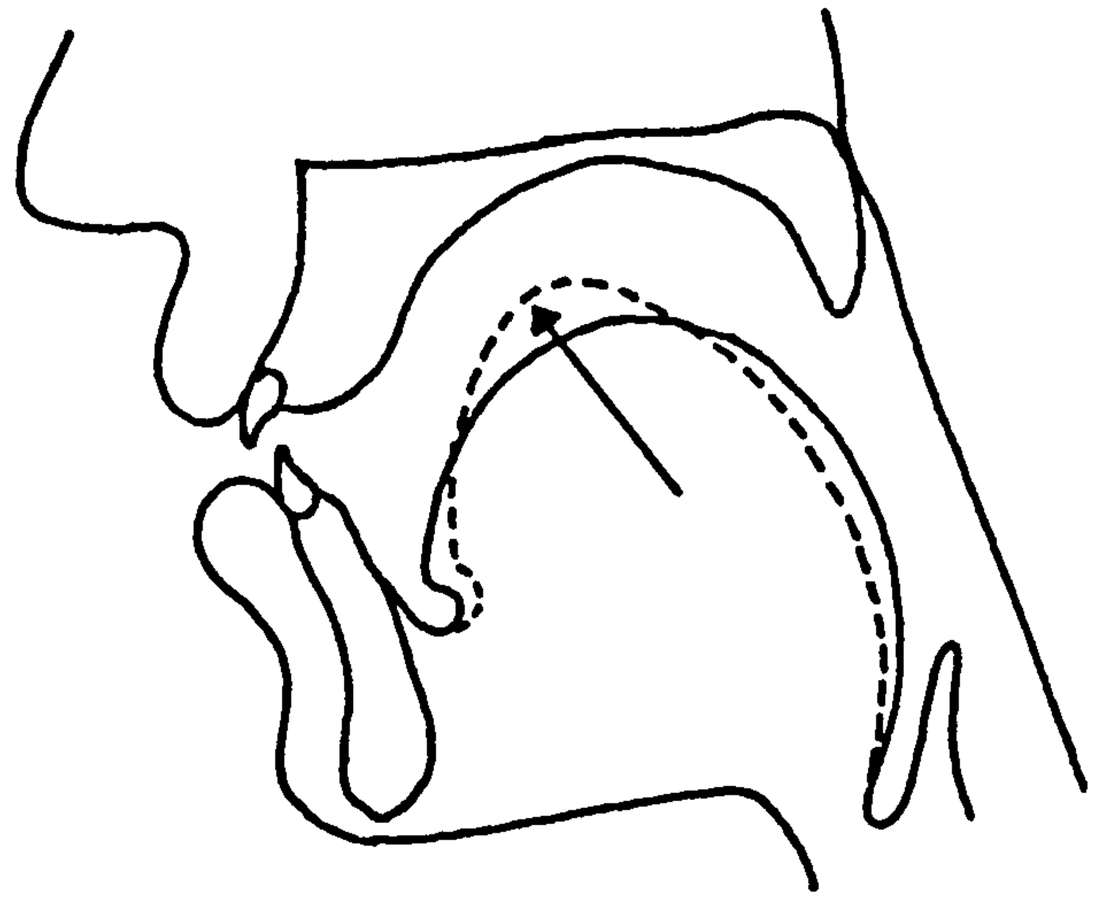
Key segments

- Vowels are the segments most susceptible to change by tongue body settings. In fronted tongue body, back vowels will be most affected, becoming progressively more fronted, so that in extreme degrees of fronted tongue body there will be no vowels in the right hand half of the vowel area. Tongue backing, in contrast, affects front vowels most, pushing all vowels backwards, towards the right of the vowel area.
- /l/ and /w/ may vary in terms of secondary articulation. Palatalization is likely to be most pronounced in speakers with fronted tongue body, whilst velarization or pharyngealization are likely to be more pronounced in speakers with backed tongue body.

Scalar degrees

Assignment of scalar degree depends on a judgement of how far the vowel area is limited to left or right (front or back). Scalar degree 4 of fronted tongue body brings the furthest back vowels forward to a central position. /u/,

A.



B.

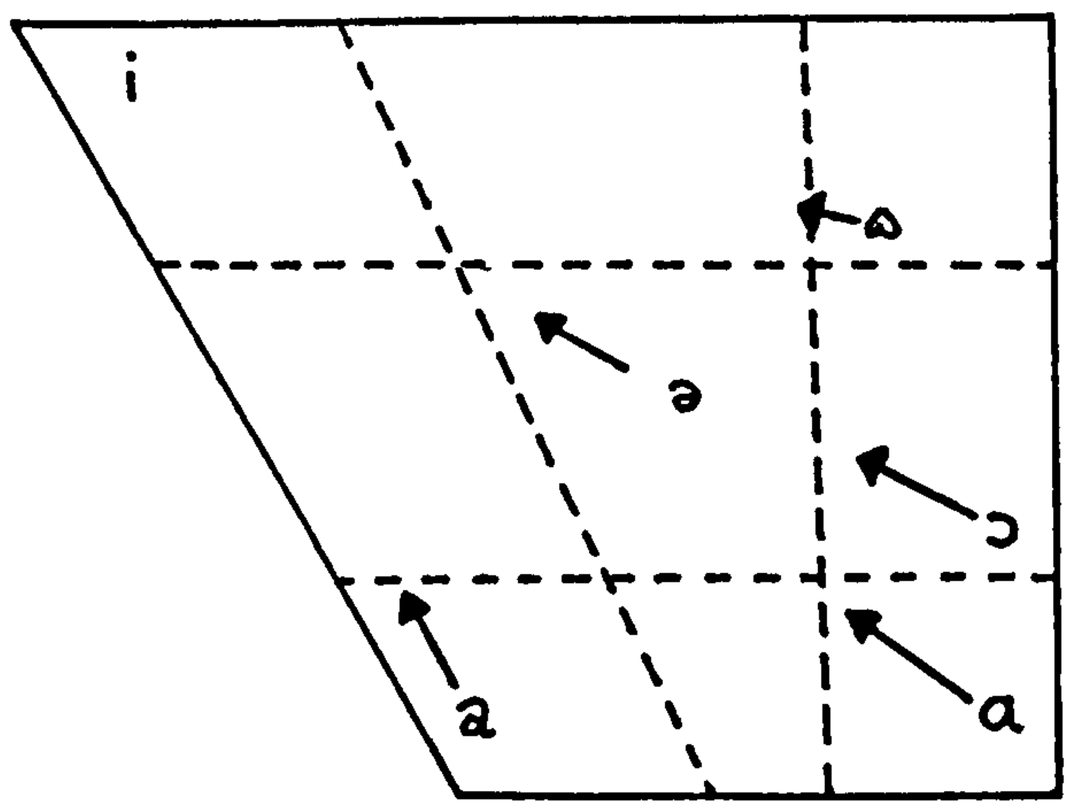


FIGURE 2.1.2/3: Diagram of changes in A. vocal tract configuration and B. vowel distribution in neutral (solid line) and fronted and raised tongue body setting (broken line)

for example, would tend to be realized as a close central vowel. In a backed tongue body setting, scalar degree 4 shifts all vowels back, so that the 'frontest' vowels are in the centre of the vowel area. /i/ would in this case be realized as a close central vowel.

Raised/lowered tongue body

The principles of judging these settings are the same as for fronted and backed tongue body. Raised tongue body makes all vowels closer, and lowered tongue body makes all vowels more open. Tongue body lowering will also affect semi-vowels /j/ and /w/, so that they may be realized as half-close variants.

Scalar degrees

Scalar degree 4 of raised tongue body will bring the most open vowels up to a borderline position between half-close and half-open. Scalar degree 4 of lowered tongue body will bring the closest vowels down to a similar position. In scalar degree 4 and beyond, /j/ and /w/ will become half-close.

5. Velopharyngeal features

Velopharyngeal settings pose some of the most complex problems for phonetics. They are complex both at the physiological level, since nasal resonance is clearly not solely a matter of whether or not the nasal cavity is coupled to the oral cavity, and at the acoustic and auditory levels. A review of some of the controversial views and findings about velopharyngeal features of voice can be found in Laver (1980:68-92). This scheme forces a simple decision between nasal and denasal resonance, but we recognise that this two way distinction may not always allow an adequate description of velopharyngeal features. Given the inconclusive nature of the findings of many

projects which have concentrated only on velopharyngeal features, it was beyond the scope of this project to provide a fully satisfactory answer to the problem. For most normal and pathological speakers the descriptive system given below allows judges to agree on the velopharyngeal setting heard.

Neutral

The neutral velopharyngeal setting is where audible nasality is present only where it is necessary to maintain phonological identity. For English that means that only /m/, /n/ and /ŋ/ will have audible nasality, and anticipatory nasality will be cut to the minimum which is physiologically necessary. In practice, neutral is virtually never heard in English, because anticipatory nasalization of vowels in pre-nasal consonant position is typically of greater duration than the physiological minimum.

Nasal

Key segments

Vowels and continuant consonants may be heard to have nasal resonance. Nasality is heard most easily on open vowels, but close vowels and eventually some consonants (e.g. voiced fricatives) will have audible nasality at higher scalar degrees.

Scalar degrees

Up to scalar degree 2 nasality will be easily heard only on open vowels. At scalar degree 3 some closer vowels will show audible nasality. By scalar degree 4 all vowels will have clearly audible nasality. Nasality begins to affect consonants at scalar degree 5, increasing at scalar degree 6 so that nasality will be clearly heard on voiced fricatives, for example.

Denasal

Key segments

- /m/, /n/ and /ŋ/ progressively lose nasal resonance.
- Vowels have a 'cold-in-the-head' quality.

Scalar degrees

In scalar degrees 1-3 the most prominent feature is the 'cold-in-the-head' effect on some vowels. In scalar degree 4 the so-called nasal stops will be clearly losing nasality. At scalar degree 6, they will have lost all nasality. The distinction between /m/, /n/ and /ŋ/ and their oral counterparts will be maintained only by having different amounts of voicing, so that severe problems of intelligibility may arise.

Audible nasal escape

Audible nasal escape is audible, fricative airflow from the nose. This is a setting which is not described in Laver (1980). Since it is considered to be abnormal in all accents of English, and presumably this is also the case for all the languages of the world, the protocol shows only scalar degrees 4-6. Audible nasal escape will tend to occur first on segments which require the maintenance of high oral air pressure, e.g. /s/, /f/. At scalar degree 4 only these segments will have fricative nasal airflow, whilst at grade 6 it will be present on virtually all segments. It should be stressed that whilst audible nasal escape occurs most commonly with high degrees of nasal resonance, this is not an invariable association. In rare instances it may even occur with a denasal setting.

6. Pharyngeal Constriction

This setting is used to describe constriction of the pharynx which results not from retraction of the body of the tongue into the pharynx, but from sphincteric contraction of the pharyngeal constrictor muscles. It lends a 'strangulated' quality to the voice, so that at high scalar degrees the empathetic listener is aware of considerable discomfort and obstruction of the pharynx.

Key segments and scalar degrees

Pharyngeal constriction is most clearly audible on vowels. The main quality shares auditory characteristics with what is normally thought of as pharyngealization, but without the tongue body or root being necessarily backed. In speakers with normal vocal anatomy, pharyngeal constriction is always the result of excessive muscle tension, and this is associated with an additional "hard" or "tinny" quality, resulting from the fact that there is little absorption of acoustic energy by the pharyngeal walls. It is difficult to specify segmentally based guidelines for assignment of scalar degrees, so the general scalar degree conventions are used.

I.A.ii) Supralaryngeal settings -articulatory range settings

Articulatory range settings specify the maximum span of movement which lips, jaw and tongue cover during speech. This should not be confused with rate of articulatory movement, although there is an obvious interaction between the two. It is, however possible to have a wide overall range of, say, jaw movement, but for the rate of jaw movement to nevertheless be rather slow.

Key segments

- Diphthongs: these will show a long travel from start to end point in extensive range settings, and very little or no travel in minimised range settings.

Scalar degrees

For ranges of lips, jaw and tongue, the end points of the scales are easily defined. Scalar degree 6 of extensive range means that the articulator must reach the most extreme positions of which it is capable, in all directions. Scalar degree 6 of minimised range means that the articulator is totally immobile. Neutral refers to the range of movement which will maintain clear intelligibility without the need for some other articulator to compensate.

I.A.iii) Supralaryngeal features - overall tension settings

Alterations in overall tension of the vocal tract tend to cause constellations of changes in configurational and range settings. Judgement of overall tension is therefore based largely on a knowlege of these constellations, which are outlined below. Problems may, however, arise in cases where physiological anomalies mean that a change in tension is not associated with the usual changes in other settings. In these cases, the listener may have to rely on an empathetic judgement of muscular tension.

Lax

Generalised laxness is often associated with the following local changes:

- Open jaw setting
- Nasal setting
- Minimised ranges of lip, jaw and tongue movement.

In addition, acoustic clues to laxness, which presumably contribute to its auditory characteristics, include damping of high frequency energy, and broad formant peaks.

Tense

Generalised tension is associated with a different set of local changes:

- Reduced degrees of nasality
- Extensive ranges of lip, jaw and tongue movement
- Pharyngeal constriction.

Acoustically, there is less absorption of high frequencies by the vocal tract walls, and formant peaks are sharper.

The general scalar degree conventions are used when scoring tension settings.

I.B.1) Laryngeal features - configurational settings

9. Larynx position

The potential range of larynx positions is quite wide, as evidenced by the displacement of the larynx which occurs during swallowing. The complex of muscles from which the larynx is slung means that alterations in larynx position may be accompanied by a wide range of other changes, and this sometimes makes it difficult to isolate the auditory effect of larynx position settings. The judge needs to concentrate on the auditory effects of lengthening or shortening the vocal tract, and try to dissociate these from features such as changes in pitch or pharyngeal constriction, which often, but not always, accompany changes in larynx position.

Neutral corresponds to the auditory quality associated with a larynx position approximately in the mid-point of its potential range.

Raised and lowered larynx

The effects of larynx position settings are most clearly audible on vowels, as a result of changes in formant values associated with vocal tract length. It is not possible to give specific guidelines for scalar degrees, so the general conventions should be followed.

10. Phonation type settings

Perfectly neutral phonation is seldom heard in normal continuous speech, but it has very clear acoustic and physiological correlates. Neutral phonation, or to give it its alternative label, modal voice, involves very regular and efficient vocal fold vibration. Only the true vocal folds are involved in phonation, and the pattern of vibration is perfectly regular; each cycle of vibration has the same duration and magnitude as its neighbours. Acoustically, it is possible to see this regularity in terms of fundamental frequency and intensity.

Phonation may deviate from neutral either by the addition of audible turbulence of the airflow, or by an alteration in the pattern of vocal fold vibration. Laver distinguishes between three classes of muscular tension which are relevant in discriminating between the physiological bases of phonation types, at least for organically normal larynges. These are shown diagramatically in Figure 2.1.2/4. Longitudinal tension is due to activity of the vocalis and/or the cricothyroid muscles, adductive tension is due to activity of the interarytenoid muscles, and medial compression is due to activity of the lateral cricoarytenoid muscles and the

lateral parts of the thyroarytenoid muscles. The importance of these muscles in controlling phonation is described more fully in Laver (1980:93ff.), but a summary of the muscle tensions and their consequences for laryngeal configuration will be included in the descriptions of non-neutral phonation types given below. Figure 2.1.2/4 tabulates the relative amounts of tension for the three tension parameters for each phonation type, and Figure 2.1.2/5 shows the associated laryngeal configurations diagrammatically.

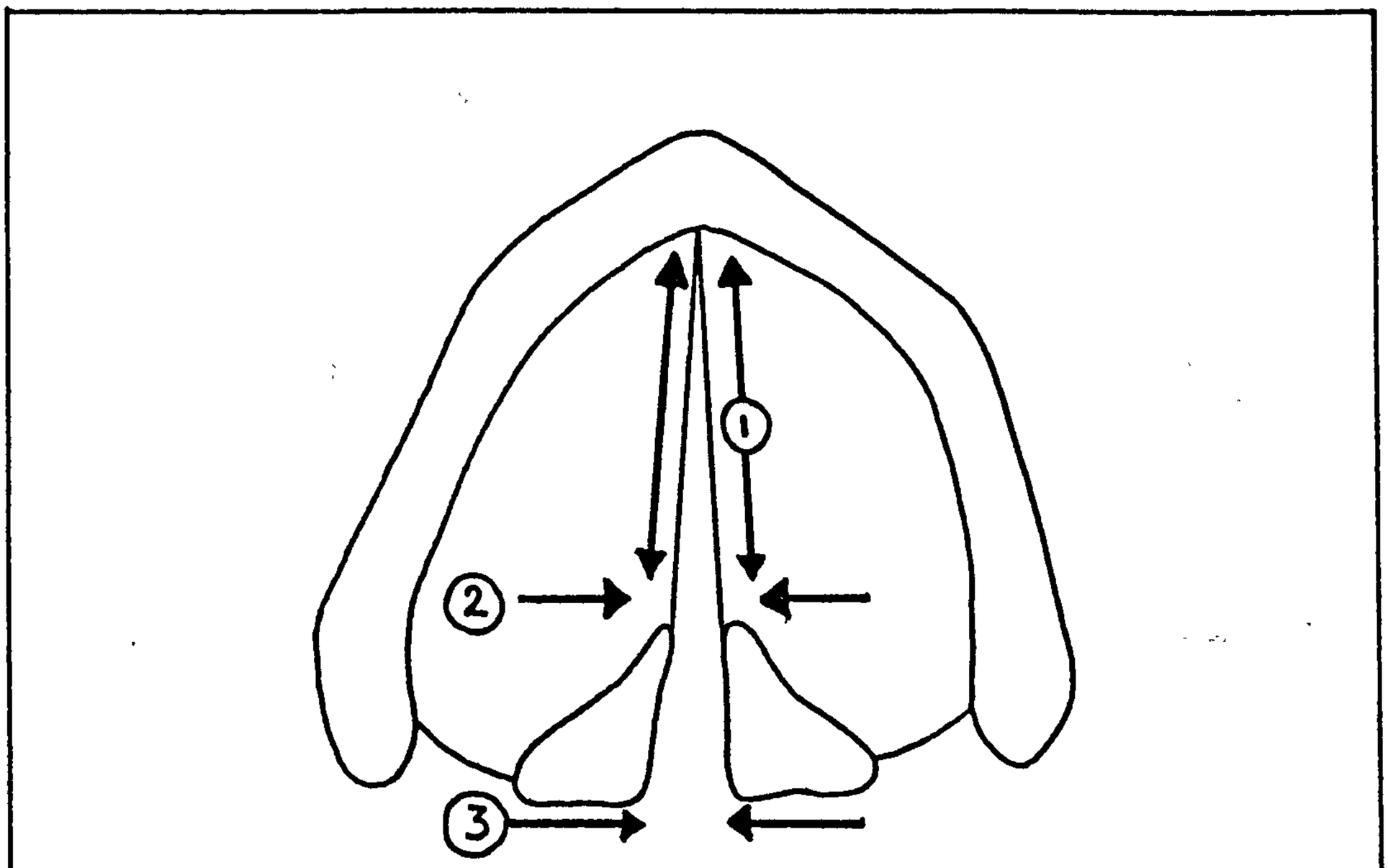
Scalar degree conventions in non-neutral phonation

Modal voice is marked simply as being present, intermittently present or absent on the protocol form. Where it occurs as a component of complex phonation types, the auditory balance between it and other components is indicated by the scalar degrees assigned to the other components. Where any phonation type is combined with voice, scalar degrees 1-3 are used where the voice component is perceptually more prominent and scalar degrees 4-6 are used if the other phonation type is perceptually more prominent. A similar rule applies when falsetto is combined with other phonation types (see below).

When modal voice occurs in combination with other phonation types in non-neutral phonation, the term 'voice' is used to describe this component. For example, the combination of modal voice with whisperiness is known as whispery voice.

Falsetto

Falsetto cannot occur at the same time as modal voice, although it can be combined with all other phonation types. This is because modal voice and falsetto require



PHONATION TYPE	LONGITUDINAL TENSION (1)	MEDIAL COMPRESSION (2)	ADDUCTIVE TENSION (3)
Neutral (modal voice)	moderate	moderate	moderate
Falsetto	high (passive)	? high	? high
Harshness	high	very high	very high
Whisper(iness)	moderate	moderate or high	low
Creak(iness)	low	high	high

FIGURE 2.1.2/4: A summary of laryngeal tension parameters in different phonation settings

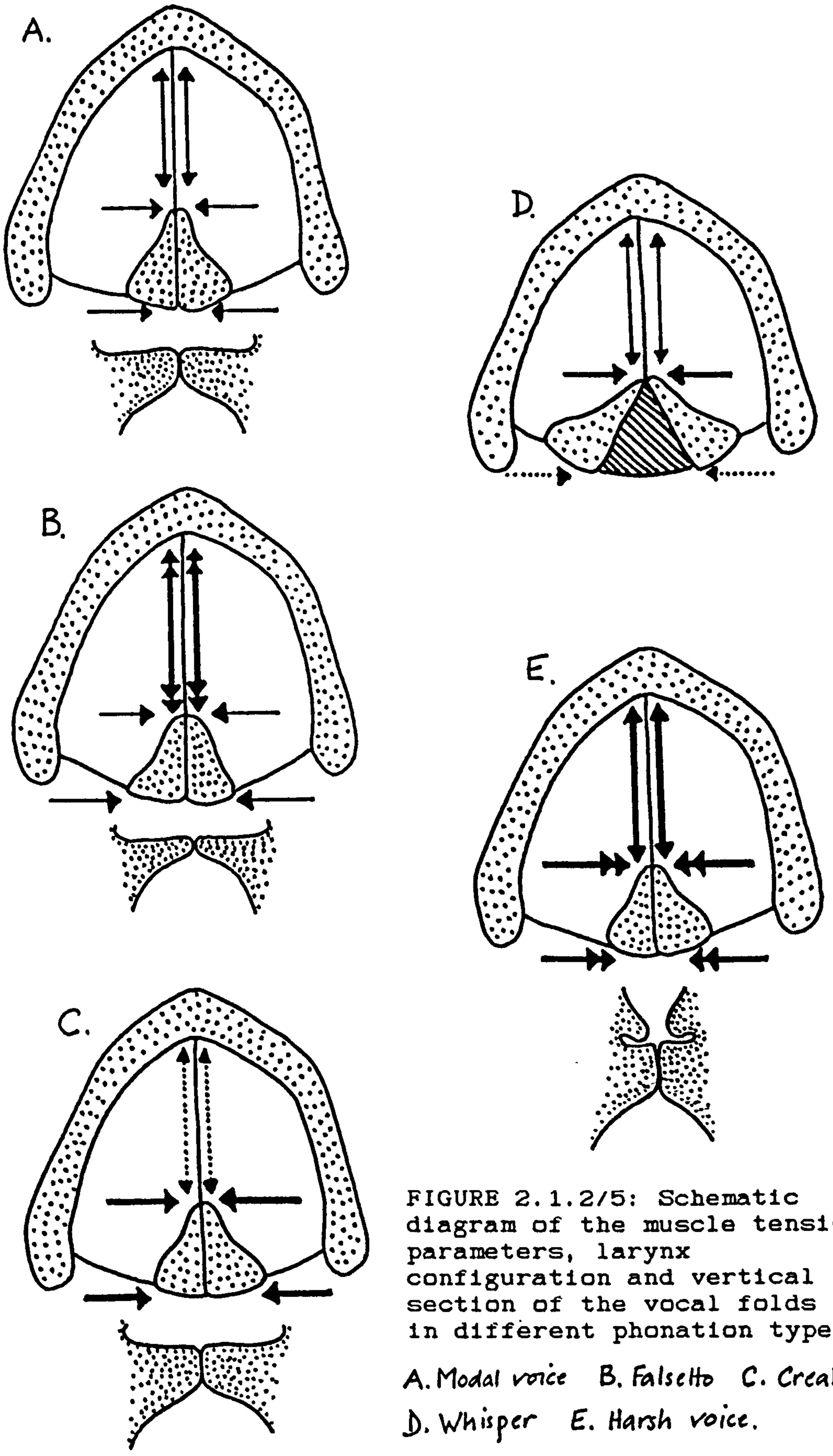


FIGURE 2.1.2/5: Schematic diagram of the muscle tension parameters, larynx configuration and vertical section of the vocal folds in different phonation types

A. Modal voice B. Falsetto C. Creak

D. Whisper E. Harsh voice.

mutually exclusive patterns of muscle activity (Hollien 1971:329, Laver 1980:118). In falsetto, there is a high level of longitudinal tension, but it is due to contraction of the cricothyroid muscles rather than of the vocalis muscle itself. This results in the vertical cross section of the vocal folds becoming rather thin. Adductive tension and medial compression are also thought to be high relative to modal voice (Van den Berg 1968:298), although this is not always specified in other descriptions (Laver 1980:118).

The fundamental frequency of falsetto tends to be higher than in modal voice, with an average pitch range of 275-634 Hz. in males, compared with 94-287 Hz. for modal voice (Hollien and Michel 1968:602). It should be noted that the pitch range does overlap with that for modal voice, and that the two phonation types are differentiated auditorily by a quality difference as well as by pitch. There does, also, seem to be a sex difference in the ease with which individuals can produce falsetto, and in the ease with which falsetto and modal voice can be auditorily discriminated. When training judges in use of the VPAS by getting them to produce the required vocal qualities themselves, it seems that some females have great difficulty in achieving the change from modal voice to falsetto as they raise the pitch of phonation, and there does not always seem to be a clearly audible transition from one phonation type to the other. This seldom seems to be a problem in males. It would be interesting to carry out a detailed physiological study to investigate this observation more fully.

Like modal voice, falsetto is marked only as present, intermittently present or absent on the protocol form.

Harshness

Harshness is a disturbance of the basic vibratory pattern of either voice or falsetto, and it can therefore only occur in combination with one or other of these phonation types. The primary acoustic characteristics of harshness are an irregularity of fundamental frequency (= jitter) and/or of intensity (= shimmer). These irregularities are heard as a general quality of "roughness", rather than as perceptible fluctuations in pitch and loudness.

In organically normal speakers, harshness can only be achieved by a large increase in tension. Medial compression and adductive tension seem to be excessively high, whilst longitudinal tension is probably less pronounced. This means that the adducting edges of the vocal folds are subjected to considerable mechanical abuse during harsh phonation.

Whisper or whisperiness

The whisper(y) setting is used whenever there is audible friction due to turbulent airflow through the glottis. Whisper can occur alone, or in combination with any other setting. It is generally agreed that, when whisper occurs alone as a phonatory setting, there is a triangular opening of the cartilaginous glottis, which allows continuous fricative airflow through the glottis. In weak whisper, this triangular opening may also include part of the ligamental portion of the glottis. This is consistent with low adductive tension together with high or moderate medial compression. The clearly defined glottal chink configuration is probably less common in combined phonation types. Whispery voice is a very common combined phonation type in British speakers, and clinical observations suggest that many speakers produce this phonation with a narrow glottal opening extending well

into the ligamental portion of the glottis, if not along its full length.

In Laver's description of phonation types, a further type of fricative airflow, which can occur only in combination with modal voice, is described as breathiness. This is differentiated from whisperiness by virtue of its very low tension, and the fact that the glottis remains open along most of its length. In breathiness, very high airflow is associated with relatively low levels of audible friction.

In the later versions of the VPAS, breathiness has been deleted from the protocol form. The main reason for this is that breathiness seems to be exceedingly rare, at least in public social interaction. In over two hundred voices recorded for the MRC project, there were no examples of breathy voice. This may be because the high airflow and low intensity of breathy voice make it very difficult to record faithfully. It seems more probable, however, that speakers simply do not use breathiness in the kind of context in which most tape recordings are made. Breathiness is used paralinguistically as a signal of intimacy, and is therefore unlikely to occur in an experimental recording session. Even patients whose physiology might lead to an expectation of breathiness, such as those with vocal fold palsies, seem to compensate in some way so as to avoid giving a false impression of intimacy. The only tape recordings analysed during the span of the project which showed episodes of breathy phonation were examples of mothers interacting with very young infants, presumably because this is one of the rare instances in which a public display of intimacy is socially acceptable.

Another reason for deleting breathiness from the protocol form is that there does in fact seem to be a continuum

between whisperiness and breathiness. The difference between the two is quite a subtle one, depending partially on vocal fold configuration and tension, and partially on subglottic pressure. In a clinical assessment tool like the VPAS it seems adequate to have only one category of laryngeal friction, and to use the degree of audible friction together with a judgement about overall laryngeal tension to discriminate between Laver's "breathy" and "whispery" categories. High degrees of audible friction with high tension levels would be equivalent to Laver's original definition of whisperiness, and low levels of audible friction with lax laryngeal tension would be equivalent to Laver's original definition of breathiness.

Creak or creakiness

The creak setting is reserved for voices in which discrete pulses can be perceived in the phonation. Alternative labels for creak, which may be found in the literature, are 'vocal fry' or 'glottal fry'. To quote Catford (1964:32), "The auditory effect is of a series of rapid taps, like a stick being run along a railing". The frequency of these taps, or pulses has an average range of 24-52 Hz. in males, with a mean of 34.6 Hz. (Michel and Hollien 1968). Like whisper, creak can occur alone, or combined with other settings.

There is some doubt about the physiological mechanism of creak, especially when it occurs in combination with other phonation types such as modal voice (see Laver 1980:122 for a review of relevant studies). Most descriptions suggest that the vocal folds are very thickened in vertical cross section, and may vibrate in tandem with the ventricular folds, which are also adducted. The mass of the folds is not necessarily very tense, and it seems likely that high levels of medial

SETTING	KEY SEGMENTS/PHONETIC CONSEQUENCES	SCALAR DEGREE CONVENTIONS
1. LABIAL; Neutral = LTA‡ lip position as for [ə]		(‡ LTA = long term average)
Lip rounding/ protrusion	[s],[θ] --> low "pitch" /i/ --> unspread or rounded /r,ʃ,t͡ʃ,d͡ʒ/--> rounded	1-3 = open rounding 4-6 = close rounding 3 = LTA position as for [ɔ] 4 = LTA position as for [o] 6 = LTA position as for [u]
Lip spreading	[s],[θ] --> high "pitch" /w/+rounded vowels --> unrounded	4 = LTA position as for [e] 6 = LTA position as for overspread /i/
Labiodentalization	[s] --> low "pitch" /p,b,m/ --> labiodental involvement	1-3 = labiodental onset/offset of labials 6 = /p,b,m/--> labiodental stops
2. MANDIBULAR; Neutral = small vertical gap between incisors, lower incisors just behind upper incisors		
Close jaw		1 = vertical gap just gone 6 = clenched teeth
Open jaw		4 = upper surface of tongue clearly visible 6 = maximum possible opening
Protruded jaw	[s],[ʃ] --> "dark", low "pitch"	4 = lower teeth just outside upper teeth 6 = lower teeth level with centre of upper lip
3. LINGUAL TIP/BLADE; Neutral = /t,d,n,l,s,z/ --> alveolar place of articulation		
Advanced tip/blade	/t,d,n,l,s,z/ --> advanced	4 = fully dental articulation 6 = interdental articulation
Retracted tip/blade	/t,d,n,l,s,z/ --> retracted	4 = post-alveolar, tip points up 6 = pre-palatal, fully retroflex
4. LINGUAL BODY; Neutral = LTA tongue body position as for [ə]		
Fronted tongue body	Back vowels --> less backed	4 = /u/ --> close central vowel
Backed tongue body	Front vowels --> less fronted	4 = /i/ --> close central vowel
Raised tongue body	Open vowels --> ½open or ½close	
Lowered tongue body	Close vowels --> ½close or ½open /j,w/ --> ½close vowels	

CONTINUED OVERLEAF

5, VELOPHARYNGEAL; Neutral = audible nasality only on /m,n,ŋ/		
Nasal	Vowels, continuants --> nasalized	1-3 = open vowels nasalized 4-6 = close vowels nasalized
Aud, nasal escape	Audible nasal friction	4 = escape on a few segments 6 = escape on all segments
Denasal	/m,n,ŋ/ --> lose nasality	1-3 = "cold-in-the-head" 6 = /m,n,ŋ/ --- oral stops
6, PHARYNGEAL; Neutral = no pharyngeal constriction		
Pharyng, constric,	Vowels --> "strangulated"	
7, SUPRALARYNGEAL TENSION; Neutral = moderate tension		
Tense	Extensive ranges, etc,	
Lax	Minimised ranges, nasal	
8, LARYNGEAL TENSION; Neutral = moderate tension		
Tense	Raised larynx, harsh	
Lax	Lowered larynx, whispery	
9, LARYNX POSITION; Neutral = larynx in middle of range		
10, PHONATION TYPE; Neutral = modal voice		
Harshness	Voiced segments --> "rough"	1-3 = voice predominant 4-6 = other phonation predominant
Whisperiness	Voiced segments --> turbulence	
Creakiness	Voiced segments --> pulses	
Falsetto		Present or absent

FIGURE 2.1.2/6: A summary of key segment characteristics for vocal quality settings

compression and adductive tension are associated with low levels of longitudinal tension. Catford (1964:32) and Ladefoged (1971:14) agree that only a small part of the ligamental part of the vocal folds is involved in vibration.

I.B.11) Laryngeal features - overall tension settings

The same general comments apply as for supralaryngeal tension settings. Lax settings often result in lowered larynx, low pitch, and moderate degrees of whisperiness. Tense settings tend to be more often associated with raised larynx, high pitch, and harshness.

A summary chart of perceptual clues for analysis of vocal quality features is shown in Figure 2.1.2/6.

II. Prosodic features

It is harder to offer objective guidelines for the judgement of prosodic features. Pitch is taken to be the perceptual correlate of fundamental frequency, but the perception of pitch is complex, and seems to relate to some degree also to spectral acoustic features. In addition, expectations are affected by the sex, age and physique of the speaker in a way which is not always easy to quantify. Loudness is the perceptual correlate of acoustic intensity, but is very hard to judge from tape-recorded material. It is therefore impossible to give clear definitions of neutral for pitch and loudness settings. It is, however, possible to give general definitions for the prosodic features, and these are summarised below. For most voices these seem to allow a reasonable level of agreement between judges, but the VPAS cannot pretend to be properly objective in this area. Various sorts of acoustic instrumentation are available which can give objective measures of

fundamental frequency and intensity, and it is recommended that these should be used wherever possible.

Pitch

Pitch Mean: this refers to the average perceived pitch for the whole speech sample. It may be judged to be neutral, high or low.

Pitch Range: this is a comment on the span between the highest and the lowest pitch used by the speaker. It may be judged to be neutral, wide or narrow.

Pitch Variability: this refers to the frequency with which a speaker moves around within his or her pitch range.

Consistency

This relates to consistency and coordination of respiratory and phonatory processes. When these break down, the audible result is often tremor. Tremor can be defined as the occurrence of audible fluctuations in pitch and/or loudness, which typically occur at a rate of 1-3 per syllable.

Loudness

The definitions of loudness settings are exactly parallel to those for pitch settings, i.e. loudness mean refers to the long-term-average loudness, loudness range refers to the span between greatest and least loudness, and loudness variability refers to the amount of movement within that loudness range.

III. Temporal organization features

This section is similar to the previous one, in that it is difficult to specify a neutral baseline, so judges should use this simply to make comments about the adequacy or otherwise of a speaker's continuity and rate.

Continuity in this context concerns the incidence of pauses within a speech sample. Marking a speaker as having an interrupted setting implies the presence of inappropriate silent pauses between words or syllables.

Rate is used to describe the actual speed of utterance at the segment or syllable level. This need not necessarily equate with a measure of words or syllables per minute, since a low number of words per minute could be due to a high incidence of pauses rather than a slow rate of syllable production.

It should be clear that these categories of the VPAS are inadequate to allow full description of speakers, such as stammerers or dysarthrics, where disrupted temporal organization is a major feature. They do, however, act as place holders, signalling the need for further specialized investigation.

IV. Comments

The VPA protocol also allows comments on breath support, rhythmicality and diplophonia. Breath support may be marked as adequate or inadequate for normal speech production. Rhythmicality is similarly scored as adequate or inadequate, although this may seem a slightly odd concept. The acceptability of the rhythm used by a speaker will obviously depend both on linguistic content, and on language or accent. Syllable timing, for example, would be appropriate, and therefore adequate, in French,

but be undoubtedly inappropriate in most British speech communities.

Diplophonia is obviously closely related to phonation type, but until there is clearer agreement about the physiological and acoustic bases for diplophonia it cannot properly be placed within a phonetic theory. The perceptual definition for diplophonia used here is that two fundamental pitches should be audible simultaneously. This excludes some voices which are often described as diplophonic, where there is rapid fluctuation of pitch, often associated with an alternation between modal voice and falsetto. Diplophonia is scored simply as being present, intermittently present (by the use of the 'i' convention), or absent.

Any other comments which are felt to be relevant to an individual's voice can be included at the bottom of this section.

Training judges to use the VPAS

The above notes, together with Laver's (1980) book, should allow a general understanding of the underlying theory and the general principles for using the VPAS, but it must be stressed that self-tuition from written material, even in conjunction with audio tapes, is not considered to be a feasible proposition. The essential feature of the VPAS is that it is a perceptual analytic system, parallel to the traditional skill of segmental analysis. Most phoneticians would accept that the perceptual skills required for segmental analysis require personal training from a skilled phonetician, and the same is true for the VPAS. For this reason, some time was spent during the MRC project in trying to ascertain the best procedure for training judges to use the scheme. To date, the author has been involved in training over 200

individuals in Britain, Holland and Australia. Training workshops varied in size from 6 to 25 individuals, most of whom were speech therapists, although some groups also included phoneticians, psychologists and drama teachers. Almost all workshop participants were already trained in segmental phonetic analysis, and this appears to be an essential prerequisite of successful training.

The exigencies of fitting training into professional lives, of both participants and tutors, meant that two basic formats seemed most appropriate. The first format involved ten two hour training sessions, at weekly intervals, taking place out of work hours. The second format involved intensive training during three day courses, usually with a follow up session a few weeks later. Full experimental comparison of these two approaches has not been attempted, because the less intensive approach was only used in earlier workshops, when the training materials used during workshops had not yet been fully developed.

Workshop participants consistently state that, wherever possible, they would favour the intensive approach, because it is easier to consolidate the new approach to phonetic analysis if they are immersed in it for several days. This is the format which is therefore now routinely used in training users of the VPAS.

Training techniques are based largely on the traditional Edinburgh approach to phonetic teaching. In other words, a combination of ear training, from live and taped material, and performance of voice settings is used. Although workshops sometimes involve up to 25 people, the aim is always to spend a substantial proportion of the time in smaller groups so that enough personal tuition is possible. To this end, larger workshops always involve at least two tutors. At the end of each workshop, all

participants complete a set of vocal profile analyses from a standard evaluation tape, so that the success, or otherwise, of the training can be assessed, from a comparison of the individual judge's analysis with that of expert judges (the MRC team).

Interjudge and intrajudge agreement

General procedure for evaluating judge agreement levels

The VPA protocol does not lend itself easily to statistical evaluation of judge reliability. The design of perceptual evaluation protocols often faces a conflict between opposing needs. In this case, statistical methodology was not always compatible with the wish for the protocol form to be an accurate reflection of the underlying phonetic theory. The compromise reached by the MRC team might be criticized as not allowing easy statistical manipulation. This is a valid criticism, and one which can be countered only by asserting that the primary purpose of the VPAS is to provide a useable, easily interpreted system for clinical evaluation of voice quality, which relates to physiological and anatomical features. This was felt to override the exigencies of statistical testing of results. For example, it would be simpler, statistically, if all setting scales were of equal length, but this seems a very artificial constraint when the phonetic basis of the scheme is considered. There are some setting scales where deviation from neutral can occur in one direction only. Examples include labiodentalization and protruded jaw, where neutral forms the first point of a 7-point scale. For other setting scales, deviations from neutral may occur in diametrically opposed directions. Examples include rounding or spreading of the lips and fronting or backing of the tongue body, where neutral effectively forms the mid-point of a 13-point scale. However much

easier it might be to have equal setting scales, the loss of the ability to reflect such real differences in the phonetic basis of these settings seems too great a sacrifice.

There are other difficulties involved in statistical analysis of VPA judgements. The most obvious of these is that the layout of the form, and the judges' understanding of the terms "normal" and "neutral" makes it very hard to formulate a nul hypothesis for any statistical test. It does not seem reasonable to expect judges to fill in the form in a completely random way, even if they are unsure or poorly trained in perception of voice quality features. Judges seem to be more inclined to mark the neutral box when in doubt, and we have chosen to use this assumption in the statistical tests described below.

These difficulties forced a very simple approach to the evaluation of judge reliability, which was used to look at both inter- and intra-judge agreement levels. For every pair of protocols which was compared, the number of setting scales on which the two judgements agreed was recorded. "Agreement" can be defined in various ways, depending upon the level of accuracy required. Three definitions were used here.

a) Complete agreement: this demands that two judgements fall in exactly the same scalar degree box.
b) Agreement within one scalar degree: for this, two judgements must fall either in the same scalar degree box, or in adjacent scalar degree boxes.
c) Agreement within two scalar degrees: for this, the two judgements must be no more than two scalar degrees away from each other. A further criterion is that, where 13-point scales are concerned, both judgements must be on the same side of neutral.

Figure 2.1.2/7 gives examples of various pairs of judgements, showing how they would be assessed in terms of these three levels of agreement.

Three sets of agreement figures give some indication of the level of agreement which can be expected for the VPAS. The first two sets of figures relate to performance of two of the MRC staff who were responsible for development and teaching of the scheme (John Laver and the author), and the third set of figures relates to the performance of newly trained judges at the end of their initial three-day training workshops. The first two sets of figures are most important, because they are necessary for interpretation of the studies reported in Sections 2.2 and 2.3.

1. Interjudge reliability (MRC staff)

Interjudge reliability was evaluated using two sets of voices; normal control speakers (N=25) and speakers with Parkinson's Disease (N=13). The reason for using the control group was that it was felt that Vocal Profile Analysis was easiest when there were no organic or physiological abnormalities, and that the control group would therefore give an indication of the best achievable agreement. The Parkinson's Disease (PD) group, which was analysed as part of a collaborative study with Ms. Sheila Scott and Professor F.I. Caird of the Southern General Hospital, Glasgow, was chosen to represent almost the worst possible case. One of the results of the neurological deficit in these patients is a tendency for voice quality settings to fluctuate during a speech sample. The consistency of settings displayed by normal speakers is lost, and this makes it very hard for judges to abstract long term average biases from the speech sample. It was felt that if reasonable levels of agreement could be achieved even for this group, then it

1.



2.



Ⓡ = Edinburgh consensus ("right answer")

A,B,C,D = Judgements being scored.

In example 1., judgements A and B are both within one scalar degree of the "right answer", and they are therefore scored as being correct at both levels of accuracy (i.e. within one scalar degree and within two scalar degrees). Judgements C and D are not correct within one scalar degree, but they are scored as being correct within two scalar degrees. In these cases, the judges have correctly identified the setting, but not the scalar degree. In other words, they are right on quality, but not quantity.

In example 2., judgement A would be correct within two scalar degrees, but not within one scalar degree. Judgement B, however, is scored as incorrect even at the two scalar degree level. This is because it is on the opposite side of neutral from the "right answer", and thus involves an error in identification of the setting. It is therefore wrong on quality, rather than just on quantity.

FIGURE 2.1.2/7: Criteria for assessing judge agreement

would be fair to say that the VPAS is a widely applicable clinical tool.

Vocal Quality Features

Interjudge agreement figures for the two MRC staff members are shown in Figure 2.1.2/8 (control group) and Figure 2.1.2/9 (PD group). Audible nasal escape (ANE), falsetto and modal voice were excluded from the analysis. This was because modal voice was present as a component of phonation type in all voices, whilst falsetto and ANE occured in none, and it was felt that high levels of agreement on these parameters would unfairly bias the overall results.

It can be seen that absolute agreement is not very high (65.3% for the control group, 40.2% for the PD group), but that the overall percentage of agreements within one scalar degree is 94.2% for the control group, and 72.1% for the PD group. These figures seem to indicate high levels of interjudge agreement, but some statistical evaluation of these results is desirable. Making the assumption mentioned earlier that judges tend to mark neutral if they are unsure, these levels of agreement were compared with the level of agreement which would be achieved if one judge scored a voice to be neutral for all parameters. A χ^2 test, the McNemar test for significance of changes (Siegel 1956:63), was used to test the following null hypothesis; the probability that JL and JM agree within one scalar degree equals the probability that JL's (or JM's) judgements are within one scalar degree of neutral. The test compares two figures:

a. the number of judgements where JL (or JM) is within one scalar degree of the other judge, but is more than one scalar degree away from neutral, and

SETTING SCALE	JL/JM		JL/JL		JM/JM	
	A.	B.	A.	B.	A.	B.
Lip rounding/spreading	40	84	40	90	40	80
Labiodentalization	100	100	100	100	90	100
Labial range	76	100	70	100	70	100
Close/open jaw	68	96	60	100	50	90
Protruded jaw	100	100	100	100	100	100
Mandibular range	84	96	80	100	50	100
Tip-blade	36	72	20	90	60	70
Fronted/backed T.B.	28	88	30	80	20	80
Raised/lowered T.B.	64	96	30	70	50	100
Lingual range	96	100	80	100	100	100
Nasal/denasal	52	100	40	100	80	100
Phar. constriction	80	100	50	100	50	100
Supralar. tension	56	100	50	90	70	100
Laryngeal tension	48	92	30	100	60	90
Larynx position	56	92	80	100	40	100
Harshness	84	96	100	100	80	100
Whisperiness	48	100	60	100	60	90
Creakiness	60	84	30	80	50	80
Total: Vocal Quality	65·3	84	58·3	80	62·2	80
Pitch mean	36	96	30	90	40	80
Pitch range	76	96	80	90	80	100
Pitch variab.	88	96	90	90	80	100
Tremor	76	100	100	100	60	100
Loudness mean	60	96	80	90	70	100
Loudness range	88	99	90	90	80	100
Loudness variab.	92	96	90	90	90	100
Total: Prosodic	73·7	95·1	80·0	93·8	71·4	94·6

FIGURE 2.1.2/8: Table showing percentage levels of inter- and intra-judge agreement for control voices (MRC project staff)
A. Absolute agreement
B. Agreement within one scalar degree

SETTING SCALE	JL/JM		JL/JL'		JM/JM	
	A.	B.	A.	B.	A.	B.
Lip rounding/spreading	46·2	61·5	30·0	70·0	40·0	60·0
Labiodentalization	100	100	100	100	90·0	100
Labial range	61·5	84·6	40·0	80·0	80·0	100
Close/open jaw	15·4	61·5	30·0	70·0	0·0	40·0
Protruded jaw	61·5	84·6	80·0	80·0	30·0	100
Mandibular range	38·5	69·2	0·0	70·0	20·0	70·0
Tip-blade	15·4	69·2	10·0	80·0	20·0	60·0
Fronted/backed T.B.	30·8	53·8	70·0	100	20·0	50·0
Raised/lowered T.B.	38·5	61·5	50·0	80·0	30·0	80·0
Lingual range	38·5	76·9	50·0	70·0	30·0	100
Nasal/denasal	53·8	84·6	50·0	90·0	80·0	90·0
Phar. constriction	23·1	38·5	30·0	60·0	20·0	90·0
Supralar. tension	23·1	53·8	30·0	90·0	30·0	50·0
Laryngeal tension	7·7	38·5	10·0	80·0	30·0	70·0
Larynx position	46·2	92·3	70·0	80·0	40·0	100
Harshness	30·8	69·2	40·0	60·0	50·0	80·0
Whisperiness	46·2	84·6	50·0	100	60·0	100
Creakiness	46·2	84·6	70·0	100	60·0	70·0
Total: Vocal Quality	40·2	72·1	45·0	82·1	40·6	79·5
Pitch mean	15·4	30·8	40·0	90·0	20·0	30·0
Pitch range	38·5	84·6	30·0	80·0	30·0	50·0
Pitch variab.	30·8	53·8	40·0	90·0	30·0	60·0
Tremor	38·5	69·2	10·0	20·0	30·0	60·0
Loudness mean	53·8	69·2	40·0	90·0	50·0	80·0
Loudness range	30·8	76·9	50·0	70·0	40·0	70·0
Loudness variab.	23·1	69·2	60·0	70·0	30·0	60·0
Total: Prosodic	33·0	64·8	38·6	72·9	32·9	60·0

FIGURE 2.1.2/9: Table showing percentage levels of inter- and intra-judge agreement for Parkinson's Disease voices (MRC project staff)
A. Absolute agreement
B. Agreement within one scalar degree

b. the number of judgements where JL (or JM) is within one scalar degree of neutral, but is not within one scalar degree of the other judge.

This test was performed for both judges, and for both subject groups, and in all cases the null hypothesis could be rejected, with a probability of less than 0.001. In other words, these two judges do agree with each other significantly better than would be expected if one judge ticked neutral for all vocal quality features.

It can be seen that the level of agreement is much higher for some setting scales than for others. The range of percentage agreement within one scalar degree is 72%-100% for the control group, but only one setting scale has less than 84% of judgements agreeing within one scalar degree. This is the tip/blade setting scale, and the relatively poor level of agreement here is probably explained by the accent characteristics of the control group. For many of these speakers from the south of Scotland, the realization of the so-called alveolar segments /t,d,n,s,z,l/ falls into at least two classes, according to the place of articulation. This makes abstraction of the overall tip/blade setting more difficult, as it is easy for the dental articulation of a small subset of these segments to become so perceptually prominent that the judge fails to analyse the remaining segments properly.

The overall level of agreement is much lower for the PD group, with only 38.5% agreement within one scalar degree for pharyngeal constriction and laryngeal tension. The poor agreement for these parameters may well reflect the fact that the normal balance of muscular tension within the vocal apparatus is disturbed in PD, so that changes in overall muscular tension are often not associated with the expected constellations of individual settings.

Prosodic Features

Analysis of agreement for the prosodic section was carried out in exactly the same way. The percentage agreement figures for the controls (see Figure 2.1.2/8) look fairly good, with absolute agreement at 73.7% for controls and agreement within one scalar degree at 96.6%, but the χ^2 test did not allow rejection of the null hypothesis at a probability level of 0.1. This lack of significance is easily explained by the observation that few control subjects were judged to have prosodic settings which deviate from neutral by more than one scalar degree.

The percentage agreement figures for the PG group was less good (see Figure 2.1.2/9), at 33.0% absolute agreement, and 64.8% agreement within one scalar degree. Again, the χ^2 test suggested that these levels of agreement were not significantly better than would have resulted if one judge ticked neutral throughout.

Intrajudge reliability (MRC staff)

In order to test intrajudge reliability, 10 of the control group voices and all of the PD group voices were reanalysed after a three month interval. It was assumed that this interval would allow the judges to forget their original judgements. Agreement was assessed as for interjudge agreement, described above. The percentage agreement figures are included on Figures 2.1.2/8 and 9.

It can be seen that the percentage agreement for vocal quality features is very similar to interjudge agreement. Absolute agreement is 58.3-62.2% for the controls and 54.0-40.6% for the PD group. Agreement within one scalar degree is 93.3-94.4% for controls and 82.1-79.5% for PD voices. Again, χ^2 tests show that these levels of

agreement are significantly better than if one set of judgements was neutral throughout.

The levels of intrajudge agreement for prosodic features are also comparable with interjudge agreement, and the χ^2 test results were similarly non-significant.

The finding that, although agreement within one scalar degree is high, absolute agreement is not, is an indication that any research application of the VPAS should be based on analysis by more than one judge. At least three judges should probably be used whenever possible.

Interjudge reliability (Training panels)

The reliablity of newly trained judges is less important as a basis for evaluating the studies outlined in this thesis, but, since it does throw some light on the general usefulness of the VPAS as a clinical tool, a brief summary of the levels of agreement reached at the end of training workshops will be included here.

At the end of each training workshop, trainees were asked to evaluate 6 voices, which were chosen to represent a range from normal to substantially pathological. The presentation of this evaluation tape was standardised, with each sample being repeated until the total exposure to each voice was about three minutes. There were short pauses between each voice sample, so that the total time spent analysing this tape was about 40 minutes.

The trainees' judgements were compared with a consensus vocal protocol for each voice, derived from the judgements of three MRC staff (JL, SW and JM). Each of the MRC judges listened to the tape separately, and the three sets of judgements were compared. Where there was

absolute agreement between all three judges, there was no difficulty in selecting the consensus judgement. In other cases, the following guidelines were used in order to construct a consensus protocol. When two judges agreed, and the other judge selected an adjacent scalar degree box, the majority decision was chosen. When the three judgements were spread over three adjacent scalar degrees, then the central scalar degree was chosen. In all other cases, the three judges reanalysed the voice. If there was still disagreement, the reasons for this were discussed and a joint decision was made. This last option was seldom used in practice, and when it was necessary, there was usually found to be some anomalous segmental articulatory feature which was causing difficulty in abstracting a long term average setting judgement from the speech sample.

These six consensus vocal protocols were then taken as the "correct" answers to the evaluation exercise, and the trainees' protocols were compared with these as described above. Figure 2.1.2/10 is a histogram showing the spread of overall agreement displayed by 106 trainees, from 6 training groups. Agreement within one scalar degree and within two scalar degrees is calculated for these judges.

It is difficult to decide what is an adequate level of accuracy for a judge to be considered a competent user of the scheme. χ^2 tests (as described in relation to MRC staff reliability) suggest that judges who agree with the "correct" answers within one scalar degree for 60% of setting scales and within two scalar degrees for 79% of setting scales are performing significantly better ($P < 0.001$) than if they ticked neutral throughout. These levels of agreement were felt to be adequate for routine use of the VPAS in a clinical situation, although trainees were encouraged to continue practicing their skills in small groups and to complete analyses in

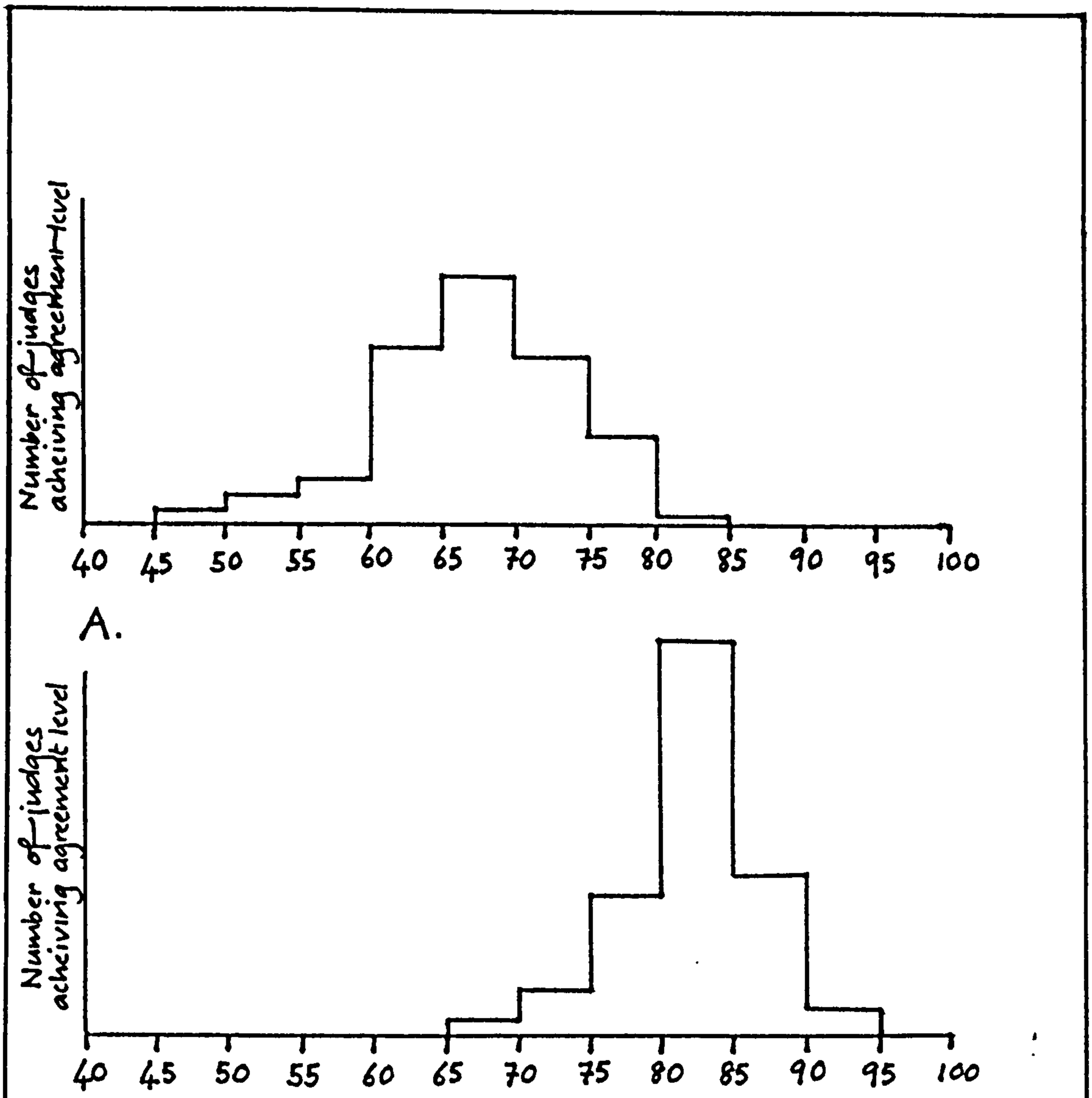


FIGURE 2.1.2/10: Histograms showing distribution of trainee judge agreement levels

A. Within one scalar degree

B. Within two scalar degrees

conjunction with other trained judges wherever possible. It was found that 77.4% of trainees attained the 60%/79% level of agreement criteria of adequate performance.

Figure 2.1.2/11 tabulates the levels of agreement by setting scale. It can immediately be seen that agreement is much higher for some settings than for others. This raises questions about whether some settings are inherently more difficult to judge, or at least take longer to learn, than others, and whether these settings should therefore be excluded from the scheme. The agreement figures for tongue body settings, for example, are quite low, and many therapists expressed doubts about the wisdom of retaining these settings on the VPA protocol. It does seem that speech therapists in Britain lack confidence in their ability to analyse vowels phonetically, and tend to concentrate on consonant segments when analysing speech data. Some difficulty in learning to differentiate tongue body settings is therefore perhaps not surprising. Evidence that tongue body settings can be judged adequately, even if they take longer to learn, is given by the observation that interjudge agreement amongst the MRC judges was reasonably high.

Applications of the VPAS

Investigation of group characteristics

During the course of the MRC project, the VPAS was used to investigate the group characteristics of several groups of speakers, two of which, normal young adults and adults with Down's Syndrome, will be described in later sections. The basic procedure for investigating the vocal profile characteristics of these groups was always the same. Firstly, the MRC judges listened to each subject's voice independently, and without reference to any medical

SETTING SCALE	PERCENTAGE CORRECT JUDGEMENTS	
	WITHIN 1 S.D.	WITHIN 2 S.D.
Lip rounding/spreading	52·98	75·94
Labiodentalization	95·60	97·64
Labial range	67·77	79·87
Close/open jaw	66·19	77·83
Protruded jaw	63·84	79·72
Mandibular range	72·64	91·19
Tip-blade	55·19	71·38
Fronted/backed T.B.	30·03	65·41
Raised/lowered T.B.	49·37	68·87
Lingual range	78·30	87·58
Nasal/denasal	67·45	79·40
Phar. constriction	66·19	81·60
Supralar. tension	58·49	74·53
Laryngeal tension	62·58	87·26
Larynx position	59·28	77·04
Harshness	63·21	82·70
Whisperiness	84·59	91·98
Creakiness	65·41	83·18
Mean: Vocal Quality	64·40	80·73
Pitch mean	54·76	69·84
Pitch range	58·73	78·77
Pitch variab.	59·13	78·57
Tremor	87·70	90·67
Loudness mean	75·99	90·08
Loudness range	71·23	90·48
Loudness variab.	70·63	88·10
Mean: Prosodic	68·31	83·79

FIGURE 2.1.2/11: Table showing percentage levels of trainee judge accuracy
A. Correct within one scalar degree
B. Correct within two scalar degrees

or biographical information. Consensus vocal protocols were then drawn up for each speaker, and these formed the data base for an examination of subject groups. For most subjects, all three MRC judges completed protocols, but unfortunately one staff member (SW) was not available to complete analyses for some of the control group voices. Given the high levels of interjudge agreement for control speakers (see above), it was felt that the VPA results for control speakers should be included in the final analyses, even where only two judges had contributed to the consensus vocal protocols.

The next step was to draw up summated vocal protocols for each subject group. Steven Hiller devised a set of programs for storing vocal protocols on computer discs, and for amalgamating specified sets of protocols to show 'summated' protocols. An example of a summated protocol is shown in Figure 2.1.2/12. The numbers in the cells represent numbers of individual subjects judged as showing the scalar degree of the setting concerned. These summated protocols allow an easy visual examination of the spread of settings displayed by any subject group, and can be used to produce various descriptive and comparative statistics.

Simple descriptive statistics which proved to be useful include the percentage of any group which display a given voice quality setting, the mean scalar degree for each setting scale, and the standard deviation. The statistical significance of mean scalar degree differences between groups can be calculated using such tests as the Mann-Whitney U test (Siegel 1956:116-127).

Longitudinal studies

The VPAS allows simple measurement of changes in voice quality over time. In times of economic stringency,

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial	12	28		Lip Rounding/Protrusion	6	12	4			
				Lip Spreading	1	5				
				Labiodentalization			1			
	8	25	7	Extensive Range	1	4	4		1	
				Minimised Range	2	6	8	3	2	1
2. Mandibular	13	27		Close Jaw	1	2	3			
				Open Jaw	3	4	9			
				Protruded Jaw	5	2				
	8	25	7	Extensive Range	3	2	2	1		
				Minimised Range	2	6	10	5		1
3. Lingual Tip/Blade	11	26	3	Advanced	4	4	2			
				Retracted	10	4	2	3		
4. Lingual Body	3	21	5	Fronted Body		6	2			
				Backed Body	2	11	8	4		1
		26	1	Raised Body	1	10	5			
				Lowered Body	4	1	5	1		
	1	16	23	Extensive Range						
				Minimised Range		3	13	14	9	
5. Velopharyngeal		14	26	Nasal			10	25	3	
				Audible Nasal Escape				4	1	
				Denasal			2			
6. Pharyngeal	5	27	8	Pharyngeal Constriction	6	10	10	7	2	
7. Supralaryngeal Tension	3	30	7	Tense	3	12	11	5	3	
				Lax	2		1			
B. Laryngeal Features										
8. Laryngeal Tension		28	12	Tense	6	10	10	10	2	
				Lax		2				
9. Larynx Position	7	32	1	Raised	7	13	2			
				Lowered	5	2	3	1		
10. Phonation Type		24	5	Harshness	3	10	11	4	1	
		32	5	Whisper(y)	2	11	19	5		
				Breathiness						
		24	3	Creak(y)	5	10	9	2	1	
		8		Falsetto	8 (7=1)					
		39		Modal Voice	39					

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

FIGURE 2.1.2/12: An example of summated Vocal Profile Analysis results: A group of 40 speakers with profound hearing impairment

health care professionals are under increasing pressure to validate and compare the efficacy of various forms of therapy, and the VPAS offers an ideal means of assessing vocal change. A simple means of assessing the statistical significance of vocal profile changes following therapy is to use the Sign test (Siegel 1956:68), and this was successfully used to assess vocal improvement following speech therapy in the study of speakers with Parkinson's Disease which was mentioned earlier.

Further applications

It should already be clear that Vocal Profile Analysis has many applications within phonetics, including such things as the investigation of accent characteristics, or interpersonal variation. The value of the VPAS in speech therapy and medicine has also been stressed, but the importance of voice quality in all vocal communication means that the VPAS is also of interest to many other disciplines. It has already been used in the study of emotion (Bezooijen 1984) and of normal mother-child interactions (Marwick et al. 1984), and it will shortly be applied to an investigation into interpersonal interactions of mothers suffering from post-natal depression.

2.1.3 ACOUSTIC ANALYSIS OF LARYNGEAL BEHAVIOUR

In the introduction to Part Two, the difficulties of using acoustic analysis for analysing a speaker's overall voice quality were mentioned briefly. There are major difficulties in the acoustic separation of the effects of combinations of settings of the supralaryngeal vocal tract. Acoustic techniques do seem to have considerable potential, however, when attention is more narrowly focussed on the larynx and phonation. Whilst the need for the vocal tract to be viewed as a whole, integrated system has been repeatedly stressed, there are some situations where it is useful to use an accurate and objective means to acquire more detail about one specific aspect of voice quality. One such situation occurs when a patient arrives in clinic with a known or suspected disorder of the vocal folds. The most direct consequence of a vocal fold disorder is likely to be some disruption of vocal fold vibration or adduction, and acoustic measurements of the laryngeal wave-form may give information which is more accurate than a perceptual judgement. This is not to say that the consequences will necessarily be confined to the larynx. Perceptual analysis of the voices of patients with vocal fold disorders shows that they are often associated with unusual voice quality settings of the supralaryngeal vocal tract as well, and the clinician needs always to be aware of this. Acoustic and perceptual techniques should be seen as complementary to each other; whilst perceptual techniques may be able to give more information about the way in which the whole vocal apparatus is behaving, acoustic techniques may be more appropriate for examining details of specific vocal features.

The automatic acoustic system which is described here was developed during the course of the second M.R.C. funded project on which the author was employed. Although the

author was involved in its development, John Laver and Steven Hiller were primarily responsible for its design, and Hiller was solely responsible for all the necessary programming. A full account of the system can be found in Hiller (1985).

The motivation for developing a computer-based system for the acoustic analysis of phonation was the hypothesis that it might be possible to use such a system to screen voices for the presence of laryngeal pathology, and, further, that it might be possible to differentiate various types of pathology using acoustic measures alone. The justification for such a hypothesis is explained in Section 2.5.

The analysis system is designed to provide two types of data, which can be related to theoretical predictions about the acoustic consequences of specific classes of laryngeal disorder. The first type of data can very loosely be described as intonational data. This includes the mean and range of fundamental frequency used in a speech sample. The second type of data involves the amount of perturbation of the laryngeal waveform. The procedure by which this data is acquired will be described very briefly below, but the reader is referred to Hiller (1985) for a complete description.

The analysis system has several characteristics which potentially improve its accuracy relative to most other available systems. One important feature is that it is capable of analysing samples of continuous speech of at least 40 seconds length. The majority of studies which have applied comparable acoustic measures to the speech of patients with laryngeal disorders have used very short speech samples consisting of sustained vowels (Iwata and von Leden 1970, Koike 1973, Katajima et al. 1975, Koike et al 1977, Deal and Emanuel 1978, Murray and Doherty

1980, Kasuya et al. 1983, Ludlow et al. 1983a, 1983b, Kane and Wellen 1985). Isolated vowels constitute speech samples which are both short and rather artificial. There is good reason to suppose that such samples are not properly representative of a speaker's habitual speech patterns, and may therefore be poor indicators of pathology.

The ability of many speakers to compensate for organic disturbances by developing new patterns of muscular activity means that the acoustic consequences of minor changes in organic state may be veiled. The question of what kind of speech sample is most likely to display such acoustic consequences most clearly is open to debate. It does seem likely that it is easier to maintain compensatory adjustments, thus masking the effects of any organic abnormality, for a short period of sustained phonation than during a longer period of continuous speech. One reason for this belief is the hypothesis that some kinds of vocal fold pathology will initially interfere most with the onset or offset of phonation, resulting in increased levels of perturbation at voicing transitions. This kind of phonatory disturbance would not be evident in a steady state vowel, but might be picked up in a sample of connected speech involving many initiations of phonation. The artificial nature of sustained vowel production makes it difficult to ensure that the fundamental frequency will be typical of a speaker's habitual pitch, and it will certainly not allow evaluation of the speaker's habitual pitch range. As Section 2.5 will show, there are strong theoretical reasons for trying to relate pitch mean and range to the type of laryngeal pathology found, so that a realistic assessment of these features is very important.

A pilot study showed that the measured values of most acoustic parameters fluctuate randomly for the first few

seconds of speech analysed, and that as much as 40 seconds of continuous speech is needed to ensure that all the acoustic parameters used in this study have stabilised (Hiller et al. 1984). In other words, if less than 40 seconds of speech is analysed, it is not possible to be certain that the acoustic values obtained are fully representative of that speaker.

The acoustic analysis system

A digitised speech waveform is derived from good quality tape recordings of 40 seconds of continuous speech. This is phase compensated, to correct low-frequency distortion introduced by tape recording equipment, and low-pass filtered to remove higher frequency resonance effects. The filter values are set at 600 Hz for males, and 800 Hz for females.

The acoustic analysis system is implemented on a VAX 11/750 computer, and proceeds through three stages. Firstly, the fundamental frequency (F0) and amplitude contours of the speech wave form are analysed in detail, using a modified version of the Gold and Rabiner (1969) parallel processing pitch detection algorithm. Secondly, these "raw" F0 and amplitude curves are smoothed statistically, to produce trend lines, which preserve the overall shape of long-term movements, but with cycle-to-cycle deviations in F0 and amplitude smoothed out. This is done using a non-linear smoother, adapted from work reported by Rabiner et al. (1975). Finally, the differences between the raw and the smoothed curves are analysed, and used to produce measures of pitch and amplitude perturbation (jitter and shimmer). These stages are shown as a flow chart in Figure 2.1.3/1. The general principles and definitions of perturbation analysis used here may be clarified by reference to Figure 2.1.3/2, which is a schematic representation of the output of the

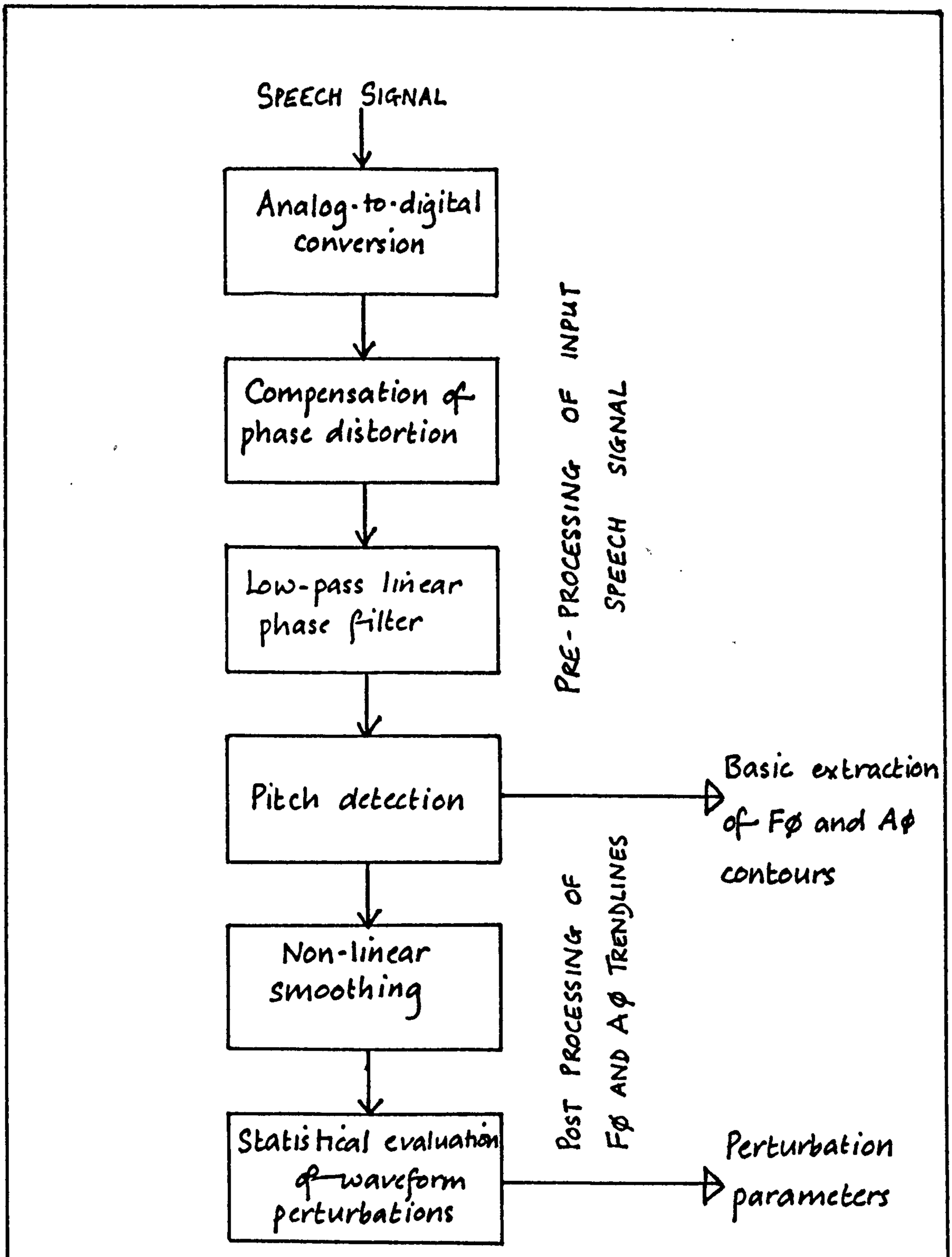


FIGURE 2.1.3/1: Flow chart of the perturbation analysis system (adapted from Hiller 1985:11)

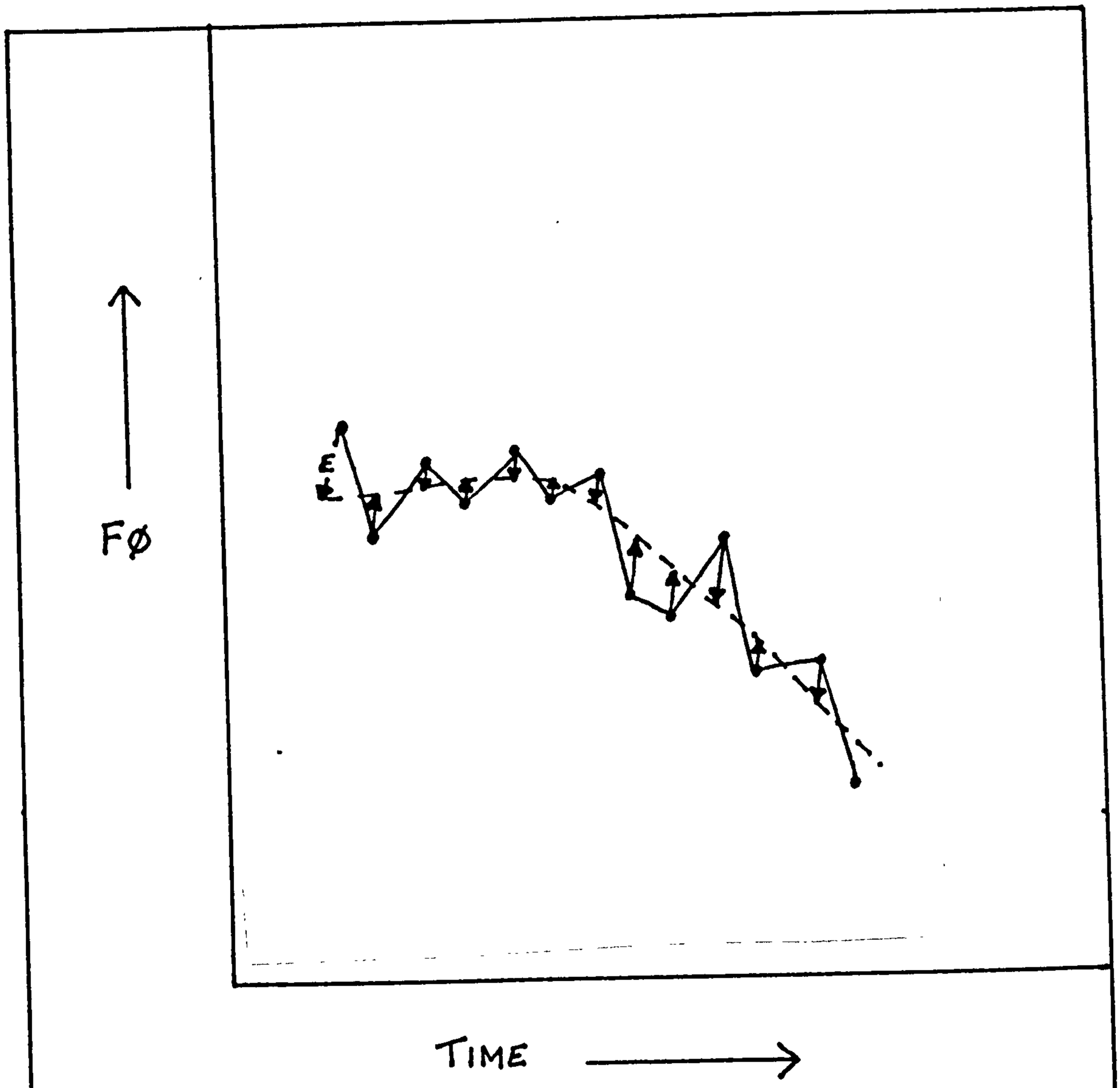


FIGURE 2.1.3/2: A diagrammatic representation of raw and smoothed F0 curves, and excursions.

first and second stages of analysis. The solid line represents a raw F0 curve, resulting from the measurement of every pitch cycle. The dotted line represents the smoothed F0 trend line. The basic units of analysis which are involved in the third stage are shown by the vertical arrows. For each pitch cycle, the difference between the raw curve and the smoothed trend line is measured. These deviations will be called excursions.

The two classes of data output by this system have already been mentioned. The first class, which is loosely termed "intonational data", is derived from the smoothed F0 trend line. The second class, the perturbational data, is derived from statistical analysis of F0 and amplitude excursions. The measures taken are summarised below.

A. Intonational data

1. F0 AV: mean fundamental frequency
2. F0 DEV: standard deviation of the fundamental frequency. This gives an indication of the pitch range used.

B. Perturbation data

The following measures are taken for both jitter (J) and shimmer (S).

1. AVEX: the average magnitude of excursions of the raw F0 or amplitude contour from the smoothed trend line.
2. DEVEX: standard deviation of the excursions from the trend line.
3. RATEX: the rate of excursions. This is the percentage of points in the sample where the magnitude of excursions is greater than, or equal to, 3% of the local trend line.
4. DPF: The directional perturbation factor. This measure, which is adapted from Hecker and Kreul (1971), is the percentage of changes

in algebraic sign between adjacent pitch or amplitude estimates in the raw curves. A 3% threshold is also applied to this measure.

The imposition of a 3% threshold for RATEX and DPF results from the observation that even speakers with healthy larynges typically show jitter levels of around 2% when producing monotone, steady-state vowels (Hanson 1978). In fact, the acoustic results obtained from normal speakers using this system suggest that higher levels of perturbation are perfectly normal when longer samples of continuous speech are analysed (see Section 2.4).

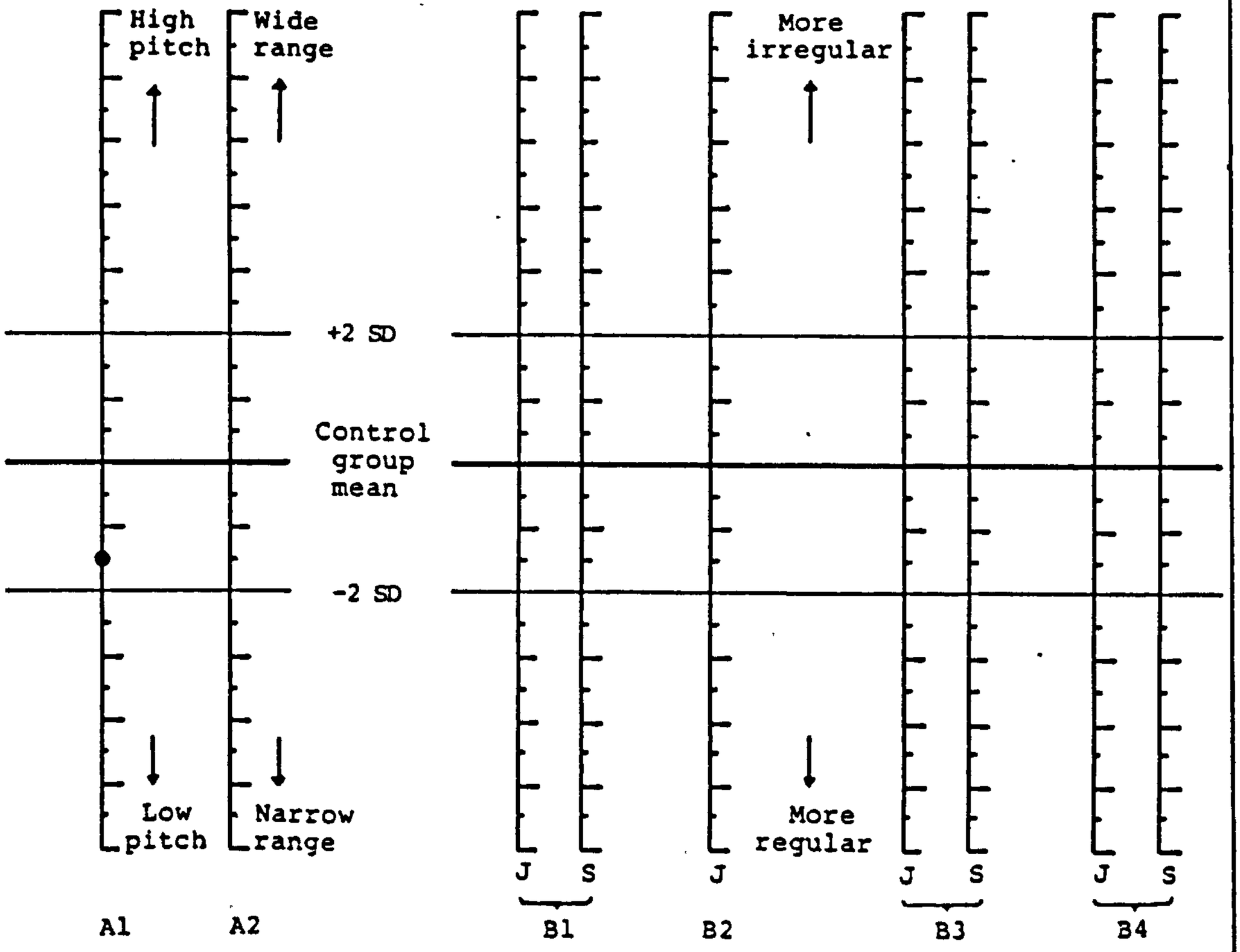
The measurements obtained from the acoustic system can be used in various ways, depending on the emphasis of the investigation. When the primary aim of a study is to study group characteristics, or to compare the characteristics of one group with another, then a variety of statistical procedures is available (see Section 2.5). When an individual, or a small number of individuals, is being assessed in detail, then it is often useful to be able to display the results in the form of an acoustic profile, which compares individual acoustic features to a normal baseline. The author has designed a protocol form for this purpose, which has proved to be quite valuable in the initial assessment of dysphonia patients, and for tracking the acoustic changes which accompany speech therapy. This has been used in conjunction with the VPAS in a collaborative series of case studies undergoing therapy at the Royal Infirmary, Edinburgh (see Nieuwenhuis and Mackenzie 1986, included as an Appendix). The acoustic profile form is shown in Figure 2.1.3/3, and will be discussed further in Section 2.5.

ACOUSTIC PROFILE

Speaker: Sex: Age: Date:

A. PITCH MEASUREMENTS
= smoothed FØ

B. MEASUREMENTS OF PHONATORY IRREGULARITY
J = JITTER (pitch irregularity)
S = SHIMMER (intensity irregularity)



A1 = Pitch mean (mean FØ)

A2 = Pitch variability (SD FØ)

B1 = Average size of irregularities (AVEX)

B2 = Standard deviation of irregularities (DEVEX)

B3 = Percentage of substantial irregularities (RATEX)

B4 = Percentage of substantial reversals in pitch/intensity contour (DPF)

"ACOUSTIC ANALYSIS OF VOICE FEATURES" Research Project.
(MRC Grant No. G8207136) Centre for Speech Technology Research,
Department of Linguistics, University of Edinburgh.

FIGURE 2.1.3/3: Acoustic Profile form

2.2 PERCEPTUAL ANALYSIS OF NORMAL VOICE QUALITY

As part of the MRC project "Vocal Profiles of Speech Disorders" a group of 50 young adult speakers were recorded to form a control group for comparison with other groups of pathological speakers (including speakers with hearing impairment, cerebral palsy, dysphonia, Down's Syndrome and Parkinson's Disease). This section will be a brief one, giving a summary of the findings for the control group only. This will act as a background for Section 2.3, which examines the vocal profile characteristics of a Down's Syndrome population, as illustration of the vocal consequences of a global disorder of growth and development. The control group study is interesting in its own right, as it allows an assessment of the distribution of voice quality settings in a normal population. The group included a variety of accent types, but since all subjects were resident in the Edinburgh area at the time of recording there was a preponderance of south east Scottish accents. Any subjects reporting a history of hearing loss or of speech or voice problems were excluded from the group.

Subjects and procedures

The group consisted of 50 young adults (25 females and 25 males), aged between 18 and 40 years. All were native speakers of English, resident in Scotland.

The procedures for completing vocal profile analysis protocols for each subject, and for summating the group characteristics, were as described in section 2.1.2.

Results

The first part of this section will concentrate on the overall group results. The second part will look at male

and female results separately, since there are indications of some sex differences in the distribution of voice features. These may be the result either of organic differences or of sociolinguistic conventions.

A. Overall group characteristics

Figure 2.2/2 shows a summated protocol for the whole group of normal speakers, and Figure 2.2/3 shows the group means and standard deviations for each setting scale. These figures do not differentiate between continuous and intermittent adoption of a setting, but the only setting which was commonly scored as being intermittent is creakiness. Creakiness is quite often heard as a regularly occuring but intermittent setting, which is most marked on intonational falls at the ends of utterances.

Several interesting features are evident from these results. The most striking fact is confirmation of the impression that neutral is most certainly not synonymous with normal. Not one of this group of normal speakers exhibited a vocal profile which was neutral for all categories, and it is clear from the summated protocol (Figure 2.2/2) that within some categories the neutral setting is actually very rare amongst this population. This is especially true for categories 5 (velopharyngeal settings) and 10 (phonation type settings), where no speakers used a neutral setting. All 50 speakers were judged to have both nasal and whispery settings at at least scalar degree 2. In addition, only one speaker was judged to have a neutral tongue body setting, nearly 80% of speakers used creak at least intermittently, and more than half of these speakers had higher than neutral overall levels of muscular tension. This last observation is probably due to the subjects suffering a certain

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial	21	29		Lip Rounding/Protrusion	19	4	1			
				Lip Spreading	4	1				
				Labiodentalization						
	45	5		Extensive Range	2					
				Minimised Range	3					
2. Mandibular	26	24		Close Jaw	17	1				
				Open Jaw	5					
				Protruded Jaw	2					
	44	6		Extensive Range	1					
				Minimised Range	5					
3. Lingual Tip/Blade	22	28		Advanced	14	6	1			
				Retracted	5	2				
4. Lingual Body	1	47		Fronted Body	10	9	1			
				Backed Body	19	9				
		29		Raised Body	14	6	1			
				Lowered Body	6	2				
	48	2		Extensive Range						
				Minimised Range	2					
5. Velopharyngeal		48	2	Nasal		13	35	2		
				Audible Nasal Escape						
				Denasal						
6. Pharyngeal	25	25		Pharyngeal Constriction	18	7				
7. Supralaryngeal Tension	20	30		Tense	25	4				
				Lax	1					
B. Laryngeal Features										
8. Laryngeal Tension	14	36		Tense	22	10	2			
				Lax	2					
9. Larynx Position	27	23		Raised	6	4				
				Lowered	10	3				
10. Phonation Type		11		Harshness	5	6				
		48	2	Whisper(y)		33	15	2		
				Breathiness						
		36	3	Creak(y)	2	20	14	3		
				Falsetto						
		50		Modal Voice	50					

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

FIGURE 2.2/2: Summated Vocal Profile Analysis results for normal subjects : male and female

SETTING SCALE	MEAN SCALAR DEGREE	S.D.
Lip rounding/spreading	0.48 (rounded)	.91
Labiodentalization	0	0
Labial range	0.02 (minimised)	.32
Close/open jaw	0.28 (close)	.67
Protruded jaw	0.04	.20
Tip-blade	0.40 (advanced)	1.03
Fronted/backed T.B.	0.16 (backed)	1.46
Raised/lowered T.B.	0.38 (raised)	1.05
Lingual range	0.04 (minimised)	.20
Nasal/denasal	2.78 (nasal)	.51
Phar. constriction	0.64	.72
Supralar. tension	0.64 (tense)	.66
Laryngeal tension	0.92 (tense)	.90
Larynx position	0.04 (lowered)	.95
Harshness	0.34	.69
Whisperiness	2.38	.57
Creakiness	1.92	1.21

None of the speakers were judged to have audible nasal escape or falsetto

FIGURE 2.2/3: Table of mean scalar degrees and standard deviations for normal speakers (male + female)

degree of unease, engendered by the unfamiliar experience of sitting in a recording studio reading a set text.

A second feature of the summated protocol is the rarity of scalar degree judgements which exceed scalar degree 3. It seems that most speakers adopt habitual vocal patterns which are around the middle of their potential vocal range. For many settings this is probably related to the need for clear articulatory separation of phones. For example, the adoption of a long term average tongue body posture which is close to the centre of its potential range, means that during vowel production there is an equally wide span of movement possible along any radius away from its habitual setting. The articulatory separation of vowels is thus relatively easy. The adoption of a habitual tongue body setting which is close to the periphery of its range forces the tongue body position during the production of all vowels towards that peripheral position. The auditory result is that the vowels are all compressed within one part of the vowel area, and vowel separation, and hence intelligibility, is impaired. Extreme deviations from neutral may thus be communicatively inefficient in those categories where neutral represents the centre of the articulatory range. These include lip spreading and rounding, close and open jaw, lingual tip/blade settings, and tongue body settings.

The few judgements which deviate from neutral by as much as four scalar degrees are all for the three settings which show the greatest average deviation from neutral, i.e. nasal, whisperiness and creakiness.

B. Sex differences in vocal profile characteristics

Whilst the overall group characteristics give some indication of the typical vocal features found in a

population of English speakers, they do not allow any conclusions to be drawn about the possible relationship between voice quality and an individual's vocal anatomy. Since there are some well documented differences between the organic characteristics of the male and female vocal apparatus (see section 1.2.4), it seemed sensible to separate the vocal profile findings by sex, and to see if any vocal differences emerged which could be related to organic factors.

Figures 2.2/4-5 compare the summated protocols for males and females, and Figure 2.2/6 compares the mean scalar degree and standard deviation for each setting scale. The significance of any differences in means and standard deviations was tested using the Mann-Whitney U test (Siegel 1956:116-127), and the results of this comparison are also shown in Figure 2.2/6.

It can be seen that, although there are slight sex differences in quite a few of the vocal quality features, only four approach high levels of significance (i.e. P<.01). These are in the tongue body, pharyngeal constriction and phonation type categories. The tongue body is judged to be markedly more raised, and slightly more fronted in females. In other words, there seems to be a tendency towards constriction in the palatal region. Pharyngeal constriction and creak are more characteristic of the male group. Without further study of other accent groups, and detailed correlational studies linking vocal tract characteristics to vocal quality settings, it is not possible to assume that these differences are entirely due to organic sex differences. Sociolinguistic conditioning may be a powerful influence in the development of vocal quality differences between males and females. It is nonetheless possible that at least some of the apparent sex differences in supralaryngeal

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial	13	12		Lip Rounding/Protrusion	6	2				
				Lip Spreading	4					
				Labiodentalization						
	19	6		Extensive Range	2					
				Minimised Range	4					
2. Mandibular	16	9		Close Jaw	6	1				
				Open Jaw	2					
				Protruded Jaw						
	22	3		Extensive Range	1					
				Minimised Range	2					
3. Lingual Tip/Blade	11	14		Advanced	11	2				
				Retracted		1				
4. Lingual Body	1	22		Fronted Body	6	8				
				Backed Body	6	2				
		15		Raised Body	9	6				
				Lowered Body						
	24	1		Extensive Range						
				Minimised Range	1					
5. Velopharyngeal		25		Nasal		8	17			
				Audible Nasal Escape						
				Denasal						
6. Pharyngeal	21	4		Pharyngeal Constriction	4					
7. Supralaryngeal Tension	8	17		Tense	15	2				
				Lax						
B. Laryngeal Features										
8. Laryngeal Tension	7	18		Tense	13	4	1			
				Lax						
9. Larynx Position	17	8		Raised	1					
				Lowered	6	1				
10. Phonation Type		4		Harshness	2	2				
		23	2	Whisper(y)		14	9	2		
				Breathiness						
		16		Creak(y)	2	9	5			
				Falsetto						
				Modal Voice	25					

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

FIGURE 2.2/4: Summated Vocal Profile Analysis results for female subjects

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial	8	17		Lip Rounding/Protrusion	13	2	1			
				Lip Spreading		1				
				Labiodentalization						
	24	1		Extensive Range						
				Minimised Range	1					
2. Mandibular	10	15		Close Jaw	11					
				Open Jaw	3					
				Protruded Jaw	2					
	22	3		Extensive Range						
				Minimised Range	3					
3. Lingual Tip/Blade	11	14		Advanced	3	4	1			
				Retracted	5	1				
4. Lingual Body		25		Fronted Body	4		1			
				Backed Body	13	7				
		13		Raised Body	5					
				Lowered Body	6	2				
	24	1		Extensive Range						
				Minimised Range	1					
5. Velopharyngeal		23	2	Nasal		5	18	2		
				Audible Nasal Escape						
				Denasal						
6. Pharyngeal	4	21		Pharyngeal Constriction	14	7				
7. Supralaryngeal Tension	12	13		Tense	10	2				
				Lax	1					
B. Laryngeal Features										
8. Laryngeal Tension	7	18		Tense	9	6	1			
				Lax	2					
9. Larynx Position	10	15		Raised	5	4				
				Lowered	4	2				
10. Phonation Type		7		Harshness	3	4				
		25		Whisper(y)		19	6			
				Breathiness						
		23		Creak(y)		11	9	3		
				Falsetto						
				Modal Voice	25					

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

FIGURE 2.2/5: Summated Vocal Profile Analysis results for male subjects

SETTING SCALE	FEMALES		MALES		Difference between groups.	Significance of difference
	Mean Scalar Degree	S.D.	Mean Scalar Degree	S.D.		
Lip rounding/spreading	0·24 (rou)	0·83	0·72 (rou)	0·94	0·48	p = ·021
Labiodentalization	0·00	0·00	0·00	0·00	0·00	N.S
Labial range	0·008 (min)	0·49	0·004 (min)	0·20	0·004	N.S
Close/open jaw	0·24 (clo)	0·66	0·32 (clo)	0·69	0·08	N.S
Protruded jaw	0·00	0·00	0·08	0·28	0·08	N.S.
Mandibular range	0·04 (min)	0·35	0·12 (min)	0·33	0·08	N.S.
Tip-blade	0·52 (adv)	0·82	0·28 (adv)	1·21	0·24	p = ·115
Fronted/backed T.B.	0·48 (fro)	1·39	0·08 (bac)	1·26	0·56	p = ·010
Raised/lowered T.B.	0·84 (rai)	0·80	0·02 (low)	0·87	0·86	p < ·0003
Lingual range	0·004 (min)	0·20	0·004 (min)	0·20	0·00	N.S
Nasal/denasal	2·68 (nas)	0·48	2·88 (nas)	0·53	0·20	N.S.
Phar. constriction	0·16	0·37	1·12	0·67	0·96	p < ·0003
Supralar. tension	0·76 (ten)	0·60	0·52 (ten)	0·21	0·24	N.S.
Laryngeal tension	0·96 (ten)	0·79	0·88 (ten)	1·01	0·08	N.S
Larynx position	0·28 (low)	0·61	0·20 (rai)	1·16	0·48	p = ·056
Harshness	0·24	0·60	0·44	0·77	0·20	N.S
Whisperiness	2·52	0·65	2·24	0·44	0·28	p = ·092
Creakiness	1·40	1·19	2·44	1·00	1·04	p < ·0026

FIGURE 2.2/6: Table of mean scalar degrees and standard deviations for males and females and statistical analysis of sex differences (Mann Whitney U Test)

settings may result from organic differences in vocal tract proportions.

The tendency for maximum average constriction of the vocal tract to be heard as being further forward in the oral cavity in females may be related to sex differences in vocal tract proportions, although it is difficult to formulate convincing explanations for this, given available anatomical data. The higher ratio of oral cavity length to pharyngeal length in females (Fant 1966:22, see Section 1.2.4) might actually lead to the expectation that a constriction in an equivalent part of the oral cavity, say at the junction between the soft and hard palate, would be heard as a more backed tongue setting in females because there is proportionally less vocal tract length between the palate and the larynx. On the other hand, it may be that the tongue bears a different relationship to the palatal vault in females, such that there is an organic tendency towards palatal constriction.

The sex difference in phonation might well be due to organic differences, given the much larger size of the larynx in males and the rather different contours of the cartilaginous framework (see Section 1.2.4). A report of sociolinguistically conditioned differences in phonation within the male population (Esling 1978) means that cultural factors cannot, however, be excluded as possible causes. Anecdotal reports that creak is much more common in many populations of American women, for example, might indicate that the higher male incidence of creak is specific to our Scottish population sample.

The conclusion of this section must be that there is room for a much fuller investigation into the general trends of vocal quality differences between males and females across a much wider range of accents and languages.

Ideally, detailed anatomical measurements of individual speakers should be correlated with vocal output, but it is hard to envisage such a study being possible, since ethical considerations inhibit the widespread collection of radiographic data without good medical indications.

2.3 VOICE QUALITY IN DOWN'S SYNDROME

2.3.1 INTRODUCTION

Down's Syndrome (DS) is a genetic disorder which occurs in one out of every five to seven hundred live births (Benda 1969:4, Strome 1981). Individuals with DS display a constellation of physical and psychological anomalies. An objective study of voice quality in DS is desirable for several reasons.

It is of interest within the context of this thesis because DS results in a global disruption of growth and development, which often has quite marked consequences for the overall size, configuration and physiological state of the vocal apparatus. The ways in which the physical characteristics of the DS population differ from normal have been well documented, so that the DS group offers an opportunity to relate voice quality findings to existing knowlege about organic state.

Voice quality is also of interest within the broader context of DS research. The voice in DS seems to be sufficiently unusual to merit some comment in a large proportion of descriptions of the disorder. Unfortunately, most comments are rather subjective and impressionistic, so that interpretation and comparison of different studies is somewhat difficult. Some examples are listed below.

"Harsh" (Brousseau and Brainerd 1928)
"Hoarseness" (West et al. 1947)
"Severe voice problem" (Strazulla 1953)
"Raucous, masculine" (Benda 1960)
"Low pitched, harsh monotone" (Blanchard 1964)
"Guttural, low-pitched" (Fraser 1978)

There is also a lack of consensus about the incidence of voice problems. West et al. (1947) reported that "hoarse" voice was found in "most" DS children. Schlanger and Gottsleben (1957) estimated that 45% of institutionalised DS subjects suffer from some kind of vocal problem. Benda (1969:27,74) suggests that appropriate treatment of thyroid deficiency in DS may have reduced the incidence of voice disorders. The difficulty, with these and other studies, is that they are based on poorly defined notions about what constitutes a voice problem.

There have been attempts to make objective measures of some vocal features, but these are rather limited in scope, with the main emphasis being on fundamental frequency (F0). Figure 2.3/1 summarises the results of some studies which compare the F0 of normal and DS children. It seems that for children, at least, subjective reports of low pitch are not substantiated by acoustic measurements. The variability between studies may be due, in part, to different types of speech samples and F0 analysis procedures. Unfortunately there seems to be little comparable data available for adult DS speakers.

Spectrographic analysis has been used by Lind et al. (1970) to analyse DS infants' pain cries. They found them to have a lower fundamental frequency than normal, with abnormal temporal characteristics and a "stuttering" phonation. Spectrographic examination of speech in older DS subjects has been focussed on articulatory or phonological patterns. The results of such studies may, none the less, shed light on long-term vocal quality settings.

Jackson (1978) collected formant data for vowels produced by six 14-18 year-olds, in order to examine consistency and distinctiveness of vowel production, and the use of

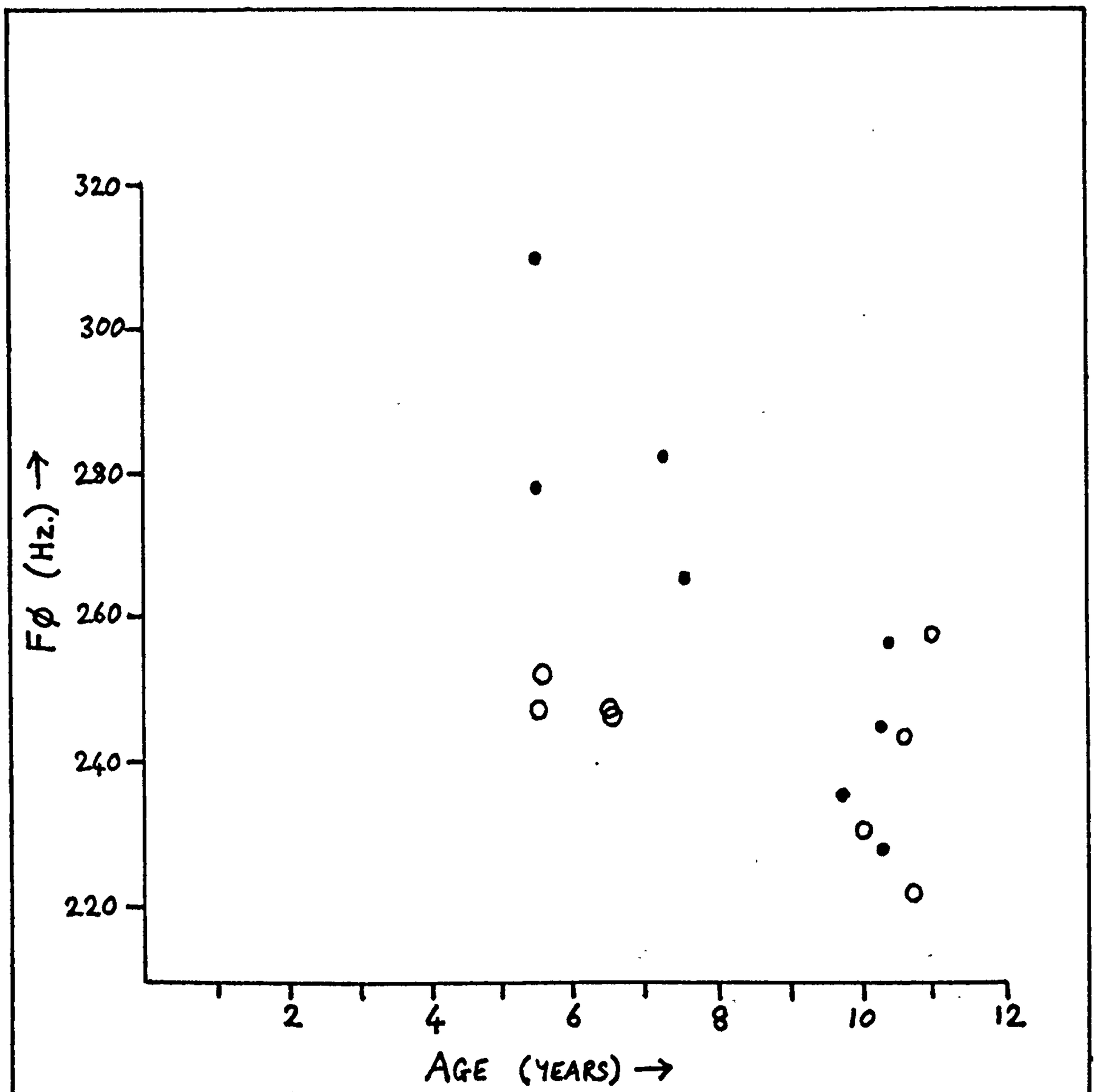


FIGURE 2.3/1: Graph of reported speaking F0 in DS and normal children

articulatory space. She found that these subjects tended to use a rather limited articulatory space, and that there seemed to be particular constraints on the production of high back vowels. The observation that these constraints were most marked in two individuals with small palatal volume measurements prompted the suggestion that the underlying problem could be an unfavourable tongue to palate size ratio. These findings are indicative of minimised articulatory range and fronted tongue body as long term vocal quality settings.

Listener response to the voice of DS children has also been investigated by several authors. Jones (1963, cited by Stoel-Gammon 1981:354) found that speech pathologists were able to discriminate between DS and non-Downs retardates on the basis of tape recordings alone. Montague and Hollien (1973) found that groups of naive listeners and speech therapists both judged tapes of 8-13 year-old DS children to exhibit more "breathiness" and "roughness" than normal children. Montague (1976) also looked at the ability of judges to assess age and sex of subjects from speech samples played backwards. He found that the "judged age" of the DS group was on average two years less than that of a sex- and age-matched group. Sex was less accurately judged for the DS group. Montague et al. (1978) used the same tapes of 8-13 year-old children to show that pitch was perceived as being lower, on average, in the DS group, but that the DS group also showed more variability between subjects. Moodie, Montague and Bradley (1978, cited by Stoel-Gammon 1981:346) used the Wilson Voice Profile Scheme to show that these same children had more deviations in pitch, more tension and more laryngeal air loss than normal.

Stoel-Gammon (1981:346) also describes work by Marriner (1980) which showed that judges were unable to discriminate between 0-18 month-old DS and normal infants

in terms of gutturality, intelligibility, speech-rate, breathiness or pitch. The discrepancy between these results and the results reported by Montague and colleagues may indicate that deviations in voice quality do not become apparent until some time after the age of 18 months. It is also possible that modern regimes of medication are succeeding in reducing the incidence or severity of voice disorders amongst DS children.

The dearth of objective studies of voice quality in DS adults is unfortunate. Deviations in voice quality seem to have a profound influence on social acceptability of speech, and it would be interesting to know how much vocal problems handicap the individual with DS. Naive listeners seem to be prepared to make far-reaching judgements about an individual, including personality traits and social or educational status, on the basis of voice quality alone. A pilot study by Saville (1983) has shown that two college students with vocal fold palsies were judged to be older than matched controls, and were given consistently lower ratings for intelligence, competence, dominance, extroversion and vitality, on the basis of tape-recorded speech samples. This kind of study highlights the possibility that if voice quality in DS is, in fact, abnormal, then adverse listener response may compound the effects of linguistic and articulatory incompetence.

Knowledge of the voice quality characteristics of DS might therefore be of value to all professions involved in the care of individuals with DS. An awareness of the risk of false attribution of psychological traits as a response to voice quality could help to minimise negative listener response.

Some work by Leudar et al. (1981) may be interpreted as suggesting that voice quality limitations of organic

origin may actually inhibit communication more in DS than in non-Downs individuals. Leudar found that DS subjects, when faced with familiar or unfamiliar interlocutors, tended to alter their non-verbal behaviour, whilst using very similar verbal structures. Non- Down's subjects, in contrast, seem to differ in linguistic output more than in non-verbal behaviour when interacting with people of differing degrees of familiarity. The experimental conditions were slightly different for the DS and the non-Down's groups, but the results may be interpreted as showing that the DS group relied more heavily on non-verbal behaviour to communicate familiarity. If this reliance on non-verbal cues is a general feature of communication by DS subjects, then impairment of any non-verbal channel, such as voice quality, would be a double handicap. It could interfere both with the listener's perception, and with the speaker's ability to communicate discriminately.

With these considerations in mind, the possibility of voice remediation must be considered, and this links back to the relationship between voice quality and organic state. An understanding of the extent to which a given deviation in voice quality is constrained by organic abnormality is essential if a speech therapist is to assess the extent to which therapy can be expected to improve matters.

2.3.2 ORGANIC CHARACTERISTICS OF DOWN'S SYNDROME

The primary object of this section is to outline the features which may be said to typify adult DS subjects, and which might influence the organic state of the vocal tract. A part-by-part account of the DS vocal tract is prefaced by a discussion of genetic factors, and some general comments on variability, and overall developmental and structural trends in DS.

2.3.2.1 Genetic Factors

The basic genetic make-up of human cells, and the pattern of chromosome replication, have already been described briefly in sections 1.1.2 and 1.2.1..At certain stages during the cycle of cell division, chromosomes from a cell can be stained, and then examined microscopically. Each chromosome has a characteristic pattern of light and dark bands, so that chromosomes can be counted and identified. The typical chromosome complement in humans is 46. This means that each cell in the body contains 23 pairs of chromosomes, with one member of each pair being derived from each parent. The only exceptions to this rule are the sex cells (cells which develop to form ova or spermatazoa), which possess only one member of each chromosome pair.

In 1959 it was discovered that DS is associated with the presence of one additional chromosome (Lejeune et al. 1959). It is now known that the usual chromosome complement in DS is 47, with three copies of chromosome number 21 instead of the normal two. The possession of three copies of any chromosome is known as trisomy, so these individuals are trisomic for chromosome 21. Occasionally only two copies of chromosome 21 are present in DS, but in these cases there is usually one chromosome which is larger than normal, and which can be shown to have a reduplicated section of chromosome 21 attached to it. The end result is the same; there are three copies of at least part of chromosome 21 (see Figure 2.3/2).

In the majority of cases of DS all cells in the body can be shown to have an abnormal chromosome complement. This would be the case if the chromosome abnormality is present in the fertilized egg, and is perpetuated at each cell division. In a few cases it is found that the chromosome abnormality is present in only some of the

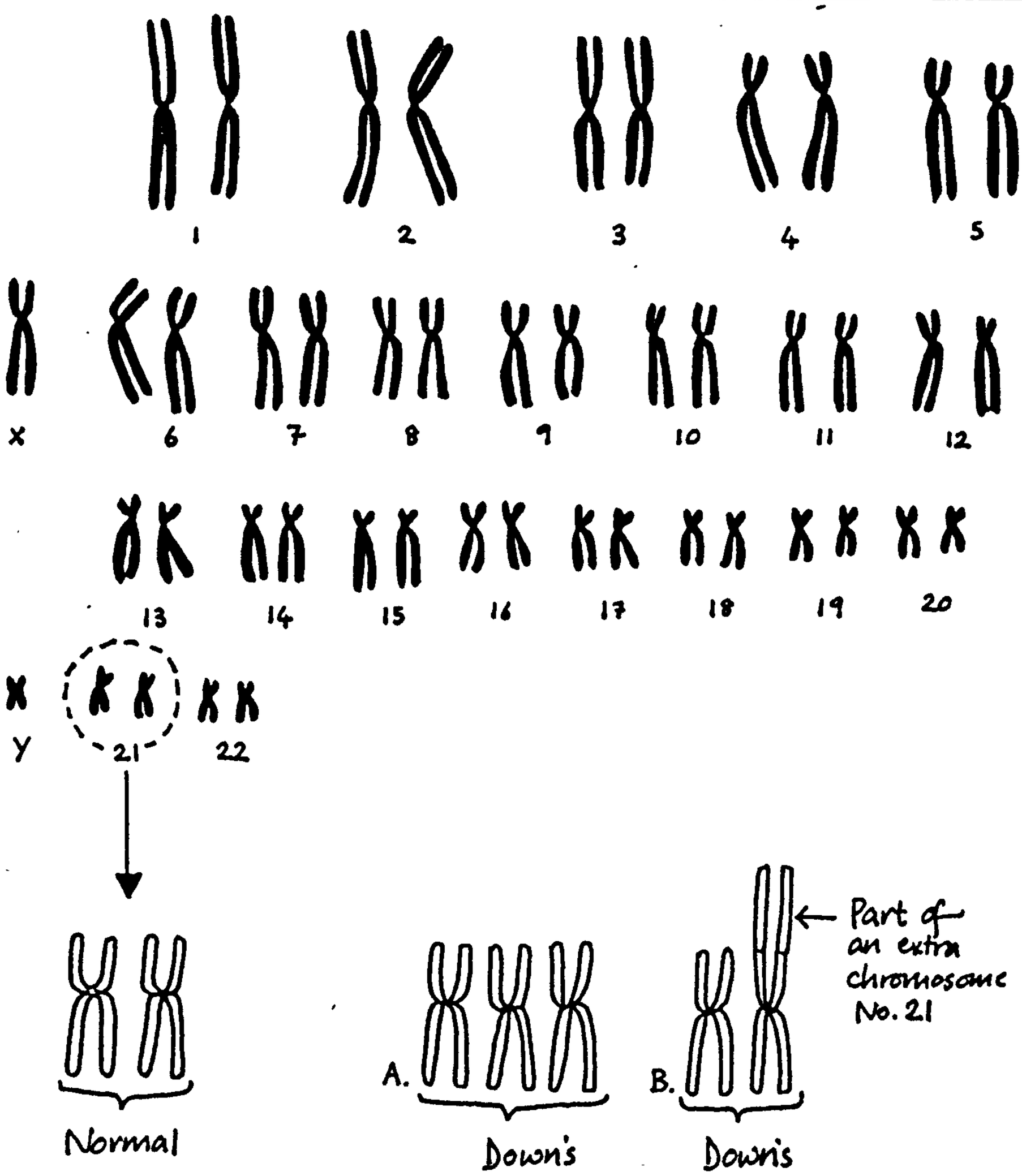


FIGURE 2.3/2: Normal human chromosome complement, arranged in pairs, and schematic representation of DS variants
A. Simple trisomy
B. Translocation

cells of the body, whilst the remaining cells are normal. These cases are known as mosaics, and are presumed to be the result of some fault in cell division during embryo development which results in a single cell with the DS chromosome complement of 47. As development progresses, all the descendant cells of this abnormal cell will also have the typical DS chromosome complement, whilst the rest of the cells develop normally. The proportion of Down's-type cells will depend on the site and timing of the faulty cell division, and will have some influence on the severity of the handicap experienced by the mosaic individual.

The mechanism by which the presence of an extra chromosome disrupts development in such a way that a pattern of physical anomalies and mental handicap ensues is not well understood. A simplistic explanation in terms of the specific affects of a triple (rather than double) dosage of the genes carried on chromosome 21 may be partially true, but is clearly not wholly adequate. Firstly, it fails to explain the unusually high variability in DS (see following section). Secondly, it makes it rather surprising that the same physical anomalies may be characteristic of several different chromosome disorders. In Goodman and Gorlin's descriptions of chromosome disorders (1970), some features seem to recur quite often. Microcephaly is said to be characteristic of syndromes involving partial or whole additions of 6 different chromosomes. An increased incidence of cleft lip and/or palate is found in 13 of the disorders listed. Congenital heart disease is found in 5 of the disorders involving chromosome additions, and several other features are also found in association with more than one chromosome disorder. These findings suggest that the presence of additional genetic material may well cause a rather general disruption of development, in addition to any specific gene effects.

The normal genetic make-up of many organisms seems to incorporate a complex system of compensatory processes, which act as buffers against environmental effects and minor genetic variations, so that development is canalized along a fairly narrow course (Waddington 1957). As a result, most physical characteristics vary within a fairly narrow range in the normal population. It may be that the additional presence of a whole chromosome unbalances the buffering processes, so that the efficiency of canalization is reduced. This would make development in DS more susceptible to disturbance by environmental and endogenous factors. This would explain the increased incidence, in DS and other chromosome disorders, of a variety of organic abnormalities which occur only rarely in the normal population. An abnormality which results only from rather major environmental influences on normal development might be caused by much milder environmental effects in DS. Shapiro (1970) develops this theme, and suggests that DS is characterised not by a particular pattern of organic abnormality, but by what he calls an amplified instability of development.

It seems likely, from the observation that different chromosome disorders do differ, whilst sharing some physical features, that an adequate explanation of DS must involve some specific gene dosage effects superimposed upon a more generalised instability of the canalization process.

2.3.2.2 Organic variability in Down's Syndrome

To many lay people the term "syndrome" may imply a condition which displays a relatively invariable set of signs and symptoms. In the case of DS, this impression may be strengthened by reading early descriptions of the

disorder, which stress the similarity between individuals (e.g. Down 1866). The reality is, as usual, more complex.

It is true that there are some organic features which occur so much more commonly in DS than in the normal population that they can be described as being, in some sense, typical of DS. It is not, however, possible to specify a constellation of physical features which is present in all cases of DS, and absent in the rest of the population. This much is made clear by an examination of various texts on clinical diagnosis (for reviews see Benda 1969, Smith and Berg 1976). The great variability in the occurrence of so-called DS characteristics was highlighted by Levinson et al. in 1955, and the implications of this for clinical diagnosis are now widely recognised. Shapiro (1973) summarises the problem as follows:

> "No unique physical abnormalities occur in Down's Syndrome. Rather, it is the frequency, intensity and multiplicity of anomalies that are characteristic."

Variability in what Shapiro calls "intensity" of anomalies is also increased in DS. This can be seen in the higher than normal standard deviations or variances which are found for many physical parameters. These include stature and skeletal maturation (Roche 1965), tooth width and various craniofacial measurements (Kisling 1966), and age of tooth eruption (Shapiro 1970).

In spite of this, it is possible to show statistically significant differences between Down's and non-Down's populations in the means of many parameters. It is also possible to show significant differences in the frequencies of occurrence of many abnormalities. The assumption that a group of DS individuals is likely, on

average, to differ from a control group in specifiable ways does, therefore, seem justifiable.

One other point relating to variability needs to be considered in interpreting reports of organic anomalies in DS. It is important to note that the majority of studies quoted do not specify whether or not the diagnosis of DS in the subjects concerned was confirmed by cytogenetic (chromosome counting) techniques. Since it is only in recent years that chromosome studies have become commonplace, and routine chromosome screening of the newborn has been confined to a few geographical areas (Emery 1979), it is reasonable to assume that a large proportion of studies have not used cytogenetic confirmation of DS. This allows the possibility that the increased variability of DS may introduce some bias into the results. If diagnosis relies on the presence of physical signs, individuals who display a high number of those signs are more likely to be given a firm diagnosis, and are therefore more likely to be included in studies on DS. Individuals who are trisomic for chromosome 21, but who display fewer clinical signs, are more likely to escape diagnosis, and would therefore be excluded from studies of DS. The effect of this may be to bias results towards the clinical stereotype of DS.

The magnitude of such biasing is impossible to assess, and it is probably not great except in the earliest studies. Unfortunately, its significance is likely to be magnified where orofacial features are concerned, because of the reliance on facial characteristics in clinical diagnosis. Six out of the ten most characteristic signs chosen by Øster (1953), for example, concern the facies, and the importance of facial characteristics in diagnosis is repeated in other texts (Hall 1964, Benda 1969:11-19, Smith and Berg 1976:156).

2.3.2.3 General Developmental Tendencies

It is possible to abstract some general tendencies of growth and development in DS. General, in this context, means that they affect the whole body, or some large part of the body, rather than being limited to any particular part of the vocal tract.

Stature

The correlation between stature and vocal tract length is not clear, even for the non-Down's population (Bristow 1980), but intuition suggests that there may be some relationship. Stature in DS shows more variability than it does in the rest of the population, but most studies show a reduction in mean height (Øster 1953, Roche 1965, Smith and Berg 1976). These studies are summarised in Figure 2.3/3. Smith and Berg suggest an average adult height of 151 cm. in males, and 141 cm. in females, which is significantly lower than normal. There are some indications that hormone imbalance may be partly responsible for the growth deficit. Benda (1969:244) made a ten year study on the effect of pituitary-thyroid treatment on DS children, and found that 67% of treated children fell within the normal height range, compared with only 28% of untreated children.

Craniofacial development

In the newborn DS infant, head measurements are usually within normal limits (Benda 1960), but the proportions of the head seem to be somewhat abnormal, and this abnormality becomes more marked later in childhood. The outstanding feature is brachycephaly, where the anterior-posterior dimension of the head is small, relative to the width. In DS there is a very marked reduction in anterior-posterior measurements of the skull, with only a

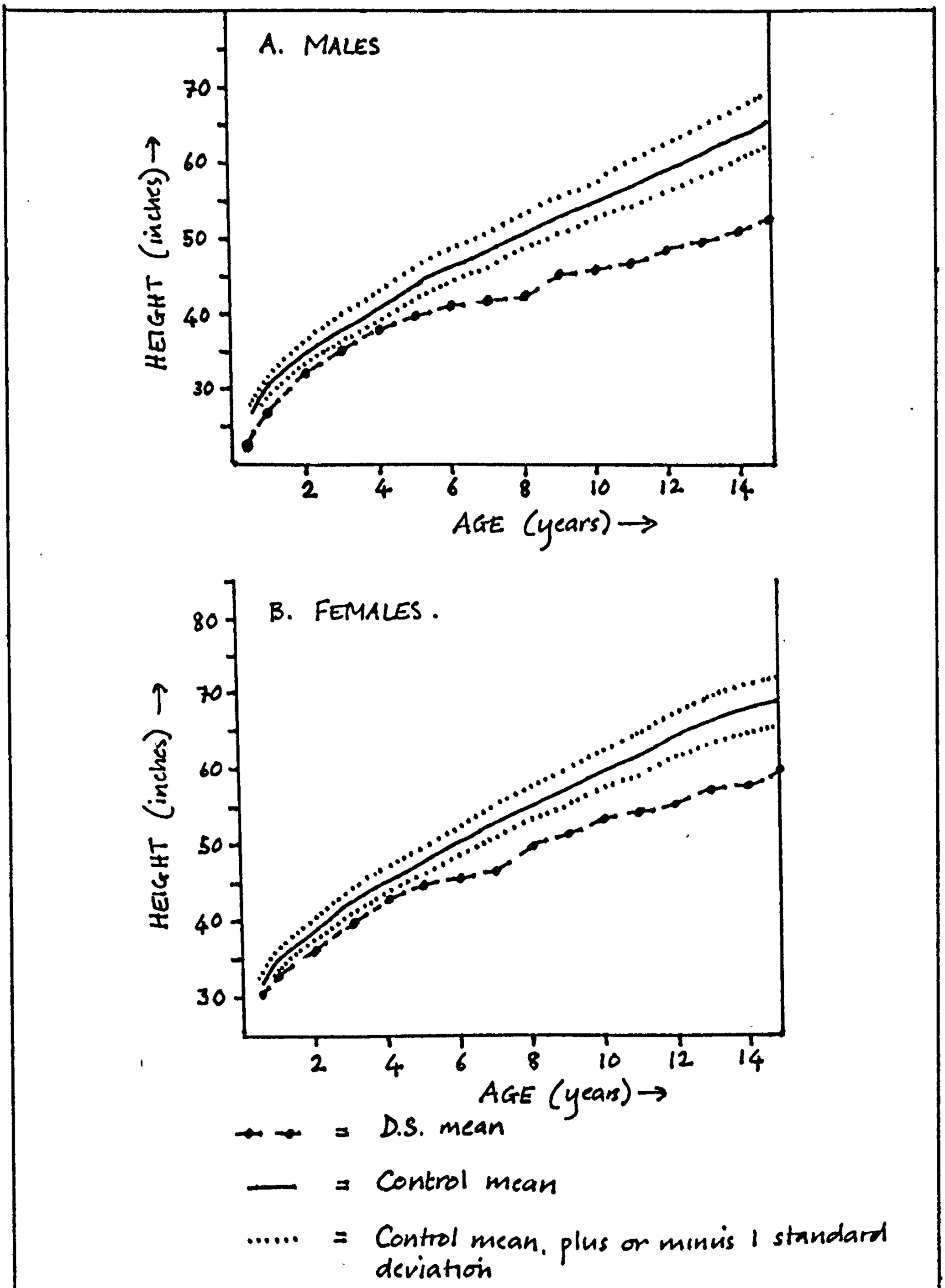


FIGURE 2.3/3: Normal and DS height growth curves (adapted from Thelander and Pryor 1966)

slight reduction in width. The whole of the midface and maxillary region tends to be rather under-developed (Gosman 1951, Penrose 1963, Kisling 1966, Thelander and Pryor 1966, Frostad et al. 1971, Smith and Berg 1976:44-50). Figure 2.3/4 shows tracings of lateral X-rays of normal and DS skulls, adapted from work by Baer and Nanda (1975:533). Figure 2.3/5 summarises some cranial measurements in DS and normal subjects.

Epidermal and mucosal structure

A variety of epidermal disorders have been reported to be common in DS (Smith and Berg 1976:38), and xerosis (dryness) has been reported in up to 90% of cases. It seems likely that the mucosal lining of the vocal tract may also be abnormal in many cases. Novak (1972) reports atrophy and dryness of the pharyngeal mucosa in DS, whilst Smith and Berg (1976:20) comment on thickening of the nasal mucosa. Fissuring of the tongue and lips (see section 2.3.2.4) are also indicative of abnormality. These dermatological and mucosal problems may be due partly to minor histological disorders, and partly to the influence of the fluid bathing the tissues. The chemical composition of saliva is said by Winer et al. (1965) to be unusual, and the same authors found an abnormally slow rate of salivary flow from the parotid gland. Hypothyroid states (see below) may also alter the fine structure of the mucosa.

Hormonal Factors

There is considerable controversy over the frequency, severity and type of hormonal disorders found in DS. Early clinical descriptions describe thyroid inadequacy in DS, which is unsurprising given that the syndrome was at one time confused with cretinism (congenital thyroid deficiency). The administration of thyroid hormone to DS

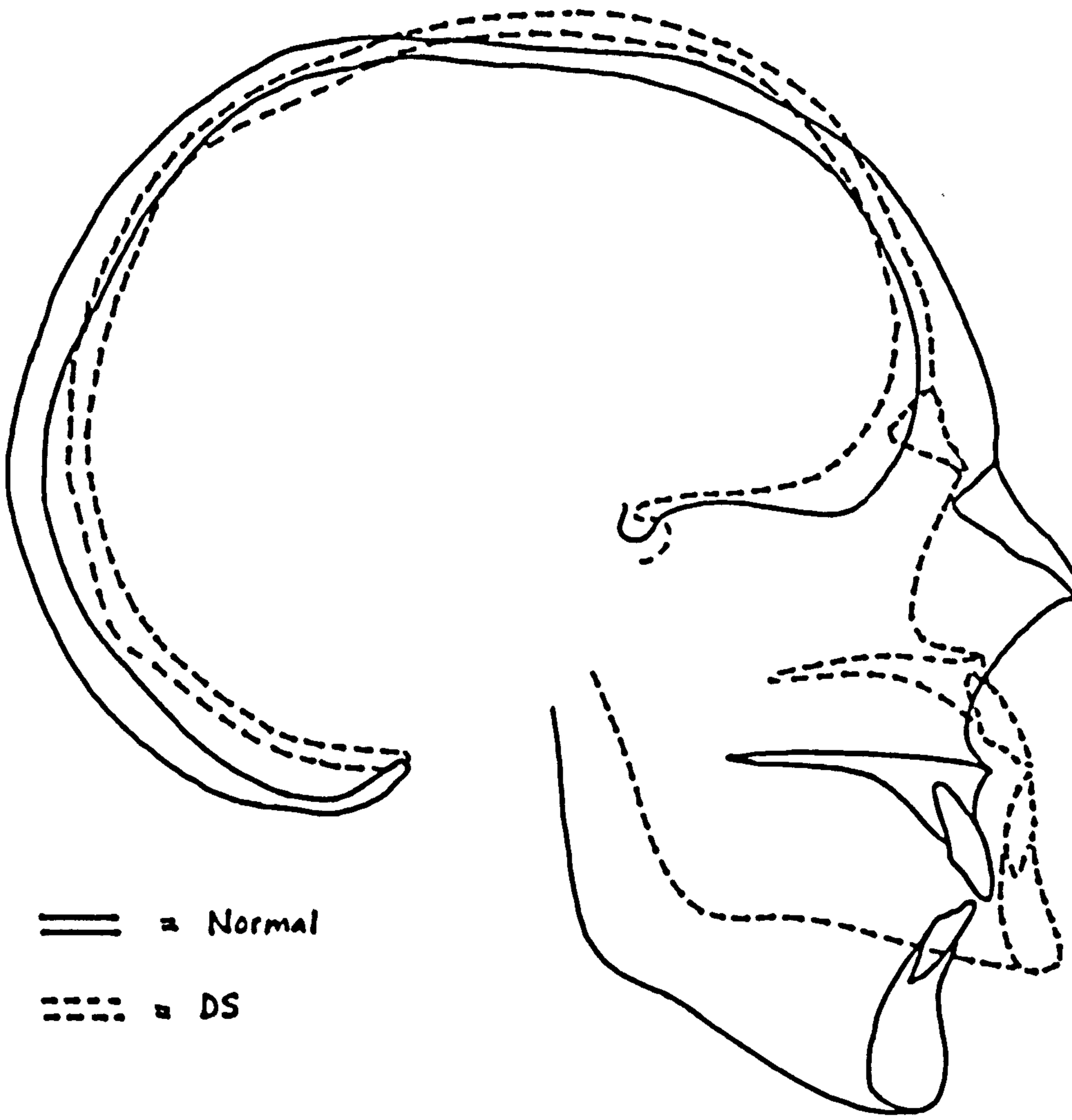


FIGURE 2.3/4: Tracings of lateral skull X-rays for normal and DS adults (adapted from Baer and Nanda 1975:533)

	Cephalic breadth (i)	Cephalic length (ii)	Cephalic height	Cephalic index (i/ii)
DS Subjects	142.5 mm.	174.6 mm.	125.7 mm.	0.82
Controls	152.5 mm.	193.7 mm.	134.6 mm.	0.78

FIGURE 2.3/5: Summary table of DS and normal cranial measurements (data from Penrose 1963)

individuals has therefore been common for many years. Benda (1969:166ff.) believes that both thyroid and pituitary inadequacy are common in DS, and reports histological evidence for abnormality in both thyroid and pituitary glands. He claims that the administration of pituitary and thyroid supplements may produce quite marked improvements in development and growth in DS. Of particular relevance here is his assertion that the typical "harsh" voice of DS is rarely found in patients who have undergone thyroid treatment over a long period (Benda 1969:71).

These beliefs are not, however, shared by other authors. Smith and Berg (1976:41) state that although clinicians often diagnose thyroid or pituitary dysfunction there is actually little evidence of endocrine gland abnormality. They cite a literature survey by Hayles et al. (1965) which discovered definite reports of only 4 cases of primary hypothyroidism and 15 cases of hyperthyroidism. According to these authors, most DS individuals seem to have normal thyroid function, and the commonest thyroid problem is hyperthyroidism rather than thyroid inadequacy. Smith and Berg (1976:266) cite a study by Koch et al. (1965) which found that the administration of thyroid hormone had no affect on linear growth, developmental quotient or general clinical status. Similarly, Berg et al. (1961, cited by Smith and Berg 1976:266) failed to substantiate Benda's reports of beneficial results following administration of pituitary extract.

Hormonal characteristics of DS are thus somewhat uncertain. It may be that even if endocrine excretion is normal in DS, the response to hormones is somehow abnormal.

Muscular hypotonia

It is generally accepted that one of the commonest signs of DS is generalised muscular hypotonia. Estimates of the incidence vary from 66% (Levinson et al. 1955) to 97.7% (McIntyre and Dutch 1964), but it seems that some degree of muscular hypotonia is likely to be present in the majority of the DS population. This has wide ranging implications for development and function of many systems in the body. Development may be affected because hypotonia will result in distortion of the normal mechanical forces acting on the skeleton and soft tissue, both as a direct result of reduced muscle tone, and as a secondary result of abnormal posture.

Several factors may be involved in the aetiology of hypotonia in DS. Crome et al. (1966) found a reduction in brain stem and cerebellar weight in DS, and suggested that hypotonia might be a consequence of an anatomical, neurological deficit. Hypothyroidism has also been implicated (Benda 1969). There is some justification for this, since hypothyroidism in otherwise normal individuals is often associated with a reduction in muscle tone, but there is some dispute about the real incidence of hypothyroidism (see above). Evidence that the cause may be at least partly biochemical comes from the observation that medication may alleviate the problem (Benda 1969).

2.3.2.4 Specific Abnormalities of the Vocal Tract

The various parts of the vocal tract will now be discussed in turn, beginning with the main airflow generators, the lungs.

Lungs

The gross structure of the lungs seems to be fairly similar to normal, with lung malformations occurring rarely (Benda 1969:208, Smith and Berg 1976:37). DS is, however, associated with a marked susceptibility to respiratory infections, which may be exacerbated by minor abnormalities in the respiratory mucosa. Pulmonary hypertension is not uncommon, and has been attributed to increased respiratory stress resulting from congestion of the upper airways.

It does seem likely that the efficiency of thoracic activity, both for respiration and for speech, may be lower than normal because of the generalised muscular hypotonia and poor posture. It is unfortunate that data on lung volumes in DS is not readily available.

Larynx

Few studies of the larynx have been reported. Benda (1969:27) examined the larynx in a small number of cases, and formed the impression that the larynx was higher in the neck than normal. It is not clear whether "higher", in this case, means that it is higher relative to other structures of the neck (e.g. the horns of the hyoid bone), or that it is less distant from the oropharynx. This is particularly difficult to interpret in the light of reports that the neck in DS is unusually short and broad (Øster 1953, Levinson et al. 1955, Benda 1969:31, Smith and Berg 1976:33).

Perhaps more significant is Benda's finding that the laryngeal mucosa in the cases he examined appeared thickened and fibrotic. Novak (1972), on the other hand, examined 32 DS subjects in the age range 7 to 19 years, and found only a "light thickness" of mucosa, and no

cases of thickened vocal folds. This apparent discrepancy in findings may be due to the small numbers studied, or to an age difference in the two groups studied; Benda does not specify the age of his subjects.

Pharynx

Few studies seem to have considered the configuration of the pharynx in any detail, although there are some references to the fact that the pharyngeal airway in DS children is often constricted by an excessive tonsillar mass (e.g. Ardran et al. 1972). Strome (1981), however, found that the mass of tissue removed during tonsillectomy in DS children was actually similar to or smaller than normal, even where visual inspection suggested an increased mass. He suggests that this is because the pharynx is narrowed at the level of the faucial pillars, and the tonsils are seldom recessed behind the pillars. Strome also observes that the nasopharynx of the children he examined was markedly narrowed in the anterior-posterior dimension, with some lateral compression. The only measurement relating to adults found in the literature concerned a single male DS subject (Rolfe et al. 1979). Measurement of the nasopharynx based on cinefluorography showed the depth of the nasopharynx in this individual to be 20 cm., compared with norms for his age of 24.2 cm. Although it would be dangerous to make generalizations about all adult DS cases from a single case, anterior-posterior compression would be consistent with the general reduction in this dimension of the cranial skeleton. It would not, therefore, be surprising to find some anatomical constriction of the pharynx persisting into adulthood as a general feature.

Oral cavity

The size and configuration of the oral cavity is largely a product of the relationship between the palatal contour and the tongue.

a. The palate

Nearly all descriptions of the orofacial characteristics of DS remark on the high incidence of abnormal palatal size and contour. Brousseau and Brainerd (1928) describe the palate as being "generally high, narrow, V-shaped or vaulted". Øster (1953) also uses the adjectives "high" and "narrow". Levinson et al. (1955) report "high, arched palate" in 74% of the 50 DS subjects (0-17 years) they examined, and "narrow palate" in 52%. Novak (1972) uses the term "gothic" to describe the palate in 20 out of the 32 children in his study. Anterior-posterior length is less often mentioned in the early literature, although Engler (1949) did suggest that the DS palate might be shorter than normal. All such reports, being based on non-metrical clinical observation, are difficult to interpret.

The first extensive metrical data was provided by Shapiro and his associates (1967), who developed instrumentation for the direct measurement of various palatal dimensions. This allowed them to study a large group (153) of DS subjects, ranging from 6 years to adulthood, and to compare these to a normal control group. Their results do not substantiate earlier reports of increased palatal height in DS. In fact, they found a small, though statistically insignificant, reduction in palatal height relative to normal. The mean width, however, was significantly reduced in the DS group, at all ages and in both sexes. The most dramatic difference between DS and control subjects was in anterior-posterior length. This

was so much shorter in DS subjects that in most cases palatal length alone was enough to differentiate DS and normal palates. These results were confirmed by Jensen et al. (1973) and by Westerman et al. (1975), who further confounded earlier reports of high palate in DS by finding a statistically significant reduction in height relative to normal.

Austin et al. (1969), in a roentgenographic study of palatal length in 10 newborn DS infants, also found a significant reduction in palatal length relative to normal, so that it seems that this characteristic is typical of DS from an early stage in development.

In summary, the DS palate seems on average to be distinctly shorter than normal, with a lesser, though significant reduction in width, and possibly also in height. These differences are found in all age groups, and are in keeping with the general trend of reduced maxillary development (see below) and brachycephaly. A summary is given in Figure 2.3/6.

The subjective impression of narrowness allied with increased height may result from the unusual palatal contour which is found in some DS individuals. Shapiro et al. (1967) observed that many of their subjects had what they describe as a "steeple-shaped" palate. This type of palatal contour, where a level shelf extends inwards from the alveolar process, and the palate then rises sharply towards the midline (see Figure 2.3/7) is rare in the normal population. It probably corresponds to some of the terms, such as "gothic" and "vaulted", which feature in the earlier literature. Benda (1969:12-13) links the palatal contour to underdevelopment of the bones connected to the nasal cavity.

	Shapiro et al. 1967 DS (N=98 males, 55 females)	Westerman et al. 1975	
		DS (N=40)	Controls (N=44)
Width	DS cases fall round a line 2 SD below normal mean. Only 7.1% males and 3.6% females wider than normal means.	28.79 ±.51	32.27 ±.47
Length	95.9% males and 94.5% females more than 2 SD below normal means. No cases longer than normal mean -1 SD.	28.97 ±.55	31.12 ±.50
Height	Palatal height does not appear abnormal	12.27 ±.35	15.13 ±.32
		All DS / control differences significant at $p < 0.01$	

FIGURE 2.3/6: Summary of reported differences in palatal dimensions between DS and normal subjects (children and adults)

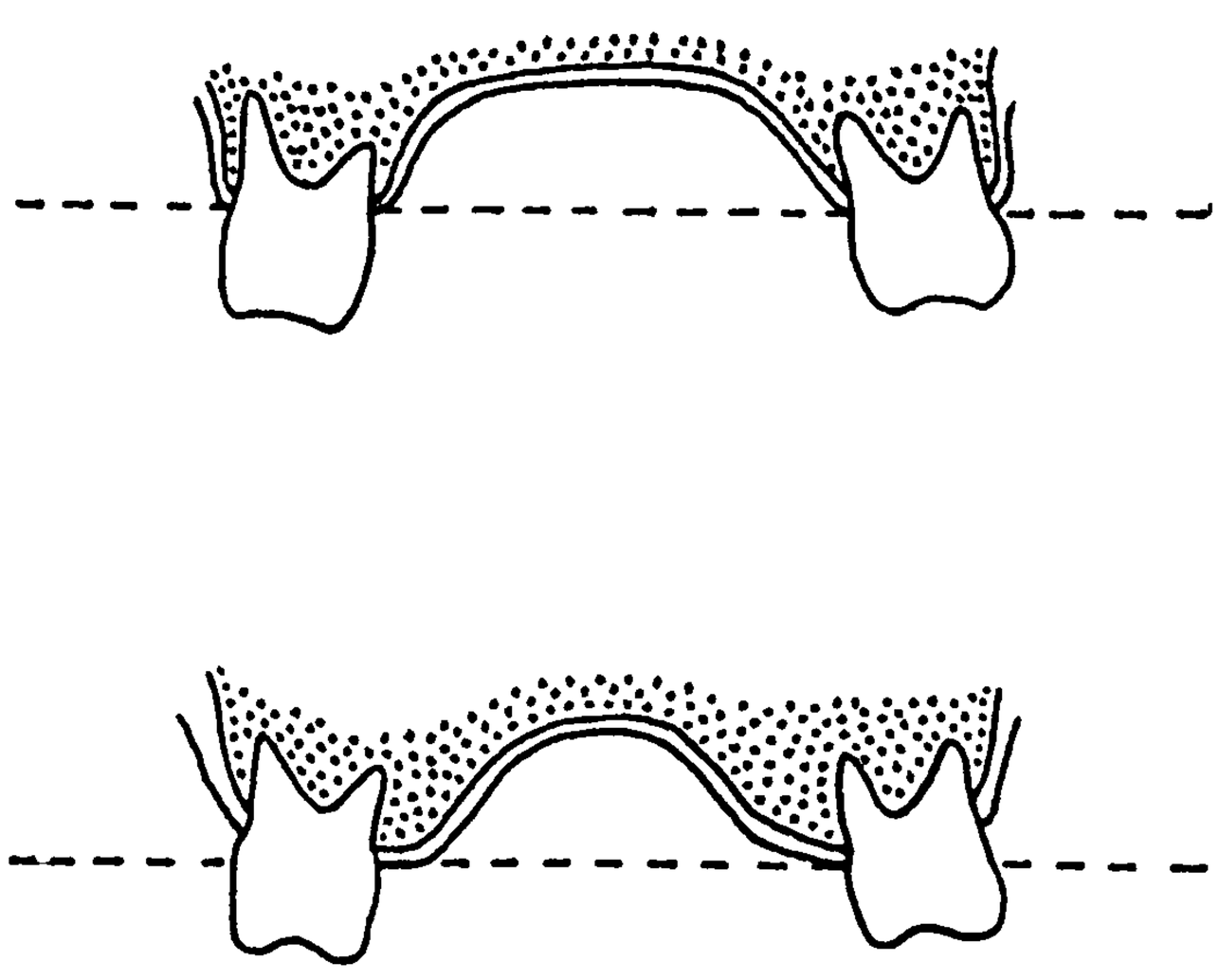


FIGURE 2.3/7: Diagram of normal palatal contour and "steeple" palate (adapted from Shapiro et al. 1967:1462)
This diagram represents coronal sections at the level of the maxillary first permanent molars.

Smith and Berg (1976:15) cite studies which show that the incidences of cleft palate and cleft lip (0.5%), submucous cleft of the palate (0.8%) and bifid uvula (4.6%) are all higher than normal. The figures are not high enough to be described as in any way typical of DS, but they do perhaps highlight the susceptibility of the maxillary area to malformation in response to genetic or environmental disturbance. In relation to this it is interesting to note that of 23 chromosome abnormalities described by Goodman and Gorlin (1970), 13 are associated with higher than normal incidences of cleft lip and/or palate. It does seem that the complex coordinate growth of the midface and palate may be espescially prone to disruption by chromosome imbalance of various kinds.

b. Tongue size and posture

Macroglossia and tongue protrusion have been commonly cited characteristics of DS for many years. Øster (1953), for example, reported overlarge tongue in 57% of his cases. Levinson et al. (1955) noted "large" tongue in 30% of cases, and "tongue protrusion" in 32%. This is controversial, however, and other writers claim that true macroglossia is in fact rare, but that the tongue may appear large in relation to the small oral cavity (Brousseau and Brainerd 1928, Benda 1969:27, Cohen and Winer 1965, Cohen and Cohen 1971). Resolution of this controversy is inhibited by the notorious problems of measuring tongue volume. The plasticity and mobility of the tongue limit the value of two dimensional representation, but this is usually the only feasible basis for objective measurement.

Ardran et al. (1972) made lateral radiographs of eight children, and found that this sample failed to substantiate reports of large tongue size. The tongues looked rather flat in profile, and none filled the oral

cavity, or protruded beyond the lower incisors. The lingual tonsils tended to look rather larger than normal, and five of the subjects appeared to have a local enlargement of the tongue in the tonsillar region. This apparent enlargement might, presumably, be an artefact of the two dimensional pictures. If the tongue is compressed laterally by the tonsils, it is likely to distort so as to appear larger in the vertical dimension. On the basis of these results, these authors conclude that the forward displacement of the tongue may be a response to a narrowing of the pharyngeal airway by the tonsils and adenoids.

Unfortunately there seems to be no comparable data for tongue size and posture in adults with DS, and the controversy about tongue size in children continues. Lemperle et al. (1980), commenting on the much disputed practice of plastic surgery to the orofacial region in DS infants, describe 63 out of 67 cases as having macroglossia which merited surgical correction. The results of this type of surgery do not, as yet, seem to have been evaluated objectively in terms of articulatory skills or general sensory and motor abilities.

In summary, it is not possible to make categorical statements about the incidence or severity of macroglossia in DS, if indeed it occurs. Some factors may, however, be called upon to explain why clinical judgements of large tongue size may be made in the absence of true macroglossia. These are:

1. hypotonicity of the lingual musculature. The prevalence of generalized hypotonia has already been discussed (see section 2.3.2.3). Engler (1949) commented specifically on lingual hypotonia, and it has been suggested that a lax tongue will tend to fall forwards in the mouth, and may protrude abnormally (Ardran et al.

1972). This would certainly be predicted if an overall laxness of the articulators resulted in a rather open jaw position, so that the tongue would tend to fall down a forward incline of the jaw.

ii. protrusion or disproportionate growth of the mandible relative to the maxilla (see below). A tongue which is normally positioned and proportioned in terms of the mandible would then tend to be carried forward in relation to the palate and upper teeth.

iii. forward displacement of the tongue to maintain an adequate airway in the presence of skeletally derived pharyngeal constriction and/or enlarged tonsils.

c. Tongue / palate relationship

The consequences of a short, narrow palate and a possibly large, forwardly displaced tongue are likely to include a constriction of the front of the oral cavity. The highest point of the tongue is likely to be closer than normal to the front of the palate or to the alveolar ridge. If the tongue is disproportionately large in relation to the palatal volume, then the whole of the oral cavity will tend to be narrower than normal in cross section.

d. Tongue morphology

"Scrotal" fissuring of the tongue and papillary hypertrophy (i.e. excessive growth of the tongue's papillae) are common findings (Benda 1969:26, Cohen and Cohen 1971, Smith and Berg 1976:15). Thomson (1907, cited in Smith and Berg 1976:16) found the tongue to be normal at birth, but with fissuring developing occasionally as early as 6 months of age. Engler (1949) believed that all DS subjects develop tongue fissuring by 5 years of age, but this has been disputed. Figures of 59% (Øster 1953),

44% (Levinson et al. 1955) and 37% (Cohen and Winer 1965) have been given as estimates of the incidence of tongue fissuring. Engler also found papillary hypertrophy from about 2 years of age, but Øster points out that this very often accompanies fissuring and is difficult to differentiate from it.

Jaw relationships and dentition

The size of the mandible in DS seems to be fairly close to normal, but the maxilla, as discussed in section 2.3.2.3, is underdeveloped. The result of this is a pseudo-prognathism, in which the mandible appears over-large, and the mandible protrudes relative to the maxilla. This is reflected in Brown and Cunningham's (1961) study of occlusion in DS, which showed that 64% of DS subjects over the age of 11 years had an Angle's class III malocclusion (i.e. the mandibular dental arch is anterior to the maxillary arch). Kisling (1966) obtained similar figures, and although Cohen et al. (1970) found that only 22% of children over 16 years had a class III malocclusion, this lower figure is still considerably higher than would be expected in the general poulation.

There have been many studies of the teeth and gums in DS, and reviews can be found in Shapiro (1970), Cohen and Cohen (1971) and Smith and Berg (1976). The major finding seems to be a high degree of variability. Teeth often show a delayed and deviant pattern of eruption, and there are high incidences of anomalous, misplaced and absent teeth. Periodontal disease appears to be a particular problem in DS, although it tends to be less severe in non-institutionalised cases (Swallow 1964), possibly because dental hygiene is more easily supervised. The incidence of dental caries, in contrast, has been said by some authors to be unusually low (Brown and Cunningham

1961, Winer and Cohen 1962). This may be partially explained by delayed tooth eruption.

Lips

Labial morphology in DS is thought to be normal at birth (Butterworth et al. 1960, Smith and Berg 1976:14), but becomes progressively more anomalous with age. Brousseau and Brainerd (1928) describe the lips in DS as follows:

> "The lips are thicker than normal, they are everted, especially the lower lip, which is unusually prominent, and they are frequently cyanotic. The lips are crossed by transverse fissures..... This mucous membrane of the lips is very sensitive, and tends to become irritated by the frequent flow of saliva....."

The main points in this description are echoed in many texts (Shuttleworth 1909, Pearce et al. 1910, Brushfield 1924, Benda 1969), although the reported incidence of these features varies somewhat. The incidence of lip fissuring, for example, is given as 35% by Pearce et al. (1910), but as 56% by Levinson et al. (1955). Levinson et al. also found 36% of their DS subjects to have broad lips and 28% to have irregular lips, whereas Øster (1953) found 29% to have broad lips, and as many as 40% to have irregular lips. Such discrepancies are not surprising given the subjective nature of such judgements.

Few studies have looked specifically at the relationship between age and labial abnormality, but Butterworth et al. (1960) do suggest a clear correlation, and Benda (1969:26) states that the mucosa of the lips becomes abnormal "early in life". Butterworth and associates found that 65% of DS cases eventually show some degree of abnormality. In the usual sequence of events, thickening

and whitening of the skin is followed by fissuring and gradual enlargement of the lips, with scaling and crusting developing later in some individuals. Permanent changes in lip structure were found to be most common in males over 20 years of age.

Factors which predispose DS individuals to dermatological problems of the lips may include avitaminosis, imperfect skin anatomy, and habitual open mouthed posture and tongue protrusion, which results in excessive bathing of the lips in saliva, followed by drying and cracking.

Comments have also been made about the small size of the mouth (Levinson et al. 1955, Joseph and Dawbarn 1970:44). Øster (1953), however, judged all out of a sample of 521 subjects to have normal sized mouths. The problem here is that it is not always clear what parameter is being commented upon. "Mouth size" could apply to the size of the labial aperture, or to the size of the area delimited by the vermilion border. The mobility of the lips makes measurement difficult, and altered proportions of the whole facial structure may influence subjective judgements. Brushfield (1924) does give a figure of 4 cm. for average mouth length in DS, but in the absence of reliable controls this adds little to the discussion.

Overall development of lip structure may be influenced by altered skeletal and muscular anatomy, combined with depressed muscle tone, which disturb the usual mechanical forces acting upon the lips. The importance of mechanical forces during lip development is illustrated dramatically by the gross malformation of the upper lip which occurs when there is no continuity of the obicularis oris in cases of bilateral cleft lip.

The causal relationship between lip posture and labial development is complex, and we are faced with something

of a chicken and egg dilemma. Habitual facial and lip posture may have a long-term affect on lip development because of the resulting mechanical constraints on labial growth. On the other hand, lip posture will itself be constrained by labial anatomy and physiology. In any case, it is interesting to note that lip posture is thought to be unusual from a very early age (Sutherland 1899 and Kasowitz 1902, both cited in Joseph and Dawbarn 1970:61, Joseph and Dawbarn 1970:45, Lind et al. 1970). Lind et al. describe a very narrow vertical opening of the lips in crying DS infants, which they say is very characteristic and in marked contrast to the normal open lip posture. Joseph and Dawbarn suggest that this crying lip posture may be of diagnostic value. In older DS subjects the tendency seems to be towards an open mouthed posture, which is consistent with generalised hypotonia, as well as with the need to maintain an airway (see comments on pharynx and oral cavity above). The incidence of habitual open mouth is given variously as 70% (Pearce et al. 1910), 67% (Øster 1953) and 59% (Gustavson 1964, cited by Joseph and Dawbarn 1970:61).

Velum

The literature contains little comment on the size, shape or function of the velum in DS, but one single case study is worthy of comment. Rolfe et al. (1979) found that cinefluorography of one adult male DS subject showed the velum at rest to be not only shorter, but also considerably thinner than normal. It would be interesting to know if these findings can be extrapolated to the DS population in general, since this would have some bearing on findings of increased nasality in this group.

Nasal cavity

Obstruction of the nasal airway is often said to be a particular problem in DS, and it does seem that the nasal cavity is often somewhat distorted. Skeletal structure of the nose seems to be very variable in DS. Levinson et al. (1955) found "flat" nose in 44% of their group, "small" nose in 54% and flat nasal bridge in 62%. This group spanned a large age range, however, and there are indications that the nature of the deviation from normal may change with age. Smith and Berg (1976:19) say that flatness of the nasal bridge due to under-development of the nasal bones is most marked in the 0-4 year age group. The nasal bones may sometimes remain underdeveloped throughout life, and Kisling (1966) found complete aplasia (lack of development) of the nasal bones in 9 out of 68 adult males.

The cartilaginous part of the nose may become fairly large in later life, giving a "pug-nosed" appearance (Benda 1969:27). The nasal septum and the conchae often deviate, and the mucosal lining may be thickened (Benda 1960, Smith and Berg 1976:20).

2.3.2.5 Summary of organic features and implications for vocal output

All of the above organic deviations may influence the overall configuration or structure of the vocal apparatus. If the idea of auditory equivalence (see chapter 2.1) is accepted, then it is possible to make some tentative hypotheses about the vocal characteristics which might be expected to reflect these particular types of organic anomaly. A summary of DS organic characteristics, together with the voice quality settings which would be expected to result if no compensatory adjustments were made, is given in Figure 2.3/8. The

ORGANIC FACTOR	PREDICTED VOICE QUALITY SETTING
Thick, everted lips	Labial protrusion
Maxillary under-development	Protruded jaw
Short, narrow palate + normal or large tongue	Advanced tip/blade Fronted and raised tongue body
Narrow lumen of pharynx	Pharyngeal constriction
Mucosal disorders	Harshness, whisperiness
Muscular hypotonia	Lax tension settings, minimised ranges, nasal, open jaw, lowered larynx

FIGURE 2.3/8: A summary of reported organic features in DS and predicted consequences for voice quality

following discussion aims to explain and amplify this summary.

i. Features affecting phonation

Phonation will be affected not only by the shape and size of the laryngeal structure, about which there seems to be little data for DS, but also by the layered tissue structure of the vocal folds, the state of the laryngeal musculature, and the efficiency of the respiratory system in providing an adequate airstream. Quite small changes in the structure of the mucosal covering of the vocal folds may cause perturbations of vocal fold vibration which would be perceived as harshness (see Chapters 2.1 and 2.5). More severe structural irregularities of the folds which are sufficient to impede adduction would be expected to cause continuous turbulent airflow throught the larynx, and hence a whispery voice quality. Hypotonicity might also result in incomplete adduction and associated whisperiness, as well as producing a lax laryngeal tension setting.

ii. Features affecting the length of the vocal tract

In a standard vocal tract, the length can be adjusted by raising or lowering the larynx, or by retracting or protruding the lips and jaw. Larynx position settings may be mimicked by anatomical deviations in larynx position, or by altered ratios between the length of the oral and pharyngeal cavities. If the larynx in DS is, as Benda suggests (see above), unusually high in the neck, then an auditory impression of raised larynx might be expected. On the other hand, muscular relaxation often allows the larynx to lower, and the prediction in a non-Down's subject would be for hypotonia to be associated with a lowered larynx setting. No clear hypotheses can be formulated here. At the outer end of the vocal tract

hypotheses are easier. The pseudo-prognathism which results from maxillary underdevelopment leads to a clear expectation of hearing a protruded jaw setting. Eversion and anterior-posterior thickening of the lips may be expected to give the auditory impression of a protruded lip setting.

iii. Features affecting the position and degree of vocal tract stricture

It is possible for a standard vocal tract to assume a configuration which approximates to a bent tube with equal cross sectional area along its whole length. There seem to be several anatomical features in DS which will tend to constrict the tube at various points. Reduction in cross section of the pharynx would be heard as a setting of pharyngeal constriction. Reduced palatal volume, and consequent reduction of the space between the tongue and the palate would be heard as a raised tongue body setting. Forward displacement of the tongue relative to the palate and alveolar ridge would lead to the auditory quality associated with fronted tongue body and advanced tip-blade settings.

iv. Features affecting nasal resonance

The anatomical and physiological correlates of nasal resonance are complex and incompletely understood (Laver 1980:68-92), so that it is difficult to predict the auditory consequences of a particular organic configuration. Overall hypotonicity tends to be associated with poor "tuning" of the velopharyngeal system, and hence increased audible nasality on segments which are not linguistically required to be nasal. If the observation of a short, thin velum reported by Rolfe et al. (1979) reflects a common feature of DS, this might also lead to poor velopharyngeal function, and hence

increased nasality. Chronic obstruction of the nasal cavities, on the other hand, might militate against this. The relative size of the entrances to the oral and nasal cavities is also important (Van Riper and Irwin 1958, Laver 1980:82-83), and Strome's observations suggesting that the proportions of both the naso-pharynx and the oro-pharynx are disturbed in DS may have some relevance here. Again, it is difficult to formulate definite predictions about the relationship between organic factors and perception of velopharyngeal settings in DS.

v. Overall tension effects

In normal speakers a reduction in overall tension of the vocal tract musculature is associated with a constellation of voice quality settings. The high incidence of hypotonia in DS leads to an expectation that all of these settings might be characteristic of DS vocal profiles. The constellation includes the following settings: open jaw, nasal setting, lowered larynx, whispery phonation, minimised range of movement of lips, jaw and tongue, and low means for pitch and loudness. In addition, of course, both laryngeal and supralaryngeal tension settings would be judged as lax. One further organic feature which may exaggerate the auditory impression of overall laxness is the presence of fissuring and roughness of the mucosa, especially in the tongue covering. This is likely to cause acoustic damping, with excessive attenuation of high frequency sounds. Since there is general agreement that one of the main acoustic consequences of lax voice is a reduction in energy in the upper harmonics (Greene 1964:53, Chiba and Kajiyama 1958:17, Laver 1980:142), any increased damping due to fissured mucosa will tend to enhance the impression of laxness in DS. Acoustic attenuation is also an acoustic correlate of increased nasality (House and Stevens 1956, Laver 1980:91), so that there may also be

implications for the perceived levels of nasality in DS (see above).

2.3.3 EXPERIMENTAL INVESTIGATION OF VOICE QUALITY IN DOWN'S SYNDROME

2.3.3.1 Subjects and procedures

DS Group

The DS group consisted of 20 female subjects. 10 of these were resident in hospitals in Fife and Lothian, whilst the remaining 10 were living in the community and attending an adult training centre in Fife. It had been hoped that an equivalent number of male subjects would be available, but only 6 male DS subjects were available for the recording sessions at these centres. It was felt that this small number did not allow sensible statistical analysis of results, so males were excluded from the study. All subjects were judged to have adequate hearing to cope with normal conversational speech. Ideally, all subjects should have been given some kind of audiometric screening, since it is possible that low levels of hearing loss may influence voice quality. The staff at the centres involved in this study were, however, understandably reluctant to subject the DS group to more disruption than was absolutely necessary. The lack of audiometric data is somewhat worrying in the light of studies which show a significant increase in the incidence of conductive hearing loss in the DS population (Fulton and Lloyd 1968, Brooks et al. 1972, Nolan et al. 1980). It is hoped that the exclusion from the study of any individuals who were suspected by familiar staff of having any difficulty in hearing will minimise any bias in the results due to audiological impairment in the population. Cytogenetic data was used to confirm the clinical diagnosis of Down's Syndrome, and none were thought to be mosaics. The age range of subjects was 20 - 36 years, with a mean of 28.9 years. Each subject was recorded in a quiet room, using a portable Uher tape

recorder and a directional microphone (Sennheisser). The speech sample consisted of spontaneous conversational speech, picture description, and serial speech (counting, days of the week etc.).

Control group

The control group consisted of 16 females, who were all native speakers of Scottish English, and who had no history of speech or hearing impairment. The age range was 18-32 years, with a mean of 20.3 years. These subjects were recorded either in a quiet room, using a portable Uher tape recorder, or in a sound proofed booth, using a Ferrograph tape recorder. The speech sample consisted of spontaneous speech, a standard reading passage (the first paragraph of "The Rainbow", Fairbanks 1960), and serial speech.

Procedure

All subjects were allowed a short time to become familiar with the interviewer (the author) before recording took place. During this time observations were made of dentition, jaw relationships, etc.. These subjective observations of the DS subjects suggest that they were fairly typical of the DS population as described in the preceding literature survey. Recording was carried out in as relaxed a context as was possible in the presence of a visible microphone, and efforts were made to ensure that an absolute minimum of 40 seconds of continuous speech was available for vocal profile analysis. This was obviously less easy in the case of the DS subjects, who tend to be less fluent linguistically, but subjects were excluded if it was not possible to obtain 40 seconds of reasonably continuous speech.

Vocal Profile Analysis

A consensus vocal profile was completed for each subject, as described in section 2.2, and these profiles provided the raw data for group comparisons. Three judges (JL, JM, S.W.) were involved in construction of the DS composite profiles. Unfortunately only two judges (JL, JM) were available for construction of the composite profiles for the control group, but the high level of agreement between these two judges (see section 2.1.2) justifies a belief that these results are nonetheless valid.

A summated protocol was prepared for each subject group as described in section 2.1.2. From these, the mean scalar degree (MSD) and the standard deviation (SD) were calculated for each setting scale.

2.3.3.2 Results

Figures 2.3/9 and 2.3/10 show summated protocols for the DS and control groups, and this information is summarised in Figure 2.3/11, which is a comparison of the MSDs and the SDs for the DS and the control group. The differences were tested for significance using the Mann-Whitney U test (Siegel 1956:116), and it can be seen that for 11 out of the 18 vocal quality scales the difference between the DS and the control groups is significant with a probability of 0.02 or less. These scales are summarised in graphic form in Figure 2.3/12. Differences in some other setting scales (lip posture and tip/blade features) were also quite marked, but not statistically significant

These results can now be related to the hypotheses proposed in section 2.3.2.5. Figure 2.3/13 summarises these predictions, and compares them with the actual findings. It can be seen that whilst many of the findings do fit remarkably well with predictions based entirely

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial	10	6		Lip Rounding/Protrusion	4	1				
				Lip Spreading	1					
				Labiodentalization						
	13	3		Extensive Range	1					
				Minimised Range	2					
2. Mandibular	9	7		Close Jaw	5	1				
				Open Jaw	1					
				Protruded Jaw						
	15	1		Extensive Range						
				Minimised Range	1					
3. Lingual Tip/Blade	6	10		Advanced	8	2				
				Retracted						
4. Lingual Body	1	13		Fronted Body	4	7				
				Backed Body	2					
		13		Raised Body	7	6				
				Lowered Body						
	15	1		Extensive Range						
				Minimised Range	1					
5. Velopharyngeal		16		Nasal		3	13			
				Audible Nasal Escape						
				Denasal						
6. Pharyngeal	13	3		Pharyngeal Constriction	3					
7. Supralaryngeal Tension	6	10		Tense	8	2				
				Lax						
B. Laryngeal Features										
8. Laryngeal Tension	5	11		Tense	6	4	1			
				Lax						
9. Larynx Position	10	6		Raised	1					
				Lowered	4	1				
10. Phonation Type		3		Harshness	2	1				
		14	2	Whisper(y)		8	6	2		
				Breathiness						
		10		Creak(y)	1	5	5			
				Falsetto						
		16		Modal Voice	16					

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192) Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

FIGURE 2.3/9: Summated Vocal Profile Analysis results for control group

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial	7	13		Lip Rounding/Protrusion	1	1	1			
				Lip Spreading	3	5	2			
				Labiodentalization						
	4	15	1	Extensive Range						
				Minimised Range		5	10	1		
2. Mandibular	3	16	1	Close Jaw	1	1				
				Open Jaw	3	1	3	1		
				Protruded Jaw	5	9	3			
	3	16	1	Extensive Range		1				
				Minimised Range	1	8	6	1		
3. Lingual Tip/Blade	4	13	3	Advanced	3	2	5	3		
				Retracted	1	2				
4. Lingual Body		14	5	Fronted Body		5	8	5		
				Backed Body		1				
		15		Raised Body	1	8	5			
				Lowered Body		1				
		16	4	Extensive Range						
				Minimised Range		4	12	4		
5. Velopharyngeal		7	13	Nasal			7	13		
				Audible Nasal Escape				2		
				Denasal						
6. Pharyngeal	7	11	2	Pharyngeal Constriction	3	7	1	2		
7. Supralaryngeal Tension	4	15	1	Tense	1	1	1			
				Lax	4	7	1	1		
B. Laryngeal Features										
8. Laryngeal Tension	1	18	1	Tense	4	7	5	1		
				Lax		1	1			
9. Larynx Position	9	11		Raised		1				
				Lowered	4	5	1			
10. Phonation Type		16	3	Harshness		8	8	2	1	
				Whisper(y)		1	6	12	1	
				Breathiness						
				Creak(y)	3	2	5			
				Falsetto						
		20		Modal Voice	20					

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

FIGURE 2.3/10: Summated Vocal Profile Analysis results for DS group

SETTING SCALE		CONTROL (n=16)		DOWN'S (n=20)	
		MSD	SD	MSD	SD
+ − Lip round/spread		+.31	.70	-.65	1.57
Labiodentalization		0	0	0	0
− + Labial range (min/ext)	**	-.06	.44	-2.2	1.57
− + Open/close jaw	**	+.36	.72	-.75	1.52
Protruded jaw	**	0	0	1.6	.94
− + Mand. range (min/ext)	**	-.06	.25	-1.85	1.42
+ − Adv./retr. tip-blade		+.75	.68	+1.45	1.93
+ − Front/back tongue body	**	+1.0	1.10	+2.60	1.47
+ − Raised/lowered T.B.		+1.19	.75	+1.50	1.40
− + Body range (min/ext)	**	-.06	.25	-3.00	.65
+ − Nasal/denasal	**	+2.81	.40	+3.65	.49
Aud. nasal esc.		0	0	0.4	1.23
Phar. constric.	**	.19	.40	1.40	1.31
+ − Supra. tense/lax	**	+.75	.68	-.95	1.67
+ − Lar. tense/lax		+1.06	.93	+1.60	1.70
− + Low/raised larynx		-.31	.70	-.75	1.16
Harsh	**	.25	.58	2.65	1.04
Whisper	**	2.63	.72	3.65	.67
Creak		1.63	1.26	1.10	1.29

FIGURE 2.3/11: Statistical comparison of summated protocols for DS and control groups. ** indicates settings where $p<0.02$ that the two groups share the same distribution (Mann-Whitney U test)

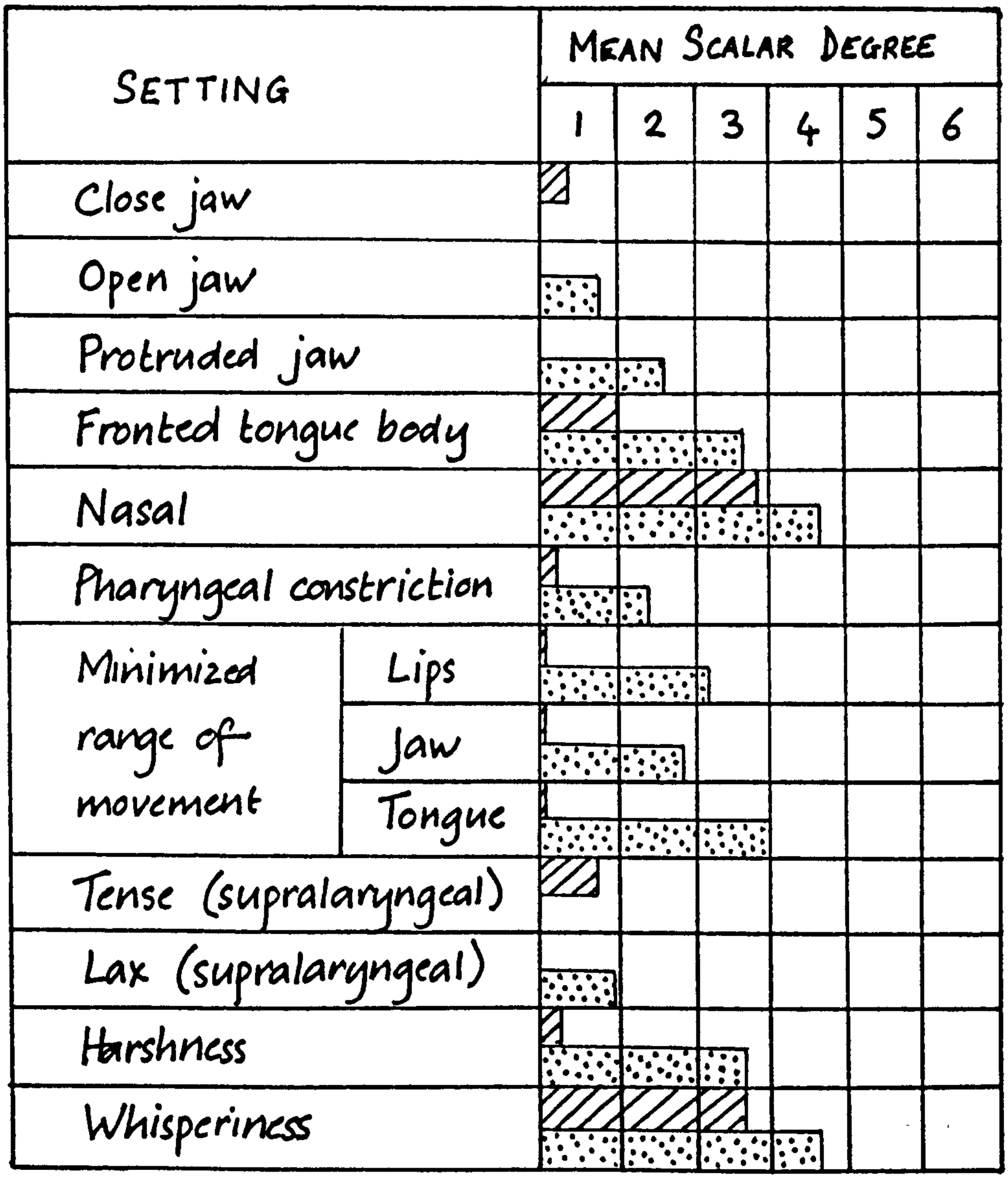


FIGURE 2.3/12: A graphic representation of significant differences in mean scalar degree of vocal quality settings between DS and control groups

PREDICTED VOICE QUALITY SETTING	VPA RESULTS
Labial protrusion	Lip spreading
Protruded jaw	Protruded jaw **
Advanced tip/blade Fronted and raised tongue body	Advanced tip-blade Fronted** and raised tongue body
Pharyngeal constriction	Pharyngeal constriction**
Harshness, whisperiness	Harshness** Whisperiness**
Lax tension settings, minimised ranges, nasal, open jaw, lowered larynx	Lax (supralar.**) Tense (laryngeal) Minimised range of lips**, jaw**, tongue** Open jaw** Lowered larynx Nasal**

FIGURE 2.3/13: A comparison of predicted voice quality settings and VPA results for the DS group
** indicates significantly different from controls ($p < 0.02$)

upon organic characteristics, there are some vocal profile features which conflict with the predictions.

A protruded mandibular setting certainly seems to be more common in the DS group than in controls, which is consistent with the increased incidence of maxillo-mandibular imbalance resulting from maxillary underdevelopment.

The unusual relationship between the tongue and the oral cavity volume is reflected in an increased incidence and degree of tongue fronting in the DS group, as would be expected given the marked reduction in palatal length in the DS population. Although the DS group did, on average, show more raised tongue body settings than the controls, this difference was not statistically significant. It may be that the auditory identification of tongue raising (i.e. reduction of oral cavity cross section in the palatal and velar area) was complicated by the existence of pharyngeal constriction (see below).

Pharyngeal constriction as an auditory setting was more common and more marked in the DS group, and was frequently combined with a neutral or lax supralaryngeal tension setting. This combination is rare in normal speakers (see Section 2.1.2), and it seems reasonable to suppose that the finding of pharyngeal constriction is a direct result of an anatomically derived reduction in the lumen of the pharynx.

Generalised hypotonia would be predicted to result in a constellation of settings, including increased nasality, minimised articulatory ranges, open jaw, and of course lax tension settings. All of these, with the exception of laryngeal tension, were found to be significantly more marked in the DS group.

Finally, phonation does seem to be significantly more whispery and more harsh in the DS group, and it is important to note that harshness quite commonly occurred in the absence of laryngeal tension. This is an uncommon finding in speakers with normal larynges (see section 2.1.2), and may thus be taken as an indication that the irregular phonation may have some organic basis, such as abnormal mucosal covering of the vocal folds.

Some predictions were not borne out by the results. The thickened and everted lip posture, for example, was predicted to be associated with the auditory judgement of a protruded lip setting in the DS group, but in fact the lip setting seemed on average to be slightly more spread in the DS group. Also, a fronted tongue posture relative to the maxillary arch would be expected to result in higher levels of fronted tip/blade settings in the DS group, but this tendency was not significantly different from the controls. This may be because the control group also displayed a sociolinguistic bias towards a fronted tip/blade setting.

One other important finding which emerges from Figure 2.3/11 is that the standard deviations of scalar degree judgements are higher for the DS group than for the control group for all scales except whisperiness. This points to a high degree of variability of vocal parameters in DS and belies any suggestion that there is a characteristic DS voice. Rather, Vocal Profile Analysis prompts an echo of Shapiro's (1973) comments on physical parameters. It seems that for voice, too, it is "the frequency, intensity, and multiplicity of anomalies that are characteristic".

One explanation of this increased variability is that it is a direct reflection of variability in organic features and of the constraints they impose on phonetic

performance. Attractive though this explanation may be in the context of this thesis, at least one alternative explanation can be proposed. This relates to the channelling effect that exposure to a speech community may have on vocal development. It has been shown quite clearly that speech communities may vary in at least some vocal profile settings (Esling 1978). This is supported by an analysis of the data presented in the previous section (2.2), which shows that the closer are a group of speakers in their geographical origins, the more similar are their vocal quality settings. It seems, therefore, that individuals tend to conform to the norms of their speech community in terms of long term vocal characteristics as well as in segmental accent features. In the case of DS speakers, it is possible that perceptual and linguistic deficits may interfere with their ability to perceive and assimilate the subtleties of the vocal models presented by their speech community. Their vocal development may thus be less narrowly channelled towards their speech community's norms.

Although it is clear that there are significant differences in group means of vocal settings between the DS and control groups, the high variability of the DS group does not allow an immediate conclusion that the overall vocal profile of any given DS individual is necessarily likely to be more similar to that of another DS speaker than to that of a control group speaker. It is therefore not possible to comment on the ability of the VPAS to discriminate between DS and control speakers.

A simplistic way of approaching this problem is to make all possible pairwise comparisons of the VPA protocols of the DS and control subjects, and to see if there is more similarity within groups than between groups. This was done, and for each pair of protocols, the number of vocal quality setting scales which differed by more than one

scalar degree was recorded. This figure can then be taken as an index of dissimilarity for any pair of individuals.

Figure 2.3/14 a., b., and c. shows the results of pairwise comparisons for:

a. all possible DS vs. DS comparisons
b. all possible control vs. control comparisons
c. all possible DS vs. control comparisons.

The results show that out of 120 control vs. control comparisons there were 17 (14.2%) "twins", where no settings differed by more than one scalar degree. On average, pairs of protocols had 2.1 vocal quality settings which differed by more than one scalar degree. This low figure is not unexpected, given that scalar degree judgements for the control group mostly fell below scalar degree 3 or 4, and it emphasises the extent of the normal "canalization" effect

In the 190 DS vs. DS comparisons there were no "twins", and on average 6.2 scales differed by more than one scalar degree for each pair. This higher figure reflects the higher standard deviations in DS.

The crucial question is whether the DS group differ even more from the controls than they do from each other. In the 340 DS vs. control comparisons there were again no "twins", and the average number of settings by which any two protocols differed was 9.3. It does therefore seem that the vocal profile of a DS speaker is likely to resemble other DS speakers more closely than controls.

2.3.4 DISCUSSION AND CONCLUSIONS

These results are consistent with predictions that organic characteristics in a population with DS will have consequences for their auditorily perceived voice

Down's Syndrome Subjects (20 ♂♂)	DOWN'S SYNDROME SUBJECTS (20 ♀♀)																			
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A																				
B	6																			
C	4	7																		
D	5	6	9																	
E	9	8	12	9																
F	8	5	7	8	7															
G	1	5	5	3	10	5														
H	5	5	7	8	6	4	5													
I	4	6	5	8	8	6	4	4												
J	4	4	2	8	9	5	6	5	5											
K	6	3	3	8	8	3	5	4	5	3										
L	9	6	7	8	6	2	6	1	6	7	3									
M	9	7	6	10	7	8	9	6	6	8	6	5								
N	6	5	7	7	6	7	5	4	7	6	6	5	8							
O	4	7	9	7	9	7	4	6	6	8	7	6	10	7						
P	4	8	5	3	8	6	5	6	6	5	2	7	9	8	6					
Q	8	8	10	8	8	6	8	7	9	8	7	7	12	8	8	6				
R	3	4	5	4	5	3	2	2	5	4	4	1	8	3	5	1	4			
S	7	6	7	6	9	5	5	8	6	5	4	6	9	4	8	5	4	5		
T	10	5	7	8	1	8	10	6	8	9	6	5	4	4	8	8	11	6	10	

FIGURE 2.3/14, A: Pairwise calculations of the number of vocal quality settings which differ by more than one scalar degree: DS vs. DS comparisons

CONTROL SUBJECTS (16 ♀♀)	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
A																
B	1															
C	1	6														
D	2	·	5													
E	1	4	2	4												
F	·	4	3	4	4											
G	1	2	1	2	1	2										
H	2	5	6	3	3	6	4									
I	1	2	1	3	2	2	-	5								
J	1	1	3	-	-	3	-	2	-							
K	1	3	1	3	3	2	·	6	·	1						
L	1	2	3	2	1	2	-	5	·	-	1					
M	1	4	3	4	3	1	4	4	3	3	3	4				
N	·	1	3	1	2	1	1	2	1	1	2	1	1			
O	·	1	3	1	1	-	2	2	2	·	2	2	1	-		
P	1	2	2	5	5	1	2	7	2	2	1	3	2	1	2	

CONTROL SUBJECTS (16 ♀♀)

FIGURE 2.3/14, B: Pairwise calculations of the number of vocal quality settings which differ by more than one scalar degree: control vs. control comparisons.

DOWN'S SYNDROME SUBJECTS (20 ♀♀)	CONTROL SUBJECTS (16 ♀♀)															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
A	10	12	10	12	8	11	13	10	9	10	10	10	11	8	10	12
B	11	9	10	10	8	11	12	10	8	10	9	8	10	7	11	11
C	11	12	11	9	10	12	14	11	10	13	11	13	13	11	12	13
D	7	9	9	10	6	7	8	6	8	6	6	6	7	5	6	8
E	7	7	9	6	8	7	6	9	6	4	8	5	8	6	6	7
F	9	11	11	10	9	10	10	11	8	9	11	10	11	10	9	11
G	12	10	9	12	7	11	11	11	7	10	10	9	10	8	11	11
H	10	12	12	9	10	11	11	13	7	9	11	11	12	9	11	12
I	11	13	12	12	9	12	13	11	10	13	12	12	13	11	12	14
J	10	11	10	9	7	9	12	11	8	11	8	11	12	10	8	11
K	12	11	11	11	8	12	12	11	9	11	10	9	11	9	10	13
L	12	11	14	9	12	11	12	13	8	12	13	10	13	11	10	13
M	13	12	15	11	13	12	13	14	10	12	14	13	14	13	12	13
N	7	9	6	6	6	8	9	7	4	8	8	7	8	5	8	10
O	12	10	11	9	8	10	12	11	8	10	11	9	11	6	10	12
P	6	8	5	6	5	8	8	7	7	6	7	6	8	5	6	7
Q	3	5	5	3	5	4	4	5	6	4	5	5	5	5	3	5
R	8	9	9	8	8	10	10	10	5	6	11	8	10	6	9	11
S	10	6	8	8	5	8	8	9	7	8	8	8	8	6	8	9
T	11	8	11	7	9	10	9	9	8	7	8	7	8	8	9	8

FIGURE 2.3/14, C: Pairwise comparisons of the number of vocal quality settings which differ by more than one scalar degree: DS vs. control comparisons

quality. This study does not, however, address several issues which could complicate interpretation of the results.

Firstly, it does not take into account the possible effects of mental handicap or linguistic deficit. Ideally there should have been a control group which was matched for mental age and linguistic competence, but which included only individuals with standard vocal tracts. Given that the relationship between mental age and linguistic competence is characteristically different in DS from other forms of mental handicap, the collection of such a control group would be a formidable task. Since this study was carried out as part of a much larger study (Laver et al. 1982), such a task could not feasibly be undertaken.

Secondly, there is a possibility that institutionalization may influence the voice settings chosen by an individual. There is some evidence that prolonged institutionalization has some effect on speech and language used, such that individuals tend to conform to the patterns they hear around them. If this effect extends to voice quality, a difference might be expected between the 10 institutionalized DS subjects and the 10 living in the community. Unfortunately the number of subjects is too small to allow a sensible statistical evaluation of this possibility.

Thirdly, this study cannot exclude the possibility that a high incidence of mild hearing loss in the DS population is instrumental in causing the observed voice quality features. The prevalence of conductive hearing loss in the DS population has already been mentioned, and whilst the subjects in this study were all judged to have adequate hearing for social interaction, full audiometric assessments were not feasible. The possibility that some

vocal profile deviations are associated with mild hearing loss cannot therefore be ruled out. In this context, it is interesting to note that the DS group does share some vocal profile features with a group of profoundly deaf young adults who were also studied in the MRC project "Vocal Profiles of Speech Disorders" (Laver et al. 1982, Wirz 1987). Both groups showed increased degrees of harshness and whisperiness, as well as minimised ranges of articulation, but some factors do indicate differing underlying causes for these features. For example, in the deaf speakers, harshness was almost invariably combined with laryngeal tension, as is the case in normal speakers. In the DS group, harshness was more often associated with a lax laryngeal setting, indicating that the harsh quality was the result of organic abnormality rather than an unusual phonetic adjustment. Similarly, the degree of supralaryngeal tension associated with minimised articulatory ranges was different in the two groups, with the deaf speakers exhibiting much higher degrees of tension. In short, the fact that although the deaf and the DS groups do share some vocal features, they nonetheless show very different overall vocal profiles supports an assertion that hearing loss is probably not playing a major role in the causality of the DS findings.

If the proposition that many, if not most, of the DS vocal profile features are the result of organic anomalies is accepted, then several valuable uses of the vocal profile present themselves. One possible application is in the monitoring of drug regimes aimed at alleviating the developmental effects of biochemical and hormonal imbalances in DS (Benda 1969).

One very controversial area of work which calls out for monitoring of the sort which the VPAS could offer is the use of cosmetic plastic surgery in DS. Plastic surgery in DS is increasing, but there seems to have been little

systematic study of the effects on speech and voice quality.

The relationship between organic factors and voice quality in populations such as the DS population has a wide range of implications. In terms of speech therapy, the link is very important. Speech therapy aimed at improving vocal patterns can only hope to be effective if the organic state of an individual allows the possibility of change. A useful illustration of this is given by a patient who presents with what is considered to be an unacceptably high degree of harshness. If the harshness is due to inefficient muscular patterns for phonation, as would probably be the case in a speaker with a normal, healthy larynx, then the potential for change is clear, and therapy aimed at acquiring more appropriate muscular control of phonation may be indicated. If, however, the harshness is primarily due to abnormalities in the mucosal covering of the vocal folds which interfere with regular vocal fold vibration, then the potential for change as a result of speech therapy alone is very limited. This may well be the case in DS. Similarly, if tongue fronting relative to the maxillary arch is due to grossly abnormal jaw relationships, then speech therapy aimed at reducing tongue fronting will also have a limited chance of success. It is thus clear that therapists working with any population of DS individuals, where the incidence of organic anomalies is rather high, should always make a careful evaluation of the extent to which speech or voice abnormalities are the inevitable consequence of organic features. Only then will they be able to make informed decisions about the potential for change in phonetic output.

The finding that the DS population does, in fact, show a rather high incidence of voice features which are different from the controls also has implications for the

social interactions of these individuals. This is because there is the possibility of confusion between organically caused vocal abnormalities and the paralinguistic use of voice. Harshness, again, is a good example of this. Paralinguistically, harshness is used in many cultures to communicate anger and aggression (Bezooijen 1984). The habitual use of harshness as a result of organic abnormality may therefore be misinterpreted as a signal of aggression, causing obvious difficulties in social interactions. On an anecdotal level, it was surprising how often staff described DS individuals as "surly" or "aggressive", and the author did begin to wonder if this was related to the use of harshness. Similarly, lax, whispery phonation, which is also common in DS, may be interpreted as an indication of depression or introversion (Saville 1983), and tongue fronting is part of society's stereotype of immaturity. The social interactions of DS individuals might be made a lot easier if all those who care for them were made aware of the potential misunderstandings resulting from organic factors.

In conclusion, it does seem possible to find clear links between organic state and voice quality in the DS population, although there is a need for this work to be extended, in order to properly evaluate the possible contributions of hearing impairment and mental or linguistic handicap. An examination of voice quality in DS individuals where the degree of organic abnormality has been limited by medication or surgery would also allow further elucidation of the relationship between organic features and voice.

2.4 ACOUSTIC CHARACTERISTICS OF NORMAL PHONATION

One of the aims of the MRC Project "Acoustic Analysis of Voice Features" was to establish a normal baseline against which speakers with known vocal fold pathology could be judged. For this reason, a control group of speakers with no known vocal fold abnormalities was recorded, and their voice recordings were subjected to acoustic analysis as described in section 2.1.3. The control group was spread over a wide age range, so a secondary aim was to investigate the effects of age upon acoustic parameters of phonation.

Subjects and data collection

183 control speakers were recorded (83 males and 80 females). All were native speakers of British English, and were mostly staff and students of the University. They cannot be said to be fully representative of a Scottish population, but an assumption underlying the study was that the effects of vocal pathology upon acoustic features of phonation are likely to be greater than the effects of accent or sociolinguistic factors. Obviously further research would be needed to fully test this assumption. It was not possible to give each subject a laryngoscopic examination, but any subjects who reported temporary throat infections or any history of laryngeal disorder were excluded from the control group. Smokers, however, were not excluded. There seems to be little data available about the incidence of observable laryngeal abnormality amongst the general population who are not attending ENT clinics, so that, in the absence of laryngoscopic data, we cannot exclude the possibility that a certain proportion of the control group may have had minor laryngeal disturbances. The male and female control groups were treated separately, since the slightly different analysis conditions for male and

female subjects make it difficult to make direct comparisons between males and females (see section 2.1.3). The age means and ranges of the control groups are shown in Figure 2.4/1, which also shows the proportion of smokers within the groups.

The speech data consisted of the first 40 seconds of "The Rainbow Passage" (Fairbanks 1960). Before recordings were made, subjects were given some time to familiarise themselves with their surroundings, and with the text, and they were asked to read at their normal speaking volume and rate. All recordings were made in a sound-treated recording studio, using a shotgun microphone (Sennheisser MKH815T) with power supply (Audio Engineering AKB11) and a REVOX A77 tape recorder. A REVOX A77 recorder was also used to play back the recordings for digitization prior to acoustic analysis.

The following acoustic parameters were collected for each subject: F0-AV, F0-DEV, J-AVEX, J-DEVEX, J-RATEX, J-DPF, S-AVEX, S-DEVEX, S-RATEX and S-DPF (see Section 2.1.3).

Results

The means and standard deviations of all acoustic parameters are shown in Figure 2.4/2. Distributions for most parameters were found to be approximately normal in shape, but one of the shimmer measures, S-DEVEX, has a highly skewed distribution. For this reason, this measure was excluded from many of the statistical procedures described later in this chapter and in the following chapter.

Sex differences

It has already been stated that direct comparison of male and female acoustic data is not really valid in view of

	FEMALES	MALES
N	80	83
Age Mean	38.13y	36.08y
Age range	18-84y	18-71y
* % smokers	28.8%	43.4%

* A smoker is defined as a subject who reports any history of regular smoking.

FIGURE 2.4/1: Age and smoking characteristics of normal subjects

	FEMALES (N=80)		MALES (N=83)	
	MEAN	SD	MEAN	SD
F0-AV	195.8	20.11	112.6	13.43
F0-DEV	40.75	8.11	20.96	5.99
J-AVEX	4.84	1.15	5.10	1.23
S-AVEX	14.24	4.19	17.19	5.86
J-DEVEX	14.85	2.25	16.07	3.28
J-RATEX	20.41	4.58	23.03	4.15
S-RATEX	47.46	7.04	58.82	6.04
J-DPF	13.13	3.45	16.22	3.25
S-DPF	21.91	4.38	27.55	5.31

FIGURE 2.4/2: Table of acoustic results for normal speakers

the different analysis conditions. It may, nonetheless, be interesting to look briefly at the acoustic differences which do emerge, and to consider whether they might be related to organic factors. The most obvious and predictable acoustic difference between male and female controls is in F0-AV. The mean figures of 195.8 Hz for females and 112.6 Hz for males is broadly in line with other studies (see Section 1.2.5), and is clearly related to the different size of male and female larynges. F0-DEV is also approximately twice as high in the female group. All the perturbation measures are higher in the male group. This may be an artefact of the analysis procedure, or it may be a true indication of greater irregularity in male phonation. The observation that harshness and creakiness are slightly more common in control males (see Section 2.2) suggests that male phonation may actually be more perturbed. If this is the case, then we must consider to what extent this may be sociolinguistically determined, and to what extent it results from differences in laryngeal structure. Considering the latter possibility, it may be that the larger male larynx is inherently less efficient mechanically, or it may be that males in our society are more exposed to factors which affect vocal fold structure deleteriously. Possible culprits include alcohol, smoking or even a tendency to use too loud a volume.

Age related differences

Previous studies have suggested that some acoustic parameters are correlated with age (see Section 1.2.5), which may reflect age related changes in the vocal apparatus. The control group was therefore divided into three age bands; 18-29 years, 30-54 years and 55 years and over. Means and standard deviations of acoustic parameters for the three age bands are shown in Figure 2.4/3. An analysis of variance (Kruskal-Wallis one-way

A. Control Females

ACOUSTIC PARAMETER	18-29yrs N=37 (mean = 21.3yr)		30-54yrs N=23 (mean = 41.0yr)		55 + yrs N=19 (mean = 68.4yr)	
	Mean	SD	Mean	SD	Mean	SD
F0-AV **	205.56	17.19	185.23	16.88	189.87	21.45
F0-DEV	40.81	6.45	37.47	6.10	44.71	11.39
J-DEVEX	15.03	2.18	14.71	2.40	14.69	2.35
J-AVEX	4.83	1.08	4.88	1.34	4.80	1.13
S-AVEX	13.85	3.39	15.04	5.90	14.04	3.25
J-RATEX	19.82	3.66	21.24	6.06	20.51	4.34
S-RATEX	45.48	5.89	49.84	7.35	48.34	8.10
J-DPF	12.29	2.40	14.23	4.48	13.43	3.62
S-DPF *	20.71	3.76	23.59	4.62	22.21	4.81

B. Control Males

ACOUSTIC PARAMETER	18-29yrs N=36 (mean = 21.5yr)		30-45yrs N=31 (mean = 39.6yr)		55 + yrs N=16 (mean = 62.2yr)	
	Mean	SD	Mean	SD	Mean	SD
F0-AV	108.53	9.30	115.69	15.08	116.00	16.08
F0-DEV **	18.24	4.42	22.00	5.63	25.12	6.97
J-DEVEX *	16.75	3.21	14.93	2.84	16.77	3.80
J-AVEX	5.17	1.22	4.82	1.12	5.46	1.40
S-AVEX	16.93	4.38	16.98	7.70	18.22	4.82
J-RATEX	22.76	4.10	22.66	3.98	24.35	4.58
S-RATEX	59.16	6.13	57.45	5.75	60.70	6.13
J-DPF	16.08	3.04	16.09	3.31	16.82	3.72
S-DPF	27.62	5.26	26.76	4.79	28.92	6.34

** = Age effect significant at $p < 0.01$ level

* = Age effect significant at $p < 0.05$ level

(Kruskal-Wallis one-way analysis of variance)

FIGURE 2.4/3: Table of acoustic results for three age bands, and statistical significance of age-related differences

analysis of variance, Siegel 1956:184) was used to determine whether acoustic parameters were significantly different within the three age bands. The results of this analysis are shown in Figure 2.4/3. For males, only F0-DEV and J-DEVEX are significantly different at a level of P = .05 or better.

F0-DEV does seem to show a clear increase with age, but it is difficult to relate this to any anatomical or physiological trends. Whilst it could be an indication of reduction in the ability to control pitch, such that inappropriately wide pitch fluctuations are more common in the older age groups, it is also possible that cultural changes have resulted in different reading styles in the different age groups. The adoption of large intonational pitch movements in reading may be less popular in the younger age groups.

J-DEVEX shows a more complex pattern, being lowest in the middle age band. It is notable that several other perturbation scores are similarly lowest in the middle age group, even though the differences do not reach statistical significance. Various post-hoc explanations of this finding could be proposed, but it is again hard to relate this finding to organic factors. It is possible that the youngest age group is still afflicted by some instability of vocal fold structure as the final stages of growth within the vocal apparatus are completed, and that this instability poses problems of neuromuscular control which make phonatory irregularities more likely. Increased perturbation in the oldest age group is less unexpected, given the increasing incidence of deleterious changes in the vocal folds in elderly populations (see Section 1.2.5).

In the female group only F0-AV and S-DPF show significant age effects. F0-AV is highest in the younger age group,

which is consistent with many other studies (see Section 1.2.5). S-DPF, and most of the other perturbation measures, is lowest in the youngest age group, in marked contrast to the male findings. This could be because maturation of the vocal apparatus is completed earlier in females, so that the vocal system is fully stable by 18 years.

In general the age effects on acoustic analysis seem to be rather small. This will be important in interpreting the results presented in the following section, where the possibility that acoustic differences between control and pathological speakers might be due to different age distributions within the two groups has to be considered.

Conclusions

This brief summary of the acoustic characteristics of a supposedly normal control group illustrates the point that phonation is typically far from regular, and that a certain level of jitter and shimmer must be considered to be normal. The data presented in this section will serve as a baseline against which the acoustic characteristics of speakers with vocal fold pathology can be compared.

2.5 ACOUSTIC ANALYSIS IN LARYNGEAL PATHOLOGY

2.5.1 INTRODUCTION

The application of acoustic analysis to laryngeal pathology has interested many workers (see review in Hiller 1985: 141-210), for good reason. An examination of the mechanics of vocal fold vibration makes it clear that the optimum, regular mode of vibration is possible only if the speaker possesses healthy vocal folds with no disturbance of the normal tissue layer relationships. Any organic lesion of the vocal folds is likely to disturb this mechanically efficient vibratory structure to some extent, with predictable effects on the mode of vibration. Many of these alterations in the pattern of vocal fold vibration should be clearly reflected in the acoustic laryngeal wave-form. If the links between organic vocal fold lesions and acoustic wave-form patterns can be established, then the way is open to develop effective acoustic screening programmes for laryngeal disorders.

An automatic acoustic system which can detect laryngeal pathology has several potential applications.

1. It could be used for PRE-DIAGNOSTIC SCREENING of voices for the potential presence of laryngeal pathology. This kind of screening of patients, before their first visit to an ENT clinic, could usefully be applied to two different populations. The first of these is an unselected population. It is envisaged, for example, that routine screening for vocal pathology might be carried out in "well-woman" and "well-man" clinics. The system would thus be used alongside existing screening tests for breast and cervical cancer, cardiac function etc.. In this way, a broad sector of the community could be

screened, whether or not a vocal problem was previously suspected.

The second type of pre-diagnostic screening, which might be more readily implemented, could be applied to a pre-selected population. An example of this is the screening of a more limited population where vocal pathology is already suspected. For instance, in some areas patients referred by GPs for laryngeal examination may face lengthy waiting lists. If these patients could be recorded at the time of referral, it might be possible to ensure that those cases where acoustic measures suggest the presence of an organic abnormality would be seen immediately. In this way, acoustic screening could be used to select priority cases.

2. It could be used to aid DIAGNOSTIC DIFFERENTIATION between various classes of vocal fold pathology, providing additional information at the time of the first visit to an ENT clinc. The first laryngeal examination at an outpatient clinic usually involves indirect laryngoscopy. There are at least two reasons why back-up acoustic analysis might be useful at this stage. Firstly, many clinics do not have stroboscopic equipment, and can therefore only obtain a static view of the vocal folds. Even where a stroboscopic view is available, the patient is in a very unusual and often uncomfortable position, so that the mode of vibration seen is probably not representative of habitual phonation. Acoustic analysis can give information about vibratory movement of the vocal folds during connected speech. This could play a part in the next stage of treatment or assessment. For example, it could help in decisions about further referral for direct laryngoscopy, or to a clinic with fiberoptic and stroboscopic equipment.

Secondly, some patients are unable to tolerate indirect laryngoscopy, so that no view of the vocal folds can be obtained. Again, acoustic information might help to decide the appropriate course of action.

3. Acoustic analysis could be used for POST-DIAGNOSTIC TRACKING of changes in laryngeal status. There are three obvious applications of such an approach. Firstly, to chart the course of therapy, whether this involves drug therapy, surgery, radiotherapy, or speech therapy. Secondly, in review patients, as an early warning system against recurrence of disease. Thirdly, it may be used to monitor deterioration in cases of progressive disease.

An acoustic system has the advantage, in all these applications, of being completely non-invasive. The recording procedure is simple, and highly portable, so that it could be operated by non-medical personnel in any relatively quiet situation in clinics, factories, schools, etc.. The procedure also causes minimal distress to subjects.

2.5.2 PREDICTED CONSEQUENCES OF VOCAL FOLD PATHOLOGY

If we accept the hypothesis that it is the mechanical disturbance of vocal fold function which will allow organic abnormalities to be detected acoustically, then it is important to consider the kinds of mechanical disruption which might result from particular classes of disorder. A detailed description of the normal tissue layers of the vocal fold and of the properties of these tissue types is described in section 1.1.3.

Given this structural framework, it is possible to suggest factors which are likely to be influential in determining the effect of structural change on vocal fold vibration. This section will decribe first the different

types of change in tissue consistency and distribution which can occur within a tissue layer, and then the changes in tissue geometry which can affect the spatial relationships between different tissue layers. A discussion of changes in the physical parameters of rigidity/flexibility, tensile stiffness/elasticity, mass and symmetry and their consequences on acoustic parameters will be followed by a proposed typology of vocal fold pathology relating to these concepts. Brief notes on individual disorders will also be included.

This section offers a somewhat simplified account of the mechanical factors involved in vocal fold pathology, but it is intended only as a tentative framework for the interpretation of acoustic findings.

a. Changes in tissue consistency and in the internal distribution of material within a tissue layer.

The consistency of a tissue can change in a number of ways. One particular instance is the effect of inflammation. This is described in more detail in section 1.2.2, but, in brief, inflammation can involve some or all of the following features: capillary dilation, an infusion of white blood cells, collection of oedematous fluid in the intercellular space, proliferation of collagen fibres and granulation tissue, and the deposition of hyaline. Another instance is keratinization (described below), where, as in the skin-forming process, the epithelium becomes stiffened by the deposition of keratin.

Changes in the distribution of cells within a tissue layer include processes such as hyperplasia (see section 1.2.3 and below), where a multiplication of cells results in a thickening of the layer, often causing the whole tissue layer to become buckled and folded. The density of

cell distribution can also change. Oedematous fluid collection, for example, by expanding the volume of the intercellular space, can cause an effective decrease in both cell and fibre density. Fibre density may also increase, in fibrosis.

b. Changes in the geometrical relationships between tissue layers.

Three kinds of disruption to the geometrical realationships between tissue layers can be described, and these are shown schematically in Figure 2.5.1. The first involves the intrusion of one layer into another, where a boundary is maintained between the two tissue types, and invasion is achieved by displacement. This is characteristic of disorders such as verrucous carcinoma and sessile polyps (see below). The second involves invasion by infiltration, where cells of the invading tissue intermingle with the cells of the invaded tissue. This happens in squamous cell carcinoma (see below). The third kind of disruption occurs when material of one layer penetrates the frontier of another to form a narrow-necked extrusion. This is seen in disorders such as papilloma and pedunculated polyps (see below).

c. Changes in physical parameters and their acoustic consequences.

A survey of various models of vocal fold vibration (Ishizaka and Flanagan 1972, Titze 1973, 1974, Hirano et al. 1982) suggest various factors which should be considered here. These are:

i) rigidity versus flexibility
ii) tensile stiffness versus elasticity
iii) mass
iv) symmetry.

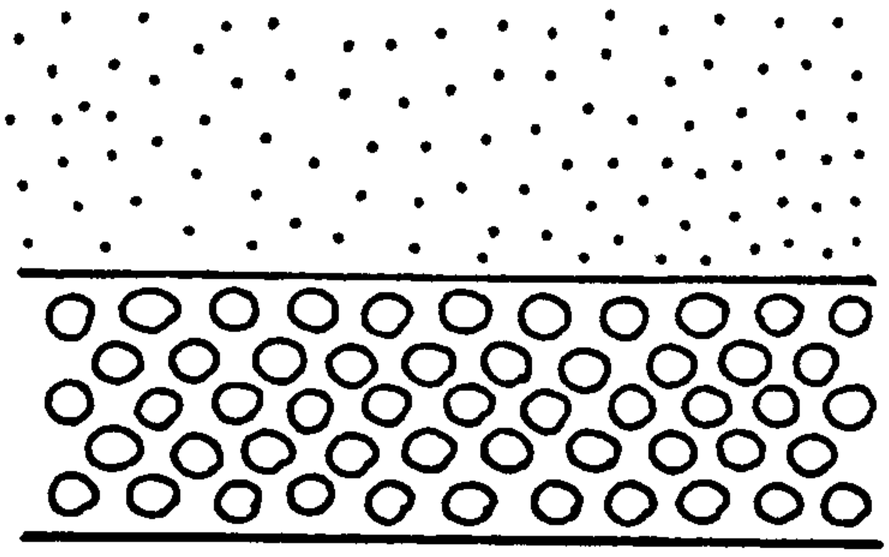


FIGURE 2.5/1: Schematic representation of tissue layer disruption
A. protrusion, B. infiltration,
C. displacement

Rigidity (i.e. resistance to bending) and tensile stiffness (i.e. resistance to stretching) can both, for convenience, be considered under the general concept of "stiffness". This seems to follow Hirano's (1981:52) undefined usage of the term "stiffness", when referring to visual examination of the vocal folds.

A further factor which can influence the acoustic output is the degree of approximation of the vocal folds, since under certain conditions of airflow inadequate approximation may induce turbulence. This would be seen in the acoustic spectrum as interharmonic energy (Laver 1980:121).

Boone (1977:47) organizes voice disorders according to mass/size changes and approximation changes, but these two factors alone allow only a rather vague prediction of phonatory quality. An expansion of this approach to the classification of organic vocal fold disorders might take the following criteria into account:

i) In which tissue layers are there structural alterations?
ii) Do these alterations involve a significant change in mass?
iii) Do they involve a significant change in stiffness?
iv) Is there a protrusion of any mass into the glottal space which might impede vocal fold approximation or cause turbulent airflow?
v) Is the structural alteration symmetrical, affecting both vocal folds equally?
vi) Are the normal geometric relationships between the different tissue layers maintained?

These criteria can be related to a number of phonatory consequences. Hirano (1981:52-53) mentions some of these

in his comments on the interpretation of strobolaryngoscopic examination. His guidelines can be briefly summarised:

i) Increased mass tends to decrease fundamental frequency and amplitude.
ii) Increased stiffness tends to increase fundamental frequency, decrease amplitude, and prevent full approximation of the vocal folds. It also inhibits the action of the mucosal wave.
iii) Localized protrusion of any mass into the glottal space will interfere with approximation of the vocal folds.
iv) Asymmetry of mass, configuration or consistency will cause dysperiodic vibration, as will any localized mass or stiffness change.

The rationale underlying these guidelines deserves some elucidation, but it is probably more useful to give a generalized summary than to embark upon detailed physical and mechanical calculations at this point.

i) Mass

An increase in mass adds inertial force to the vocal fold, which will tend to decrease the speed of oscillation. It may be expected to exert its effect most strongly at the onset of phonation, when the vocal fold is accelerating from a relatively stationary position. The influence of mass on amplitude is less straightforward, and it should be noted that Hirano et al. (1981) contradict the above guideline, where they suggest that a larger mass should increase both speed and amplitude of vocal fold excursion. The theoretical link between mass and decreased fundamental frequency, at least, does seem to be borne out by one study. Oedematous increases in mass, as associated for example with chronic laryngeal inflammation, should be expected to show a

lower fundamental frequency. Fritzell et al. (1982) demonstrate that this is in fact the case.

The precise location of any increase in mass also needs to be taken into account. A local increase in mass will cause the greatest inertial resistance to vocal fold displacement when it is close to the point of maximum excursion, i.e. close to the longitudinal midpoint, and near the surface of the fold.

ii) Stiffness

It is reasonable to expect that increasing the stiffness of a vibrating body should inhibit the vibratory movement, causing a decrease in amplitude of excursion. The mucosal wave, which is visible during normal vocal fold vibration, is a travelling wave in the mucosal layer (see section 1.1.3). This requires that the superficial layer of the lamina propria should behave as a semi-fluid, relatively independent of the deeper tissues. Increased stiffness of this layer, or of the epithelial layer (as in keratosis) will therefore tend to limit the mucosal wave. Changes in stiffness of the underlying tissue layers would not neccessarily have the same effect. Since the mucosal wave is thought to be a necessary part of normal regular vibration (see Section 1.1.3), its inhibition may increase perturbation levels.

iii) Protrusion

Protrusion of a mass into the glottal space will only interfere with vocal fold approximation if it is relatively localized. A uniform swelling along the whole length of the vocal fold may actually improve approximation, as seems to be the experience of some speakers with mild inflammation of the vocal folds during upper respiratory tract infections. A distinction must therefore be drawn between localized and non-localized protrusions. An example of a localized protrusion is a vocal polyp, which may become wedged between the vocal folds, thus preventing the folds from meeting.

In considering the effects of localized protrusions, the site and type of attachment of the protruding body need also to be considered. A protrusion which restricts approximation only of the cartilaginous portion of the vocal folds will have a rather different effect from one which prevents approximation of the ligamental portion. Protrusions with flexible, stalk-like attachments, such as are sometimes found in pedunculated polyps and papillomata, may be displaced by the transglottal airflow, thus causing only intermittent obstruction.

iv) Asymmetry

Asymmetry of vocal fold structure may cause the two vibrating folds to move out of phase with one another, with complex implications for the acoustic waveform. Asymmetry will therefore disrupt the fine co-ordination between airflow and vocal fold configuration, causing perturbations in the laryngeal waveform. Structural asymmetry is a feature of many laryngeal pathologies, including carcinoma, most vocal polyps, and papillomata.

v) Tissue layer integrity

In addition to the above comments, the degree of integrity of the tissue layers which make up the vocal folds needs to be taken into account. A degree of independent behaviour of the body and covering tissues is important in determining the fine detail of phonatory vibration (Smith 1961, Perello 1962). Any loss of integrity between the tissue layers can therefore be expected to affect vibratory patterns by limiting this independence. It is likely that such disturbance will result in increased perturbation levels.

d. Proposed typology of vocal fold disorders

To be susceptible to acoustic registration, a vocal disorder must show either a structural or a functional change from the characteristics of the healthy, normal larynx. This typology will concentrate on organic pathologies only, where the disorder involves a structural alteration of the vocal fold. Further, it will include only the more commonly encountered organic pathologies. Phonatory problems which arise in the absence of any structural alteration will not be considered in any detail. These include neuromuscular disorders, such as paralysis of the vocal fold musculature, as well as a range of psychogenically induced disorders where there is no organic change. Since disorders of the true vocal fold are the most likely to have direct consequences for phonation, sub-glottic and supra-glottic disorders will not be included, although it is acknowleged that such disorders may, under some circumstances, interfere with vocal fold function. The scope of the typology is further limited by the exclusion of all disorders which are specific to childhood. The reasons for this are two-fold. The first relates simply to the needs of the study described in this thesis, which used speech samples drawn exclusively from an adult population. The second reason is that the mature layered structure of the larynx is not fully developed until after puberty (see Section 1.2.4).

An examination of the literature on vocal fold pathology reveals that classification of disorders usually uses criteria related either to the underlying pathology, or to the presumed aetiology. The term "pathology" is used here to describe processes acting within the tissues in the development of a disorder, such as inflammation or neoplastic change (see sections 1.2.2 and 1.2.3). The term "aetiology" can then be reserved for factors which

arise externally to the tissues, as in infection or mechanical abuse of the tissues.

The overiding concern of the medical profession is, quite properly, to identify the pathological processes involved in a given disorder, since these play a large part in determining the most appropriate treatment. The medical literature is therefore typified by classifications based on the underlying pathology, such as that shown below:

1. Inflammatory conditions
   i. acute
   ii. chronic

2. Neoplasms (tumours)
   i. benign
   ii. malignant

3. Congenital malformations

4. Traumatic injury

(e.g. Hall and Colman 1975, Birrell 1977).

There are some demarcation difficulties with this approach, in that there is no clear agreement about the borderline between chronic inflammatory conditions and some benign tumours. Vocal polyps, for example, are considered by some authors to be inflammatory in origin (New and Erich 1938, Arnold 1962, Friedmann and Osborn 1978), and by others to be instances of benign tumours (Birrell 1977).

The speech therapy literature is understandably more concerned with the extent to which poor phonatory habits may be involved in the aetiology of a vocal fold lesion. Hence, a distinction is often drawn between those

disorders which arise apparently independently of any vocal misuse, versus those which are considered to be the sequel of faulty habitual phonation. The latter type are often called "functional" or "psychogenic" disorders (Luchsinger and Arnold 1965, Greene 1972, Perkins 1977), in contrast to the former group of "organic" disorders.

This approach also has a demarcation problem. There seems to be general agreement that vocal nodules, for example, are "functional", in that they arise most often in speakers who habitually misuse their vocal folds. They may therefore be classified with disorders like conversion aphonia (hysterical loss of voice) or spastic dysphonia (extreme adductive compression of the vocal folds), which exhibit no obvious structural abnormality. Vocal nodules are, however, clearly "organic", in the sense that there is an easily observable structural abnormality of the vocal folds. When fully developed, they may even be indistinguishable, both macroscopically and histologically, from certain types of tumour (Shaw 1979). It is also very difficult to disentangle the relative contributions of "organic" predisposition and "functional" misuse in the causation of many disorders. Arnold (1962) considers the role of various predisposing factors in vocal nodule formation, and even in this most "functional" of vocal fold lesions it seems that general bodily health and infection may play an important part.

The proposed system of classification which is outlined here tries to utilise the predictive features summarised above. This is not intended to be the definitive solution to the problem of devising a phonatory classification of organic vocal fold pathology. It should be seen rather as a preliminary attempt to highlight some of the mechanical factors which must be considered in order to predict the vibratory and acoustic characteristics of any disorder. This will serve as a basis for discussion of the acoustic

results of the study which will be reported later in this chapter.

Figure 2.5/2 summarises the proposed classification system. The primary bases for classification are the geographical site of the disorder, and the tissue layer(s) involved. The organic vocal fold pathologies which are most commonly described in the medical literature are listed in Figure 2.5/3, and each is assigned a classificatory code which corresponds with the codes shown in the previous figure. Brief descriptions of these disorders can be found below. This is by no means an exhaustive list of all the disorders which involve structural changes of the larynx, but it should serve to give some idea of the possibilities and limitations of an acoustic screening procedure for detecting vocal fold pathologies.

Allocations of disorders to categories within this framework are often tentative, because it has not been easy to gather sufficient histological details about all the disorders mentioned. It should also be stressed that such a framework does not always necessarily relate directly to medical and pathological considerations, and this potentially limits the usefulness of acoustic techniques as an aid to clinical decision making. For example, the differing structures and mechanical properties of vocal polyps and polypoid degeneration demand that they be given different classificatory codes in this system, because we predict that they will have different acoustic characteristics. They may, however, both be seen as forms of chronic inflammatory reaction to chemical or mechanical irritation, thus sharing common underlying aetiology and pathology (Luchsinger and Arnold 1965, Boone 1977, Aronson 1980).

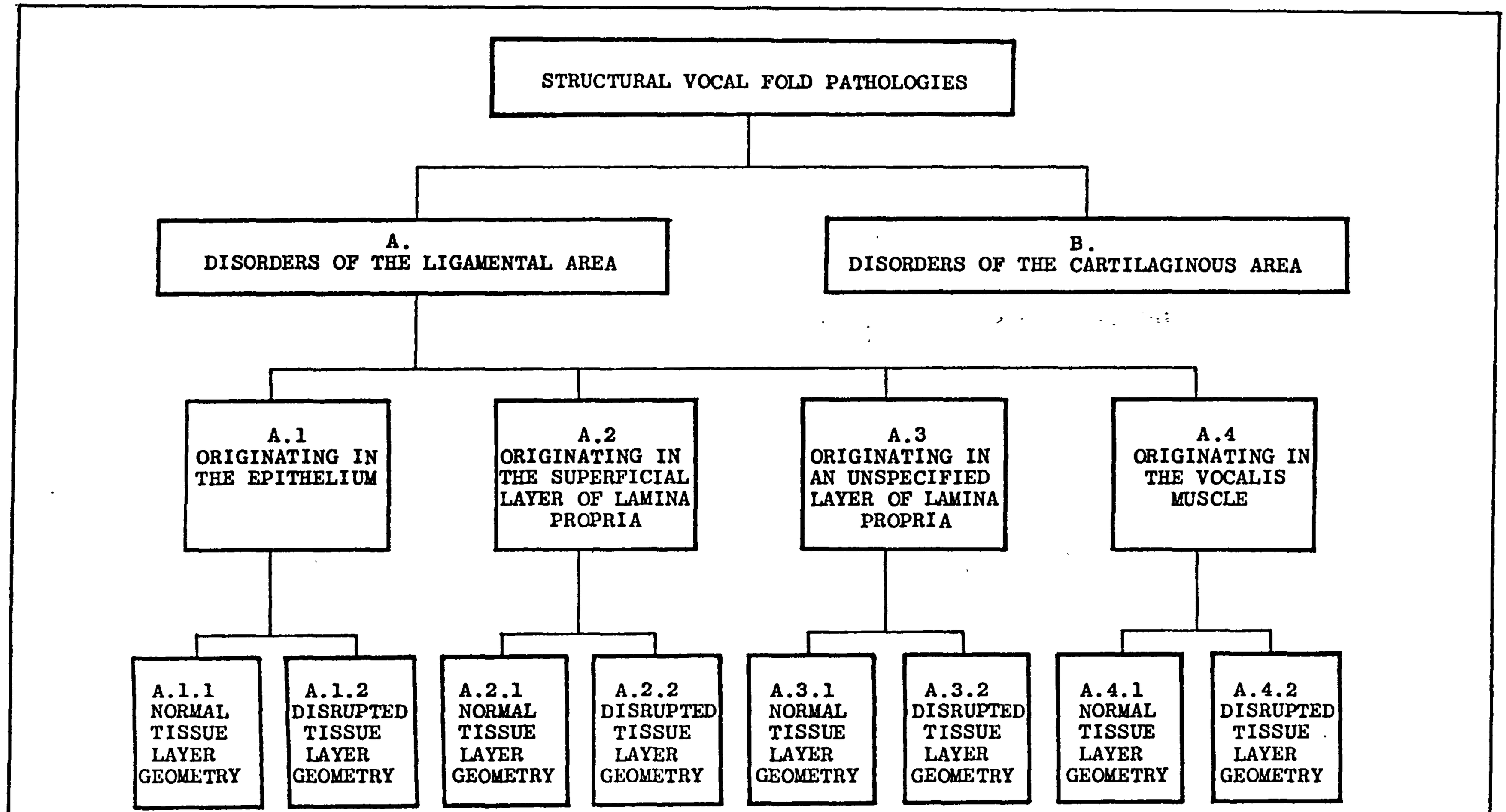


FIGURE 2.5/2, A: Classification system for vocal fold pathology; ligamental area

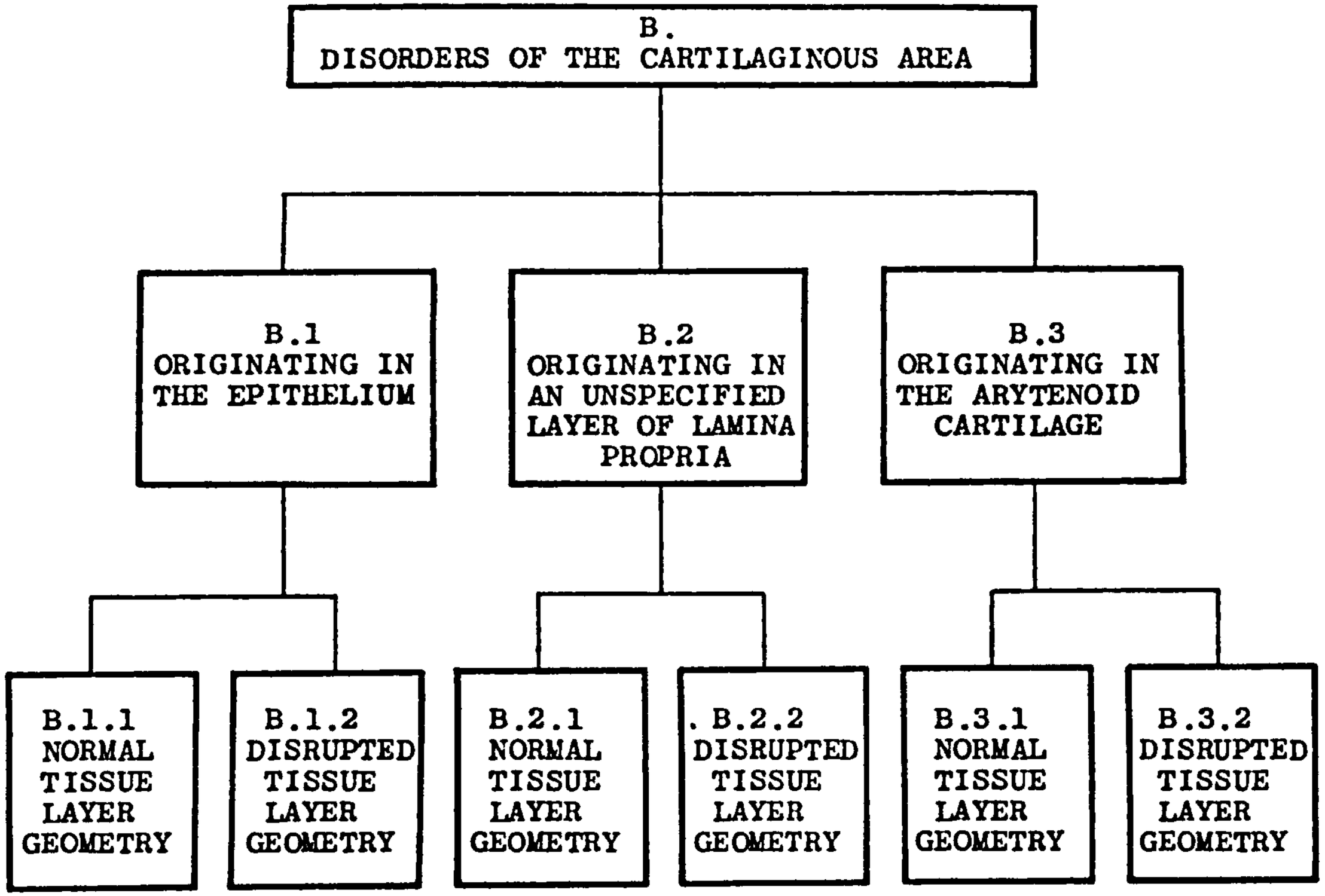


FIGURE 2.5/2, B: Classification system for vocal fold pathologies; cartilaginous area

A. Disorders of the ligamental portion

A.1. Disorders originating in the epithelium

A.1.1. Normal tissue layer geometry

Hyperplasia
Keratosis
Carcinoma-in-situ

A.1.2. Disrupted tissue layer geometry

Squamous cell carcinoma
Verrucous carcinoma (a specific form of squamous cell carcinoma)

A.2. Disorders originating in the superficial layer of the lamina propria

A.2.1. Normal tissue layer geometry

Reinke's oedema

A.3. Disorders originating in any unspecified layer of the lamina propria

A.3.1. Normal tissue layer geometry

Vocal nodules
Sessile vocal polyps
Acute laryngitis
Chronic laryngitis
Chronic hyperplastic laryngitis
Fibroma

A.3.2. Disrupted tissue layer geometry

Pedunculated polyp

A.4. Disorders originating in the vocalis muscle

A.4.1. Normal tissue layer relationships

Sarcoma

B. Disorders of the cartilaginous portion

B.2. Disorders originating in any unspecified layer of the lamina propria

B.2.1. Normal tissue layer geometry

Acute oedema

B.2.2. Disrupted tissue layer geometry

Contact ulcer

FIGURE 2.5/3: Structural vocal fold pathologies arranged according to the classification system outlined in Figure 2.5/2

As is often the case with classificatory systems, the divisions laid out in Figure 2.5/2 are also to some extent over-specific, in that they imply a rather more orderly situation than exists in reality. Many disorders show so much variation in form, in different individuals or at different stages in their development, that they could have been allocated to more than one category. The proposed framework imposes somewhat artificial boundaries in these cases, but the allocation to categories attempts to reflect the most characteristic form of each disorder. In clinical practice, classification would have to be based on careful observation of each individual's presentation.

The combination of disorders originating in any of the three separate layers of the lamina propria into one overall category (see categories A3 and B2) is suggested because medical writers are often not specific about which layers are involved in an organic change. It may be that such distinctions are of little medical significance, because of the lack of any biological boundaries between these layers, even though there are possible consequences for the details of vibratory pattern. Further examination of individual cases could allow a more detailed categorization.

Figure 2.5/4 summarises pathologies in terms of the presence or absence of mass and stiffness changes, protrusion into the glottal space, symmetry, and tissue layer geometry. An important point emerges from this, concerning the potential power of acoustic screening to differentiate between disorders. Some clinically separable disorders may be expected to impose rather similar mechanical constraints on vibration, and hence on acoustic output, so that they are unlikely to be easily separable by an acoustic assessment procedure alone. An example of this is the grouping of papilloma, squamous

PATHOLOGY	Altered Fø		Increased perturbation
A. LIGAMENTAL PORTION A.1. EPITHELIAL			
A.1.1. Hyperplasia	Low		+
A.1.1. Keratosis		(+)	+
A.1.1. Carcinoma-in-situ	?	(+)	+
A.1.2. Squamous carcinoma	?	+	+
A.1.2. Verrucous carcinoma	?	+	+
A.1.2. Adult papilloma	?	+	+
A.2. SUPERFICIAL L.P.*			
A.2.1. Reinke's oedema	Low		
A.3. UNSPECIFIED L.P.*			
A.3.1. Vocal nodules	Low	+	(+)
A.3.1. Vocal polyps (sessile)	?	+	(+)
A.3.1. Acute laryngitis	Low		
A.3.1. Chronic laryngitis	Low		
A.3.1. Chronic hyper-plastic laryngitis	?		
A.3.1. Fibroma	?	+	+
A.3.2. Vocal polyps (pedunculated)	?	+	+
A.4. VOCALIS MUSCLE			
A.4.1. Sarcoma	?Low		+
B. CARTILAGINOUS PORTION B.1. EPITHELIAL (as under A.1.) B.2. UNSPECIFIED L.P.*			
B.2.1. Acute oedema	Low	+	
B.2.2. Contact ulcer	?	+	(+)

L.P.* = lamina propria

(+) = possible or variable presence

+ = presence of a factor

FIGURE 2.5/4B: A summary of predicted acoustic characteristics of vocal fold pathologies

PATHOLOGY	Disrupted tissue layer geometry	Mass change	Stiff-ness change	Protru-sion	Asymmetry
A. LIGAMENTAL PORTION A.1. EPITHELIAL					
A.1.1. Hyperplasia		+			+
A.1.1. Keratosis		(+)	+	(+)	+
A.1.1. Carcinoma-in-situ		+	+	(+)	+
A.1.2. Squamous carcinoma	+	+	+	+	+
A.1.2. Verrucous carcinoma	+	+	+	+	+
A.1.2. Adult papilloma	+	+	+	+	+
A.2. SUPERFICIAL L.P.*					
A.2.1. Reinke's oedema		+		N.L.	
A.3. UNSPECIFIED L.P.*					
A.3.1. Vocal nodules		+		+	(+)
A.3.1. Vocal polyps (sessile)	(+)	+	+	+	(+)
A.3.1. Acute laryngitis		+		N.L.	
A.3.1. Chronic laryngitis		+		N.L.	
A.3.1. Chronic hyper-plastic laryngitis		+	+	N.L.	
A.3.1. Fibroma		+	+	+	+
A.3.2. Vocal polyps (pedunculated)	+	+	+	+	(+)
A.4. VOCALIS MUSCLE					
A.4.1. Sarcoma		+	?		+
B. CARTILAGINOUS PORTION B.1. EPITHELIAL (as under A.1.) B.2. UNSPECIFIED L.P.*					
B.2.1. Acute oedema		+		+	
B.2.2. Contact ulcer	+	+	+	+	(+)

L.P.* = lamina propria
+ = presence of a factor
(+) = possible or variable presence
N.L. = non-localised protrusion, not expected to prevent vocal fold approximation

FIGURE 2.5/4A: A summary of the mechanical characteristics of vocal fold pathologies

carcinoma and verrucous carcinoma, all of which may show an increase in mass and stiffness originating in the epithelium, with protrusion into the glottis and altered tissue layer geometry. Whilst the pattern of invasion by the tumour cells is rather different in each disorder, this is unlikely to have a clearly differentiable acoustic effect until invasion is rather advanced, if at all.

2.5.3 ORGANIC VOCAL FOLD PATHOLOGIES

This section includes brief notes on the individual vocal fold pathologies classified above.

Inflammation

Many of the disorders described here involve some degree of inflammation. This may play a major role in the development of a disorder, as in the various forms of chronic laryngitis, or it may occur as a secondary peripheral response, like that seen in the tissues adjacent to an advancing verrucous carcinoma (Ferlito 1974). The descriptions of individual disorders may therefore be more fully understood if the basic characteristics of the inflammatory process are remembered. This process is described in section 1.2.2.

A. Disorders of the ligamental area of the vocal fold

A.1 Disorders originating in the epithelium

Terminology

A survey of epithelial disorders is complicated by the lack of a standardized terminology to describe some common types of structural disorder within the epithelium. The terms "hyperplasia", "keratosis", "hyperkeratosis" and "leucoplakia" seem all to be applied

to a rather similar group of epithelial conditions which are thought to be aggravated by prolonged chemical or mechanical irritation. The common link between these conditions is the presence, in varying balance, of two types of structural change. The first, which I shall call hyperplasia, is a simple increase in cell number resulting from excessive cell division (see section 1.2.3). The second, keratosis, is the inappropriate formation of keratin within the tissue. These two processes are described as separate disorders below, but they do commonly occur in combination. It is assumed that alone, or in combination, hyperplasia and keratosis cover all the labels listed at the beginning of this paragraph.

There is considerable controversy over the question of whether or not these conditions should be considered as precursors of malignant change. As long as individual cells appear to have normal structure there is no evidence of malignancy, but there does seem to be a continuum from simple hyperplasia and keratosis, where cell structure is normal, to carcinoma-in-situ, where a large proportion of cells are abnormal in structure and malignancy must be suspected (see section 1.2.3). Differential diagnosis is therefore often highly problematic. (Saunders 1964, Hall and Colman 1975, Birrell 1977, Friedmann and Osborn 1978)

A.1.1 Hyperplasia

Tissue of origin: epithelium.
Mechanical factors: an asymmetric increase in mass, with normal tissue layer geometry.
Site of occurrence: anywhere within the laryngeal epithelium. Common at the centre of the ligamental area of the vocal fold.

Hyperplasia is an increase in cell number, resulting from rapid division of the basal cell layer. The disproportionate increase in basal cell number may cause buckling and distortion of the basement membrane, but the stratified arrangement of the cells is maintained, and the cells appear normal.

A.1.1 Keratosis

Tissue of origin: epithelium.
Mechanical factors: an asymmetric increase in stiffness, with normal tissue layer geometry. Eventually there may be a significant increase in mass and protrusion into the glottal space.
Site of occurrence: as for hyperplasia.

Keratosis is a condition in which the squamous cells of the epithelium begin to produce keratin, which is laid down as a horny layer at the surface of the epithelium. It may form a large, whitish mass, which protrudes into the glottal space and can then interfere with vocal fold approximation. Smoking seems to be a major aetiological factor in the development of keratosis. (Auerbach, Hammond and Garfinkel 1970)

A.1.1 Carcinoma-in-situ

Tissue of origin: epithelium.
Mechanical factors: an asymmetrical increase in mass, with normal tissue layer geometry. Variable increase in stiffness and protrusion into the glottal space.
Site of occurrence: anywhere within the laryngeal epithelium.

Carcinoma-in-situ is usually regarded as the earliest recognisable stage of cancer of the larynx, although it is not an inevitable precursor of invasive cancer, and not all cases of carcinoma-in-situ necessarily progress to become fully invasive. The difficulty of making a differential diagnosis between simple hyperplasia, keratosis and carcinoma-in-situ has already been mentioned. This is because carcinoma-in-situ always involves hyperplasia and it may also co-exist with some degree of keratosis. The feature which sets carcinoma-in-situ apart, and which indicates the onset of malignancy, is the presence of a high proportion of abnormal cells, and the loss of the normal orderly arrangement of cells within the epithelium. This disorder therefore displays a histological pattern of haphazardly dividing cells which may have quite bizarre structure. The abnormality spreads laterally within the epithelium, but the basement membrane seems to act as a barrier, preventing spread into the lamina propria. The lamina propria may, however, be inflamed. (Auerbach, Hammond and Garfinkel 1970, Bauer and McGavran 1972, Ferlito 1974, Friedmann and Osborn 1978)

A.1.2 Squamous cell carcinoma

Tissue of origin:	epithelium.
Mechanical factors:	an asymmetrical change in mass and stiffness, with disrupted tissue layer geometry and protrusion into the glottal space.
Site of occurrence:	anywhere within the larynx. Most common in the ligamental portion of the vocal fold.

The commonest type of laryngeal tumour is carcinoma arising in the squamous epithelium. Carcinomatous change is characterized by a loss of the normal control of

epithelial cell division (see section 1.2.3). The epithelial cells divide at an abnormal rate and form a disorderly mass. The cells are recognised as being malignant by their abnormal structure, and by their tendency to infiltrate not just the surrounding epithelial tissue, but also the underlying tissues. Squamous cell carcinomas vary greatly in their structure, and in their pattern of invasion, so that it is difficult to generalise about their expected mechanical correlates. An increase in mass is almost always found, except in some cases with ulceration. Ulceration may occasionally expose and destroy even the laryngeal cartilages, so that a considerable amount of tissue is lost. Stiffness depends on cell density, and on the extent of concomitant keratosis, both of which are very variable. The size of the lesion may also fall within a wide range. Some specific forms of squamous carcinoma are recognised, one of which is described below (verrucous carcinoma). (Ferlito 1974, Michaels 1976, Friedmann and Osborn 1978, Shaw 1979)

A.1.2 Verrucous carcinoma (= a specific type of squamous carcinoma)

Tissue of origin: epithelium.
Mechanical factors: an asymmetrical increase in stiffness and mass, with localised protrusion into the glottal space, and disrupted tissue geometry.
Site of occurrence: anywhere within the larynx. Commonest in the ligamental portion of the vocal fold.

This tumour is a specific type of squamous cell carcinoma, which presents as a slowly growing warty mass, and may be multicentric. The epithelium becomes hyperplastic and highly keratinized, with folds and

finger-like protrusions extending deep into the lamina propria. Epithelial pearls (= dense deposits of keratin) may develop, forming localized areas of extreme stiffness. Verrucous carcinoma is of relatively low malignancy, and advances by displacement of cells rather than by infiltration. Adjacent tissue usually shows a marked inflammatory response. The tumour may grow large enough to cause dysphagia (swallowing difficulty) and respiratory obstruction. (Ferlito 1974, Michaels 1976, Friedmann and Osborn 1978, Maw et al. 1982)

A.1.2 Adult form of papilloma

Tissue of origin:	epithelium.
Mechanical factors:	an asymmetrical increase in mass and stiffness, with disrupted tissue layer geometry and localised protrusion into the glottal space.
Site of occurrence:	commonest at the edge of the ligamental portion of the vocal fold or at the anterior commisure.

Papilloma is a benign warty tumour which, in adults, forms multiple branch-like projections of highly keratinized epithelium. There may be extrusion of thin columns of lamina propria into the tumour, so that tissue layer geometry is substantially disrupted. Papillomata are usually unilateral and solitary, and most are pedunculated. These growths are not common in adults, and their medical significance derives from reports that a small proportion of papillomata undergo malignant transformation. (Hall and Colman 1975, Birrell 1977, Friedmann and Osborn 1978, Shaw 1979)

A.2 Disorders arising in the superficial layer of the lamina propria

A.2.1 Reinke's oedema (polypoid degeneration, chronic oedematous laryngitis)

Tissue of origin: superficial layer of the lamina propria.

Mechanical factors: a symmetrical mass increase with non-localised protrusion into the glottal space. Tissue layer geometry is normal, but with weakened adherence between layers.

Site of occurrence: both vocal folds usually affected along their full length.

Reinke's oedema is a specific form of chronic laryngitis which is characterized by a loosening of the attachment between tissue layers in the ligamental portion of the vocal fold. This allows oedematous collection of fluid along the full length of the vocal fold. The overlying epithelium is normal, or only slightly hyperplastic, and if fluid is allowed to drain away, the lamina propria appears to be relatively normal. Only in long standing cases does mild hyperaemia occur. Reinke's oedema is a disorder of middle age, which seems to be exacerbated by smoking and alcohol. It is interesting that clinical descriptions of Reinke's oedema suggest similarities with the normal age related changes described by Hirano et al. (1982 - see section 1.2.4). One of the main vocal symptoms is a decrease in fundamental frequency, consistent with the mass increase. (Saunders 1964, Luchsinger and Arnold 1965, Kleinsasser 1968, Birrell 1977, Friedmann and Osborn 1978, Salmon 1979, Aronson 1980, Fritzell, Sundberg and Strange-Ebbeson 1982)

A.3.1 Vocal nodules (early stage)

Tissue of origin:	lamina propria (probably the superficial layer).
Mechanical factors:	a symmetrical or asymmetrical increase in mass, with localised protrusion into the glottal space and normal tissue layer geometry. Stiffness is increased only slightly.
Site of occurrence:	usually at the edge of the vocal fold in the centre of the ligamental portion.

Vocal nodule formation is thought usually to be precipitated by local mechanical trauma. The first stage is probably a haemorrhage of the small blood vessels within the lamina propria, which is followed by a localised inflammatory response. The nodules appear as soft, red swellings, and they are commonly bilateral, at the centre of the ligamental portion of the vocal fold. Nodules may recover spontaneously if further mechanical abuse of the larynx is avoided. If they become established fibrosis, epithelial hyperplasia or capillary proliferation may occur, creating a much firmer growth. There is some disagreement about the relationship between vocal nodules and vocal polyps (see below). Some writers consider polyps to be chronically established nodules which have undergone a late stage inflammatory change, so the following section on polyps can be taken to represent a later stage in nodule development. (Arnold 1962, Luchsinger and Arnold 1965, Michaels 1976, Perkins 1977, Boone 1978, Friedmann and Osborn 1978, Salmon 1979, Aronson 1980)

A.3.1 Sessile vocal polyps

Vocal polyps may be sessile (broad based) or pedunculated (stalked). Pedunculated polyps have disrupted tissue layer geometry, and must therefore be placed in the category A.3.2. Histological characteristics of both forms are, however, similar, so they will be discussed together below.

Tissue of origin:	lamina propria (probably the superficial layer).
Mechanical factors:	an asymmetrical (or rarely symmetrical) increase in mass and stiffness, with localised protrusion into the glottal space. Tissue layer geometry is significantly disrupted only if the growth is pedunculated.
Site of occurrence:	usually at the edge of the ligamental portion of the vocal fold.

Long term mechanical abuse of the vocal folds may result in the establishment of localised chronic inflammatory changes. These appear as small, stiff swellings on the edge of the vocal fold, which may be unilateral or bilateral. In bilateral cases the swellings are seldom the same size, so that true symmetry is rare. The extent and constancy of protrusion into the glottal space varies, because polyps may be sessile or pedunculated. Stiffness depends on the histological make-up of the individual polyp. Some are predominantly fibrotic, with a dense, disorganised network of collagen fibres, and this type may eventually develop patches of hyalinization. Others are built up largely from vascular tissue, and may be much less stiff than the fibrotic type. The epithelium overlying a polyp may also become hyperplastic. (Arnold 1962, Luchsinger and Arnold 1965, Kleinsasser 1968, Greene 1972, Hall and Colman 1975, Michaels 1976, Birrell

1977, Boone 1977, Perkins 1977, Friedmann and Osborn 1978, Salmon 1979, Aronson 1980)

A.3.1 Acute laryngitis

Tissue of origin: lamina propria.
Mechanical factors: a symmetrical increase in mass, with normal tissue layer geometry. Approximation may be limited by associated oedema affecting the cartilaginous area of the fold.
Site of occurrence: the whole of the larynx may be involved.

Acute laryngitis, which may have many causes, including infection, sudden irritation or mechanical abuse, shows all the features of a generalized acute inflammation. There is hyperaemia throughout the larynx, and infiltration of leucocytes, so that the vocal folds appear to be rounded and thickened in cross section. The swelling due to oedema is usually most marked in the mucous membrane covering the arytenoids (see B.2.1; acute oedema), so that approximation of the ligamental area of the vocal folds may be prevented. In severe cases the epithelium may become necrotic (necrosis = localized tissue death), and ulceration results as the dead tissue is sloughed off. The underlying muscle may also become inflamed. (Luchsinger and Arnold 1965, Hall and Colman 1975, Birrell 1977, Boone 1977, Friedmann and Osborn 1978, Salmon 1979, Aronson 1980)

A.3.1 Chronic laryngitis

Tissue of origin: lamina propria.
Mechanical factors: a symmetrical increase in mass, with non-localized protrusion into the glottal space, and normal tissue

layer geometry.
Site of occurrence: the whole larynx may be involved.

Chronic inflammation of the larynx shows various forms. The simplest presentation includes hyperaemia and swelling, with an increase in mucous secretions covering the folds. In severe cases the inflammatory response may involve the vocalis muscle. Chronic laryngitis may be a response to long standing exposure to irritants such as dust or smoke, or to habitual vocal abuse and misuse. Other forms of chronic laryngitis are described elsewhere (see Reinke's oedema and chronic hyperplastic laryngitis). (Saunders 1964, Hall and Colman 1975, Friedmann and Osborn 1978, Aronson 1980)

A.3.1 Chronic hyperplastic laryngitis (chronic hypertrophic laryngitis)

Tissue of origin: lamina propria.
Mechanical factors: a symmetrical increase in mass and stiffness, with non-localized protrusion into the glottal space, and normal tissue layer geometry.
Site of occurrence: the whole larynx may be involved.

Some authors differentiate a type of chronic laryngitis which is characterized by a generalised hyperplasia of the epithelium, and in terms of mechanical factors it makes sense to follow this example here. The vocal folds are swollen and hyperaemic, as in other forms of laryngitis, but this is associated with changes in the overlying epithelium. The ciliated epithelium above and below the vocal fold becomes hyperplastic, and takes on a squamous pattern, whilst the squamous epithelium at the edges of the vocal folds becomes keratinized. The vocal folds become progressively more irregular and swollen,

and may appear very dry. (Kleinsasser 1968, Birrell 1977, Salmon 1979)

A.3.1 Fibroma

Tissue of origin: lamina propria.
Mechanical factors: an asymmetrical increase in mass and stiffness, with localised protrusion into the glottal space, but no significant disruption of tissue layer geometry.
Site of occurrence: anywhere within the larynx. Most common on the edge of the ligamental portion of the vocal fold.

This rare, benign tumour usually presents as a smooth, sessile body on the edge of the vocal fold. It contains a network of collagen fibres, and may be difficult to distinguish from a fibrous polyp. (Birrell 1977, Perkins 1977, Shaw 1979)

A.3.2 Pedunculated polyp

See earlier (A.3.1 - sessile vocal polyps)

A.4 Disorders originating in the body of the vocal fold

A.4.1 Sarcoma

Tissue of origin: vocalis muscle or lamina propria.
Mechanical factors: an asymmetrical increase in mass.
Site of occurrence: not specified.

Sarcoma is a very rare type of malignant tumour, which may affect connective tissue and muscle. Sarcoma arising from the vocalis muscle is one of the very rare disorders (excluding atrophy due to paralysis) which originates in

the body of the vocal fold. The rather brief comments in the references below allow only tentative suggestions about mechanical correlates. (Friedmann and Osborn 1978, Shaw 1979)

B. Disorders of the cartilaginous area of the vocal fold

B.1 Disorders originating in the epithelium

All of the epithelial disorders already described under A.1 may also affect the epithelium overlying the arytenoid cartilages. Most, however, are very much more common in the ligamental area.

B.2 Disorders originating in the lamina propria

B.2.1 Acute oedema of the larynx

Tissue of origin: lamina propria.
Mechanical factors: symmetrical mass increase, with non-localised protrusion into the glottal space, and normal tissue layer geometry.
Site of occurrence: the mucosal covering of the arytenoid cartilage.

Oedema is a symptom with many possible underlying causes. These include chemical or thermal irritation, infection, allergy, and cardiac or renal failure. It merits some special comment, however, because of its characteristic distribution. Fluid tends to collect first in the mucosa overlying the arytenoid cartilages, and whilst it may spread upwards to the ventricular folds and epiglottis, the firm adherence of the tissue layers in the ligamental portions of the vocal folds limits its anterior spread. The ligamental area, therefore, tends not to be affected except where chronic oedema leads to loss of tissue layer

adherence in Reinke's oedema. The swelling will usually be symmetrical, and is likely to prevent full approximation of the vocal folds in the unaffected ligamental portion. (Birrell 1977, Friedmann and Osborn 1978, Salmon 1979)

B.2.2 Contact ulcer (contact pachydermia, contact granuloma)

Tissue of origin:	superficial layer of the lamina propria.
Mechanical factors:	an increase in stiffness, with a redistribution of mass, localised protrusion into the glottal space, and disrupted tissue layer geometry. The degree of symmetry is variable.
Site of occurrence:	the mucosa covering the vocal processes of the arytenoid cartilages.

Contact ulcer is generally thought to develop from a localised area of inflammation over the vocal process of the arytenoid cartilage, which is the point of maximum impact during adduction of the cartilages for phonation. A pile of granulation tissue forms, and further excessive impact may cause the centre of this to be worn away until the cartilage is exposed. The result is a central crater, surrounded by an outgrowth of connective tissue and epithelium. The epithelium may be markedly hyperplastic and keratinized. Contact ulcers are usually bilateral, but there is often some discrepancy in size of the ulcers on the two folds. Vocal abuse and psychogenic factors have both been implicated in the aetiology. (Luchsinger and Arnold 1965, Birrell 1977, Boone 1977, Perkins 1977, Salmon 1979, Aronson 1980)

EXPERIMENTAL INVESTIGATION

Acoustic analysis of 116 patients with known pathology of the larynx was carried out as part of the MRC project, and three aspects of the study will be considered in this section. The first question to be asked is whether the speakers with any unspecified laryngeal pathology can be separated from the normal control group on the basis of acoustic characteristics. This question is vital if a screening system is to be developed for the detection of laryngeal pathology. As the introduction to vocal fold abnormalities has indicated, complications are likely to stem from the fact that different organic changes are predicted in many cases to have very different acoustic consequences. The second question which will be addressed therefore concerns the extent to which the predictions about acoustic correlates of specified disorders are met, and hence, the value of acoustic analysis in separating different classes of laryngeal disorder. The final aspect of the research which will be discussed is an evaluation of the potential value of acoustic analysis in tracking changes which may occur during treatment of patients with vocal pathology, whether by surgery, drugs or speech therapy.

Subjects and procedures

The collection of pathological data was carried out with the help of ENT and speech therapy departments at the Royal Infirmary, Edinburgh and the Radcliffe Infirmary, Oxford. Both hospitals conduct outpatient clinics for patients who are referred with voice problems, but the routine and recording equipment used in the two clinics was somewhat different.

In Edinburgh, the author was involved in setting up the voice clinic, which was conducted by Mr. A Maran (ENT

consultant) and Mrs R. Nieuwenhuis (speech therapist). The author attended this clinic, and some general ENT clinics, on a regular basis, and was present during medical examination of most patients. Whenever possible, she also observed the larynx during these examinations. Laryngeal examination of most patients in the Edinburgh group was carried out with a mirror, and no stroboscope, so that only a static view of the larynx was possible. This has marked limitations when information about vocal fold cover stiffness is required. This is because the first sign of increased epithelial stiffness is often the absence or interruption of the mucosal wave (Hirano 1981:53, Harris, personal communication), and this can only be seen if vocal fold movement can be visualized. A small proportion of patients was examined using a flexible nasolaryngoscope. All patients with observable organic abnormalities of the vocal folds were then taken for recording and further case history taking. Full medical notes were made available, and histological reports were included whenever biopsies were performed following the initial examination. Recordings were also made of patients with no observable lesions who were to be referred for speech therapy because of "functional" dysphonia, but the acoustic results of these patients will not be discussed in any detail here.

In Oxford the voice clinic already had a well established routine, involving close collaboration between an ENT registrar (Mr. T. Harris) and a speech therapist (Mrs. S. Collins). Recordings were made of all patients attending the clinic, and they were then examined using a rigid laryngoscope and a stroboscopic light source, so that observations of vocal fold movement during phonation could be made. As in Edinburgh, all relevant medical and histological information was collated.

Unfortunately, the recording conditions and equipment used was rather different at the two centres. In Edinburgh, patients were recorded in a sound treated booth, so that background noise was minimised, using a Dynamic Cardioid microphone (Sennheisser MD421), either a microphone mixer/amplifier (Shure M67-2E) or a balanced pre-amplifier, and a REVOX A77 recorder. The quality of recordings from this centre was uniformly high. In Oxford, it was more difficult to arrange regular access to a sound treated booth, and many recordings had to be excluded from the study because of background noise and interference from other electrical equipment in the hospital. The recording equipment consisted of a similar microphone to that used in Edinburgh, with a portable recorder (Uher 4000 report IC). Differences in acoustic waveform which may have resulted from the different recording equipment were minimised to some extent by appropriate phase compensation prior to application of the pitch detection program (see section 2.1.3).

The final pool of subjects consisted of 60 females and 56 males, with known vocal fold pathology. A breakdown of the group by type of laryngeal pathology is shown in Figure 2.5/5. For comparison, the group of normal speakers described in the previous chapter was used as a control group. Figure 2.5/6 summarises age characteristics and smoking habits of the two groups.

40 seconds of speech was analysed as described in section 2.1.3, and 10 acoustic measures were recorded for each subject (F0-AV, F0-DEV, J-AVEX, J-DEVEX, J-RATEX, J-DPF, S-AVEX, S-DEVEX, S-RATEX, S-DPF). S-DEVEX was found not to be a reliable measure when control group results were examined, and it showed a very non-normal distribution, so that it is excluded from some of the statistical analyses described below.

Results and discussion

A. A comparison between the whole pathology group and controls

Several approaches to the problem of separating these two groups were considered, of which three will be presented here. The first is a simple graphic technique, using bivariate plots, to compare pathological speakers with control group distributions. The second approach uses linear discriminant analysis, which can include data from all acoustic parameters simultaneously. The third statistical method, which can also utilise the results of all acoustic parameters simultaneously, is a pattern recognition technique, based on the maximum likelihood principle.

A simple graphic layout, plotting perturbation measures against fundamental frequency, was initially tested because it has the advantage that an individual's acoustic results can be related to predictions about acoustic consequences of mechanical alterations in the vocal folds (see earlier in this chapter). The value of this type of plot simply to separate pathological speakers from controls needs to be considered.

In order to facilitate the comparison of pathological speakers and controls, all subjects' scores were transformed to z-scores. The deviation of each score from the control group mean was thus expressed as a multiple of the control group standard deviation. The sexes were treated separately in these calculations. Assuming normal distributions (see Section 2.4), two standard deviations on any one parameter should include approximately 90-95% of control subjects. A subject whose score for any parameter deviates from the control group by more than two standard deviations may therefore reasonably be

DISORDER TYPE	MALE	FEMALE
1. Ligamental area disorders		
A. Epithelial disorders		
- hyperplasia	1	1
- keratosis	2	-
- squamous carcinoma	9	-
- verrucous carcinoma	1	-
- papillomata	2	-
B. Lamina Propria disorders		
- polyps/nodules	11	20
- Reinke's oedema	-	5
- acute laryngitis	2	2
- chronic laryngitis	3	4
- hyperplastic laryngitis	-	1
- oedema	2	4
- cysts	-	4
- mild redness, thickening	6	6
2. Cartilaginous area disorders		
- papilloma	-	1
- oedema	-	3
- polyps/nodules	-	4
- contact ulcer	4	-
- chronic inflammation	1	-
- cyst	1	-
3. Vocal fold palsies	11	5
TOTAL	56	60

FIGURE 2.5/5: Laryngeal disorders diagnosed in the pathological subject group

	FEMALES		MALES	
	CONTROL	PATH.	CONTROL	PATH.
NUMBER	80	60	83	56
AGE MEAN	38.1	52.4	36.1	52.0
AGE RANGE	18-84	22-81	18-71	23-82
% SMOKERS*	28.8%	58.3%	43.4%	76.8%

* "Smoker" = anyone reporting a history of regular smoking

FIGURE 2.5/6: Subject group information

* (which would be expected to increase perturbation in the absence of compensatory adjustments),*

considered to be abnormal in some way, and this may be taken as an indication of a risk that laryngeal pathology is present. Figure 2.5/7 shows the numbers of control and pathological speakers who deviate from the control group mean by more than 2 standard deviations for each parameter. This demonstrates that no single acoustic parameter is able to separate more than 76.4% of pathological speakers from the controls, which is inadequate for the purposes of screening. The combination of two parameters in a graphic technique seems to be more effective (Laver et al. 1984, 1986).

Figure 2.5/8 shows the scores of male pathological speakers plotted on a scattergram of F0-AV versus S-DPF. The axes are marked in units of standard deviation, and the origin of both axes corresponds to the control group mean for each parameter. S-DPF was used as the measure of perturbation because it was the best single discriminator between control and pathological speakers for both sexes. A similar scattergram was constructed for female speakers. F0-AV was chosen for the other axis because, although it is a poor discriminator in its own right, there are theoretical reasons why some pathological speakers may be able to maintain normal perturbation levels whilst deviating from normal in their average F0. Firstly, any symmetrical increase in stiffness or mass may result in F0 changes, but not significantly increased perturbation. Secondly, some pathological subjects may be able to maintain normal perturbation levels in the face of asymmetrical lesions, but only at the expense of slightly increased laryngeal tension and hence higher than normal F0. The ellipses around the origins represent the results of principal components analysis of the control group data, which indicates the covariance between S-DPF and F0-AV. The ellipse is drawn at the 2 standard deviation level, and can be used as a screening threshold for the detection of pathology.

	FEMALE		MALE	
	CONTROL (N=58)	PATHOLOGICAL (N=54)	CONTROL (N=63)	PATHOLOGICAL (N=55)
FO-AV	3.4	22.2	4.7	21.8
FO-DEV	8.6	14.8	3.2	12.7
J-DEVEX	3.4	27.7	3.2	25.4
S-DEVEX	3.4	37.0	6.3	45.4
J-AVEX	5.2	29.6	4.7	38.2
S-AVEX	5.2	40.7	9.5	50.1
J-RATEX	5.2	5.5	3.2	5.4
S-RATEX	1.7	20.4	4.7	29.1
J-DPF	3.4	57.4	3.2	61.8
S-DPF	1.7	64.8	3.2	76.4

FIGURE 2.5/7: Table showing the percentage of each group which deviate from the control group mean by more than 2 standard deviations for each of 10 acoustic parameters

(N.B. these figures refer to smaller groups than the rest of the results presented in this section, as they were calculated at an earlier stage of the project)

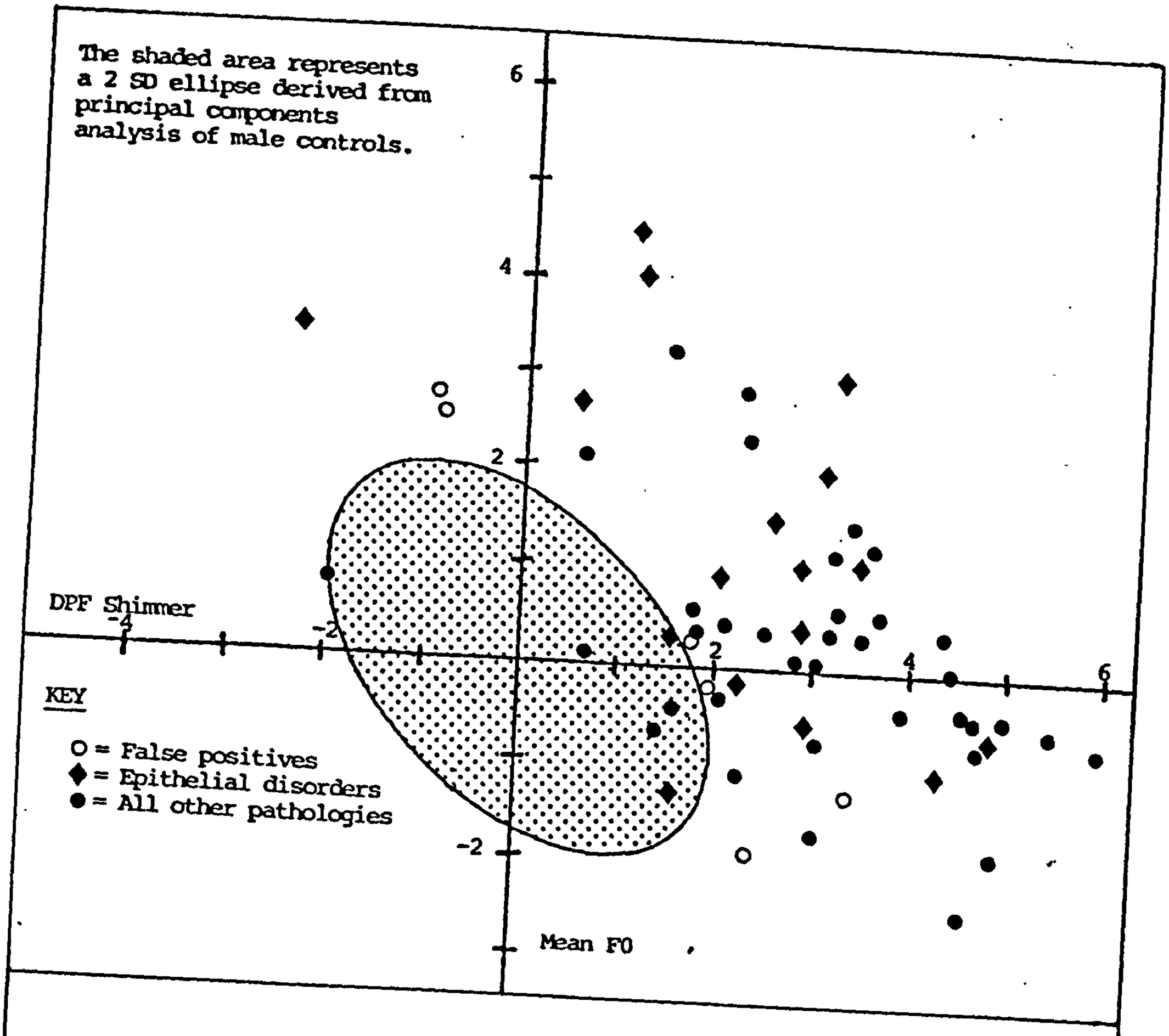


FIGURE 2.5/8: Scattergram of F0 vs. S-DPF: male pathological subjects. The shaded area represents a 2 SD ellipse derived from principal components analysis of male controls

Using this method, 80.35% of the pathological males fall outside the ellipse, compared with 10.8% of the controls. In other words, 80.35% of known pathologies would have been correctly identified as being pathological, whilst 10.8% of the controls would be picked up as "false positives". It must, of course, be stressed that since control subjects were not given laryngeal examinations a proportion of these "false positives" may actually have had minor laryngeal abnormalities. Whilst this is a reasonable success rate, and certainly high enough to suggest that the system has some potential as a screening tool, some serious pathologies were still missed. In medical terms, it is not important if some benign laryngeal disorders are missed, but it is important that all cases of cancer or potentially precancerous states should be detected. These mostly arise in the epithelium, so an ideal screening tool would pick up all changes in the epithelium. When the epithelial disorders were examined separately, 88.2% of cases were correctly identified as being pathological. This will be discussed further in the second half of this section.

The detection rate for pathology in the female group was rather lower, with 76.67% of the pathological group being correctly identified, and 12.5% of the controls being classed as false positives. This lower detection rate may be due to the different distribution of laryngeal pathologies within the male and female pathology groups.

Some alternative bivariate plots were tested to see if better separation of control and pathological speakers could be achieved, and the results are summarised in Figure 2.5/9. Using the two best single discriminators, J-RATEX and S-DPF, together did increase the overall detection rates for pathological speakers to 82.14% for males, and 88.3% for females, but at the expense of also increasing the false positive rates to 14.46% for males

	PERCENTAGE SUBJECTS CLASSIFIED AS PATHOLOGICAL	
	CONTROLS	PATHOLOGICAL
S-DPF vs F0 (♂♂)	10.8	80.4
S-DPF vs F0 (♀♀)	12.5	76.7
J-RATEX vs S-DPF (♂♂)	14.5	82.1
J-RATEX vs S-DPF (♀♀)	12.5	88.3
J-RATEX vs F0 (♂♂)	14.5	62.5
J-RATEX vs F0 (♀♀)	11.3	55.0

FIGURE 2.5/9: Discrimination success of selected bivariate plots

and 12.5% for females. J-RATEX plotted against F0-AV produced much lower detection rates.

It was therefore decided that a more complex statistical procedure which could utilise information about all acoustic parameters simultaneously should be used. The following statistical results are presented with thanks to Edmund Rooney, who was responsible for selecting and running the appropriate programs, and for helping to interpret the results.

Linear discriminant analysis (Klecka 1980) is a technique for discriminating between groups on the basis of several parameters simultaneously. The parameters are weighted and combined to produce a discriminant function which will separate the groups as far as possible. The first step is to see if there is enough difference between the groups on the available parameters to justify the analysis. This is done by calculating Wilks' λ (an inverse measure of group difference), and a X^2 test of statistical significance. A significant Wilks' λ means that the first discriminant function to be derived will itself be statistically significant. The substantive utility of the discriminant function is indicated by its canonical correlation. This is the association between the function and the nominal categories representing the groups present in the data. A canonical correlation of 0.7 or above indicates that the function is discriminating the groups quite successfully. The discriminant function scores for each subject can then be used to classify the subjects, so that the percentage of each group which are correctly classified can then be calculated.

Discriminant functions for each sex were derived from subjects raw scores on all ten parameters, using the DISCRIMINANT subprogram from the Statistical Package for

the Social Sciences (1983). Stepwise analysis of the data was then performed to select an optimal subset of parameters. The results of these analyses are presented in Figure 2.5/10. The detection rates for pathological speakers are 85.7%-87.5% for males and 88.3%-91.7% for females. False positive rates are 6.0%-8.4% for males, and 5.0%-7.5% for females.

These results indicate that linear discriminant analysis is rather more successful as a screening procedure than the bivariate technique, but the results need to be treated with some caution. Linear discriminant analysis assumes that the data show a multivariate normal distribution, but the heterogeneous nature of the pathology group makes this very unlikely. The technique is, however, fairly robust, and the consequence of such violations is probably quite small. A more serious problem is the relatively small numbers of subjects in each group, given that so many parameters are used in the analysis.

The final statistical technique which was applied to these results was a pattern recognition technique (maximum likelihood principle) (Davis 1976). Using 9 acoustic parameters (excluding S-DEVEX because of its non-normal distribution) the detection rates for pathology were 87.5% for males, and 95% for females. These high rates were unfortunately balanced by unacceptably high false positive rates, of 14.3% for males and 25.0% for females. Using an optimal subset of parameters, the false positive rate is reduced to 5.4% for males and 11.7% for females, which is still rather high for the female group. The comparison of statistical results shown in Figure 2.5/11 suggests that discriminant analysis is probably the best screening option using this data.

	PERCENTAGE SUBJECTS CLASSIFIED AS PATHOLOGICAL	
	PATHOLOGICAL	CONTROLS
ALL 10 PARAMETERS (males)	87.5	8.4
ALL 10 PARAMETERS (females)	88.3	7.5
ALL 10 PARAMETERS (males) (LOG SCALE FOR S-AVEX)	87.5	6.0
ALL 10 PARAMETERS (females) (LOG SCALE FOR S-AVEX)	91.7	7.5
OPTIMAL SUBSET (males)	85.7	7.2
OPTIMAL SUBSET (females)	90.0	5.0

FIGURE 2.5/10: Results of linear discriminant analysis

STATISTICAL PROCEDURE	% subjects classified as pathological					
	Controls		Pathological		NP-Dysphonics	
	males	females	males	females	males	females
Bivariate analysis (S-DPF vs F0-AV)	10.8	12.5	80.4	76.7	86.7	88.6
Discriminant analysis						
- all parameters	6.0	7.5	87.5	91.7	-	-
- optimal subset of parameters	7.2	5.0	85.7	90.0	-	-
Pattern recognition (maximum likelihood principle)						
- all parameters	14.3	25.0	87.5	95.0	86.7	94.3
- optimal subset of parameters	5.4	11.7	87.5	90.0	-	-

FIGURE 2.5/11: A comparison of the ability of three statistical procedures to discriminate between control and pathological voices

B. Differential acoustic characteristics of specific laryngeal disorders

The original intention of relating acoustic features to details of vocal fold state was obstructed by two factors. The first was the problem of acquiring good voice recordings and adequate laryngeal observations simultaneously. It has already been mentioned that the clinic which was able to provide consistently good tape recordings was less well equipped to provide detailed information about laryngeal state and movement. Neither clinic was able to provide photographic records of the subjects' larynges. The second factor was that, because so many recordings were not of adequate quality to allow acoustic analysis, the final numbers within each pathology group were too small to allow proper statistical evaluation of acoustic differences. There is, however, enough information available to make some comments on apparent tendencies, and to suggest possible approaches to further study. A few single cases will also be examined in more detail to show how predictions of acoustic behaviour may be related to actual findings.

The imprecise nature of some of the diagnostic information meant that only rather broad classifications of pathology types are possible. Three classes of pathology for each sex will be considered here. These are epithelial disorders, polyps/nodules and disorders of the cartilaginous area for males, and: Reinke's oedema, polyps/nodules, and disorders of the cartilaginous area for females. It will be clear from the introduction to this chapter that these classifications may group together subjects who show unfortunately high levels of heterogeneity in terms of the mechanical characteristics of the vocal folds. The averaged acoustic data, expressed as raw scores and z-scores, for these groups is tabulated in Figure 2.5/12. Figures 2.5/13 and 2.5/14 compare the

A. Males

PARAMETER	EPITHELIAL		POL./NODULES		CARTILAGINOUS	
	RAW	Z-SCORE	RAW	Z-SCORE	RAW	Z-SCORE
FO-AV	134.4	1.62	122.8	0.76	114.5	0.14
FO-DEV	28.3	1.23	22.9	0.32	21.8	0.14
J-AVEX	7.0	1.54	5.6	0.44	5.61	0.42
S-AVEX	19.4	0.38	17.9	0.11	19.9	0.46
J-DEVEX	17.6	0.45	16.2	0.03	16.4	0.09
J-RATEX	32.5	2.28	25.9	0.68	27.8	1.16
S-RATEX	65.6	1.11	67.9	1.50	66.9	1.34
J-DPF	24.2	2.44	20.3	1.26	22.3	1.88
S-DPF	36.0	1.59	39.0	2.16	38.8	2.12

B. Females

PARAMETER	REINKE'S		POL./NODULES		CARTILAGINOUS	
	RAW	Z-SCORE	RAW	Z-SCORE	RAW	Z-SCORE
FO-AV	151.6	-2.20	176.6	-0.96	207.3	0.57
FO-DEV	29.9	-1.34	39.3	-0.18	42.8	0.25
J-AVEX	4.9	0.13	6.3	1.30	4.0	-0.74
S-AVEX	19.0	1.13	20.8	1.56	14.3	0.02
J-DEVEX	15.2	0.16	16.9	0.91	12.98	-0.83
J-RATEX	22.1	0.38	27.1	1.46	16.7	-0.81
S-RATEX	63.1	2.22	62.2	2.09	56.1	1.23
J-DPF	15.3	0.63	19.4	1.82	11.7	-0.41
S-DPF	36.2	3.26	35.3	3.05	26.9	1.14

FIGURE 2.5/12: Table of average acoustic values for different pathologies

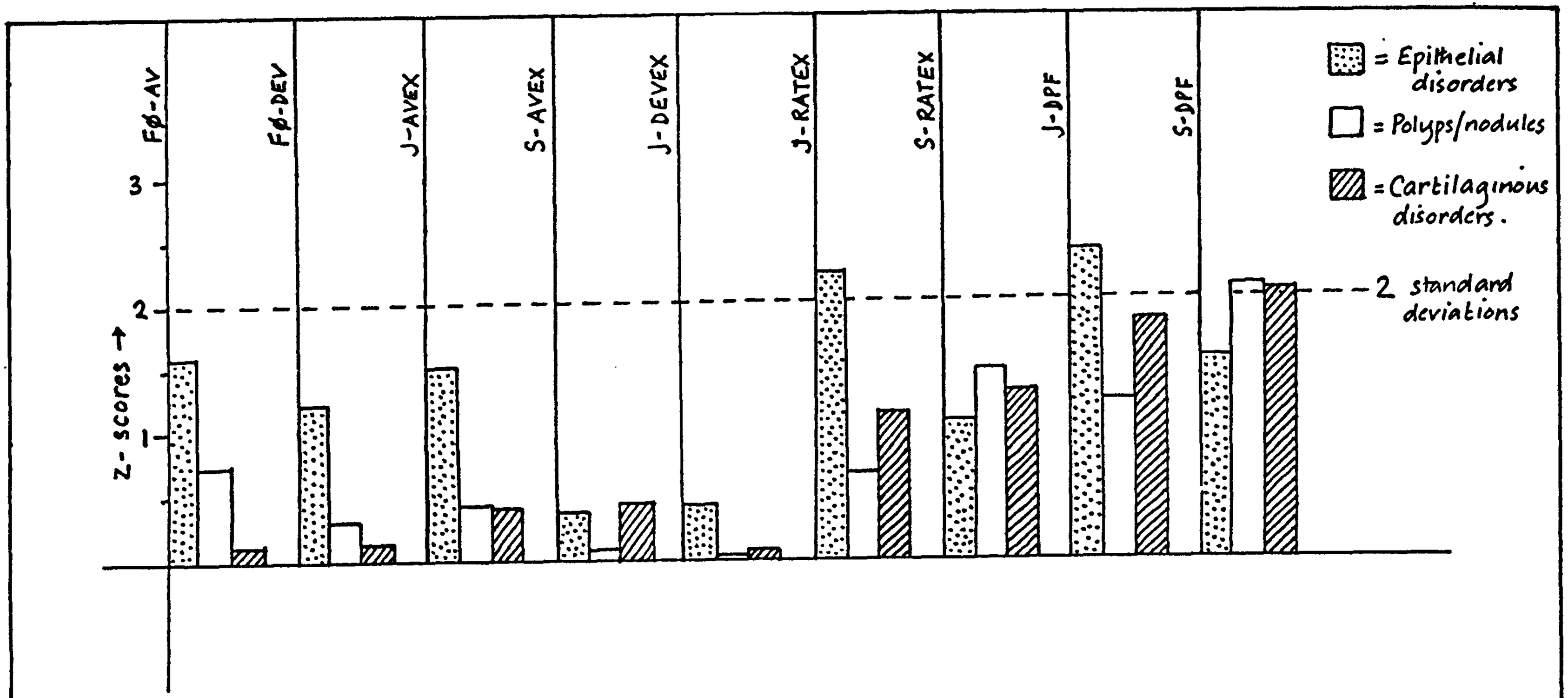


FIGURE 2.5/13: Average acoustic profiles in different pathologies: males

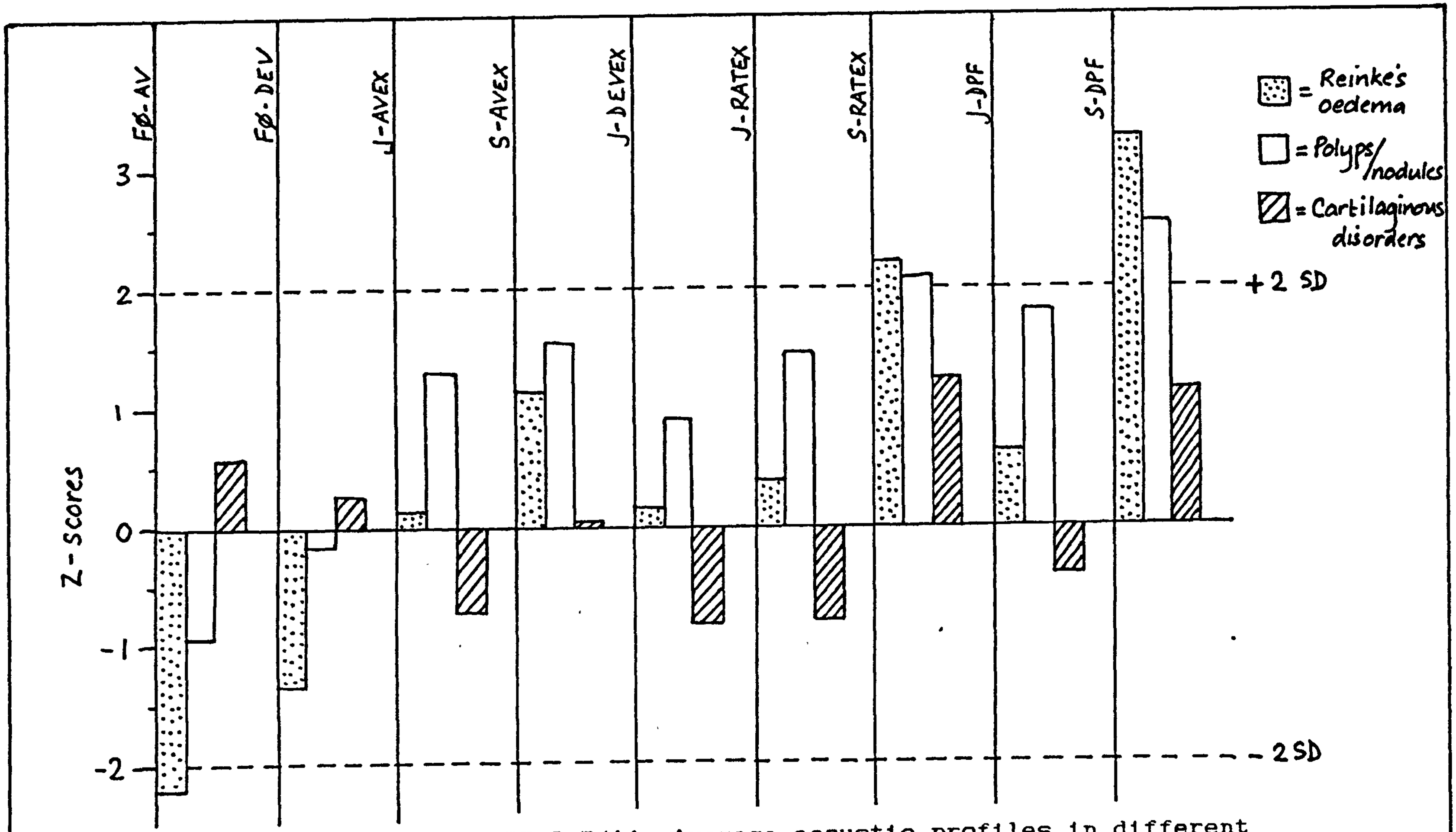


FIGURE 2.5/14: Average acoustic profiles in different pathologies: females

average acoustic profiles for these groups, and several points emerge from these.

The only really homogeneous pathology group in terms of details of the structural abnormality is the Reinke's oedema group, so that it is difficult to relate the other results to the predictions made earlier. It is worth noting that all pathology groups, with the exception of the female cartilaginous disorders, seem to be associated with increased average levels of perturbation. This fits with the theory that normal vocal fold vibration is very sensitive to changes in the mechanical state of the ligamental portion of the vocal folds, but that alterations in the cartilaginous portion have much less effect on vocal fold vibration. Most of the lesions included in this study involved some degree of mass increase, which would be expected to lower the average FO. In fact, we find that mean FO-AV is actually higher than normal in all the male pathology groups, and in the female group of cartilaginous disorders. This may be due to the fact that many of the disorders also involved some increase in stiffness, which might balance the mechanical consequences of mass increase. This is especially true of the epithelial disorders, many of which involve keratosis, and hence an increase in stiffness of the vocal fold cover, and it is therefore not surprising that the greatest mean FO-AV is found in this group. Another factor which might lead to increased FO-AV levels would be a tendency for speakers to boost overall laryngeal tension in response to changed laryngeal structure.

One striking feature is that the ratio between mean values of shimmer and jitter for RATBX and DPF shows marked differences between pathology types. In males, epithelial disorders of the ligamental region are characterised by relatively higher jitter scores, whilst polyps, nodules and disorders of the cartilaginous region

tend to have higher shimmer scores. Two possible explanations for this may be proposed. The first, which is mechanically based, is that in some way which is not altogether clear, alterations in the epithelium have a different perturbatory effect from alterations in other tissue layers. The second possibility is that the different shimmer/jitter ratios result from different types of phonetic adjustment of the larynx. Many cases of polyps and nodules, as well as contact ulcers (which account for 2 thirds of the cartilaginous disorders in this group), are thought to have an element of vocal misuse in their aetiology. All are thought to be precipitated by long term mechanical abuse of the vocal fold edge, brought about by inappropriately hard adduction of some portion of the vocal fold. This is almost inevitably associated with an increase in muscular tension. Epithelial disorders, in contrast, are not generally thought to be primarily due to excessive tension. Even though these patients may later exhibit increased muscular tension as they attempt to phonate normally in the face of abnormal vocal fold structure, the balance of tension used is likely to be somewhat different from that used by a speaker who has habitually misused his voice over a long period of time. One hypothesis, therefore, is that the kind of habitual tension patterns which may trigger vocal nodule and contact ulcer formation are associated with higher values of shimmer than jitter.

Given the greater importance of detecting epithelial disorders, it is encouraging that there is an apparent tendency for these cases to show a different acoustic profile from any other disorders.

In females, both Reinke's oedema and vocal polyps and nodules seemed to be associated with higher values of shimmer than jitter. This rather contradicts the phonetic

adjustment argument, since habitual vocal misuse is not generally implicated in the aetiology of Reinke's oedema. There is, however, a case for suggesting that the gradual onset of Reinke's oedema may be associated with a gradual increase in habitual tension levels as women attempt to compensate for the unacceptable lowering in pitch which the large mass brings about. The prediction that pitch should be lowered by the mass increase of the vocal folds in Reinke's oedema is clearly supported by the results shown here.

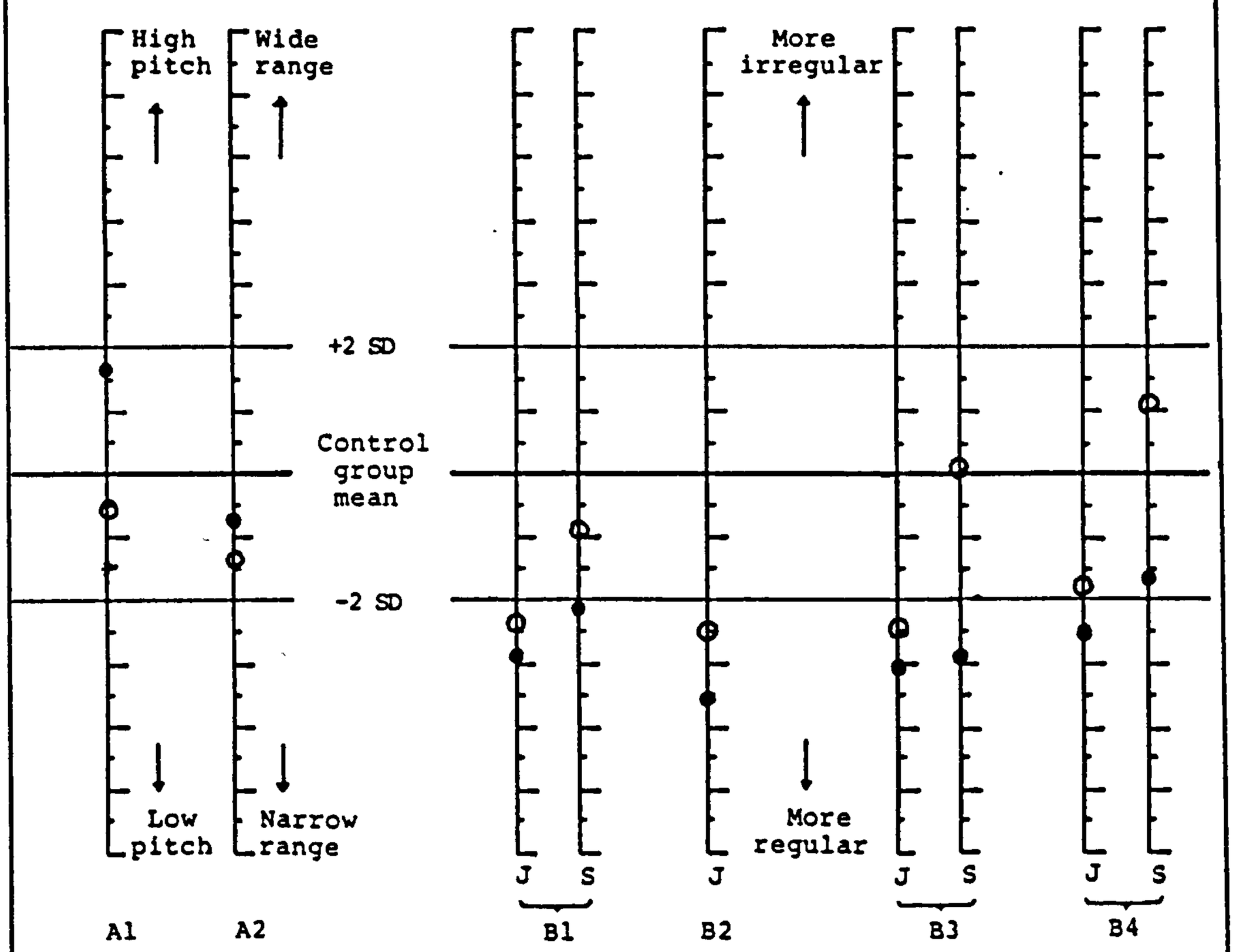
One finding which is not obvious from the averaged results is that some individuals in the pathology group were discriminated from the control group by virtue of having lower than normal levels of perturbation. This was a rather unexpected finding, which is illustrated by the acoustic profiles of two cases shown in Figure 2.5/15. A detailed examination of 8 such cases suggested some common features. Firstly, most such cases displayed relatively minor changes in vocal fold structure, usually early nodules with no apparent increase in tissue stiffness. Secondly, a perceptual analysis of phonation showed that most had phonation types which did not deviate very much from neutral, modal voice. The typical vocal profile for this group showed whisperiness at scalar degree 1-2, which is at the low end of the normal range (see Section 2.2), with moderate levels of laryngeal tension. Most cases seemed to exhibit a peculiar phonatory quality, which was not properly described using the standard VPAS labels. This became known as "incipient creak", because trained judges often commented that the speaker sounded as if he or she was always about to start using a creaky phonation, but never quite slipped into a creaky setting. Another comment which recurred frequently was that phonation in these speakers had a "mechanical" quality. This may be a reflection of the greater than normal regularity of vocal

ACOUSTIC PROFILE

Speaker: ● = Female, O = Male } Both with polyps Sex: Age: Date:

A. PITCH MEASUREMENTS
= smoothed FØ

B. MEASUREMENTS OF PHONATORY IRREGULARITY
J = JITTER (pitch irregularity)
S = SHIMMER (intensity irregularity)



A1 = Pitch mean (mean FØ)

A2 = Pitch variability (SD FØ)

B1 = Average size of irregularities (AVEX)

B2 = Standard deviation of irregularities (DEVEX)

B3 = Percentage of substantial irregularities (RATEX)

B4 = Percentage of substantial reversals in pitch/intensity contour (DPF)

"ACOUSTIC ANALYSIS OF VOICE FEATURES" Research Project.
(MRC Grant No. G8207136) Centre for Speech Technology Research,
Department of Linguistics, University of Edinburgh.

FIGURE 2.5/15: Acoustic profiles of two speakers with unusually regular phonation

fold vibration, which listeners tend to associate with synthetic speech. Perceptually, this phonation type may be similar to what Catford (1977:32) describes as anterior voice, where the arytenoids are closely adducted and only the ligamental portion of the vocal folds are involved in phonation.

A hypothesis which was prompted by finding these "super-regular" voices is that there is a clear relationship between overall laryngeal tension and perturbation levels. Figure 2.5/16 shows a graphical representation of this hypothesis. It is well documented that excessive laryngeal tension in organically normal speakers tends to be associated with harshness, and increased levels of perturbation (Berg 1955:63, Laver 1980:144). The suggestion illustrated in the figure is that slight or moderate increases in laryngeal tension tend to decrease perturbation, as long as there is no major organic change in the larynx. Beyond this level, there is a rapid loss of vibratory efficiency, so that perturbation levels suddenly increase above normal. There is little research reported to support this hypothesis, but there is some circumstantial evidence to support it. One observation which is relevant here stems from VPAS training courses. If normal speakers are asked to produce neutral phonation (modal voice), they have to reduce the level of whisperiness and minimise any phonatory irregularities (i.e. harshness and creakiness) which are present in their habitual voices. Almost invariably, they report that this is only possible if they increase laryngeal tension, and it is very common for early attempts at modal voice production to fluctuate erratically between a reasonable approximation to modal voice and markedly harsh voice. This suggests that there is a critical tension level, at which there is a rapid transition from "super-regular" phonation to harshness.

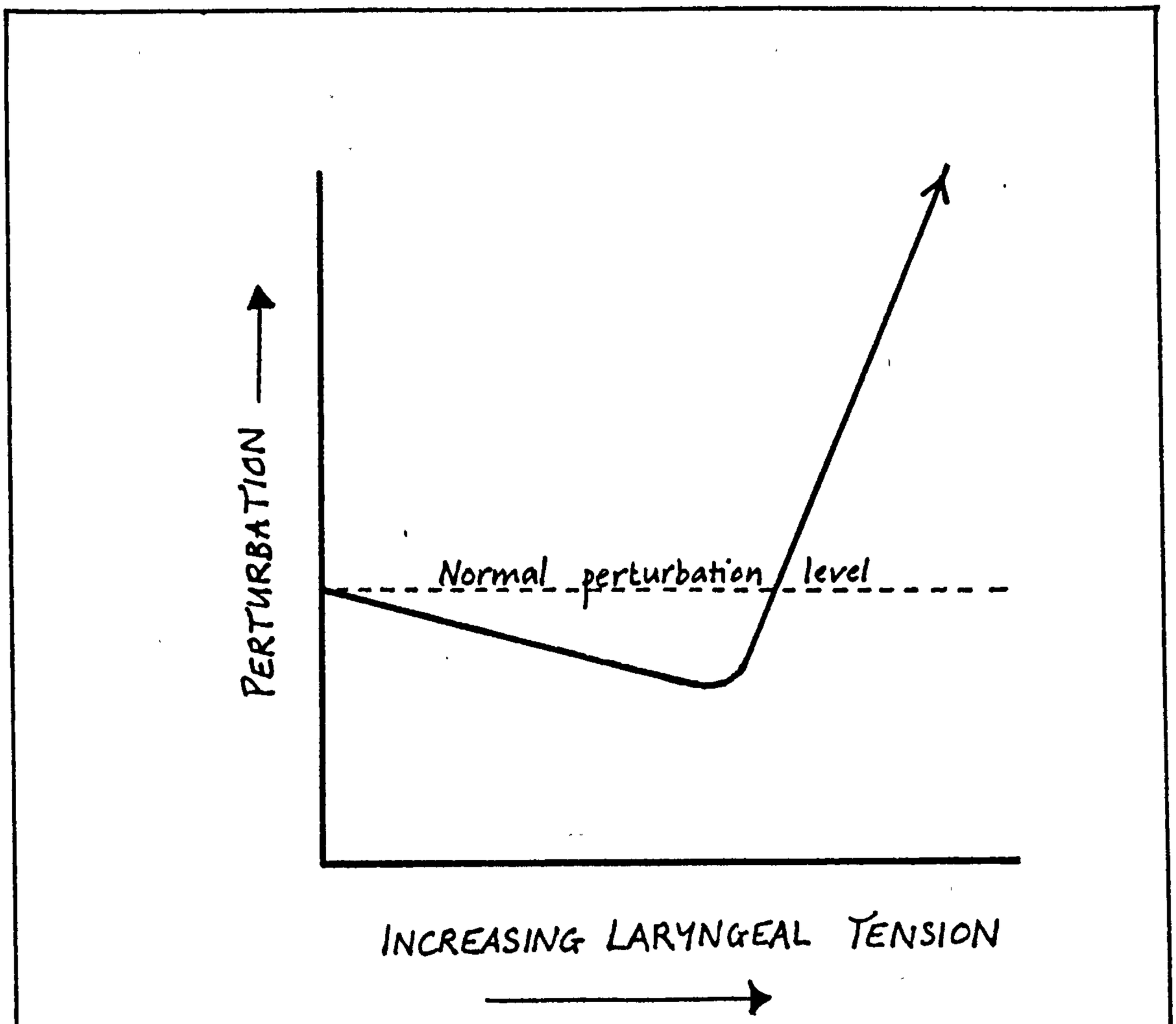


FIGURE 2.5/16: Graphic representation of hypothesised relationship between laryngeal tension and perturbation

Although the small numbers of cases with abnormally low perturbation scores allows only a rather anecdotal report of their characteristics, it would be interesting to follow up this line of investigation in a further study, since it might be a useful strand in the development of acoustic analysis as a supplementary diagnostic tool.

In the absence of adequate group data on pathology classes, the following four case studies may serve as illustrations of the variety of acoustic patterns which may be found in a pathological population, and of the way in which acoustic profiles may relate to the predictions outlined in the first part of this section. Although the use of bivariate plots of Mean F0 versus perturbation was initially introduced as a simple graphic method of relating theoretical predictions about mechanical consequences of vocal fold disorder to acoustic results, these have now been abandoned in favour of full acoustic profiles. These may be less immediately interpreted visually, but they do show the relationship between jitter and shimmer as well as the relationship between F0 and perurbation scores. Given the possible importance of jitter to shimmer ratios in differentiating classes of pathology, graphic complexity seems a necessary price to pay for adequate diagnostic information.

Case 1: Female patient with Reinke's oedema.

This 65 year old woman presented with a history of hoarseness lasting for several years. She admitted to smoking 20 cigarettes per day. Indirect laryngoscopy on the day of recording for acoustic analysis showed bilateral swelling of the vocal folds, but with no obvious epithelial abnormality, and a tentative diagnosis of Reinke's oedema was made. Direct laryngoscopy two weeks later confirmed this diagnosis. There was considerable fluid accumulation within the mucosa at the

glottal edge of both vocal folds, but no stiffening or significant thickening of the epithelium. The predicted acoustic consequence of such a symmetrical mass increase is that mean F0 would be reduced (see earlier in this section), without any necessary increase in perturbation. The acoustic profile for this patient is shown in Figure 2.5/17, and it can be seen that the only acoustic parameter which falls outside the 2 SD range is mean F0, which is indeed lower than normal. The acoustic results thus fit the predictions quite well. This profile also shows a consistent pattern of low jitter to shimmer ratios, with all perturbation measures being somewhat lower than the control mean. Following the above discussion, this may be interpreted as suggesting a boost in tension as the patient attempts to compensate for the lowering in F0 due to increased vocal fold mass.

Case 2: Female patient with unilateral sessile vocal polyp.

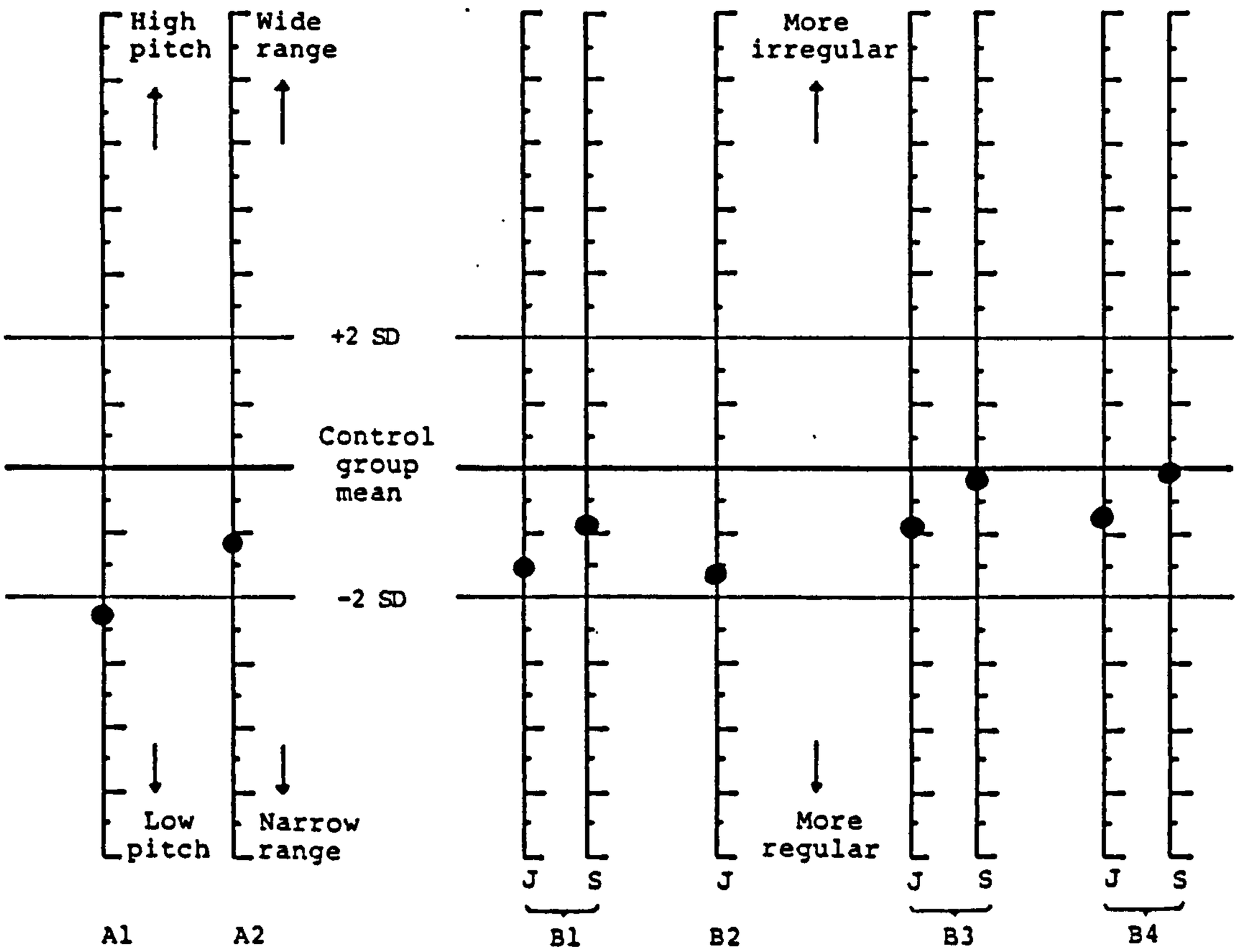
This 44 year old woman had a six month history of hoarseness following a short period of complete voice loss associated with influenza. There was no indication of excessive habitual voice use, but she did smoke about 20 cigarettes per day. Indirect laryngoscopy on the day of recording showed a large sessile vocal polyp in the centre of the ligamental portion of the left vocal fold, occupying about one third of the total length of the vocal fold. The polyp appeared to be very flexible, moving up and down in the glottal space. Histological examination following biopsy three weeks later showed inflammatory tissue beneath normal epithelium, with no significant stiffening due to hyaline formation. A large asymmetrical mass increase of this type with no increase in stiffness would be expected to cause a reduction in mean F0 and probably an increase in perturbation. Figure 2.5/18 shows the acoustic profile of this patient, which

ACOUSTIC PROFILE

Speaker: Sex: Age: Date:

A. PITCH MEASUREMENTS
= smoothed FØ

B. MEASUREMENTS OF PHONATORY IRREGULARITY
J = JITTER (pitch irregularity)
S = SHIMMER (intensity irregularity)



A1 = Pitch mean (mean FØ)

A2 = Pitch variability (SD FØ)

B1 = Average size of irregularities (AVEX)

B2 = Standard deviation of irregularities (DEVEX)

B3 = Percentage of substantial irregularities (RATEX)

B4 = Percentage of substantial reversals in pitch/intensity contour (DPF)

"ACOUSTIC ANALYSIS OF VOICE FEATURES" Research Project.
(MRC Grant No. G8207136) Centre for Speech Technology Research,
Department of Linguistics, University of Edinburgh.

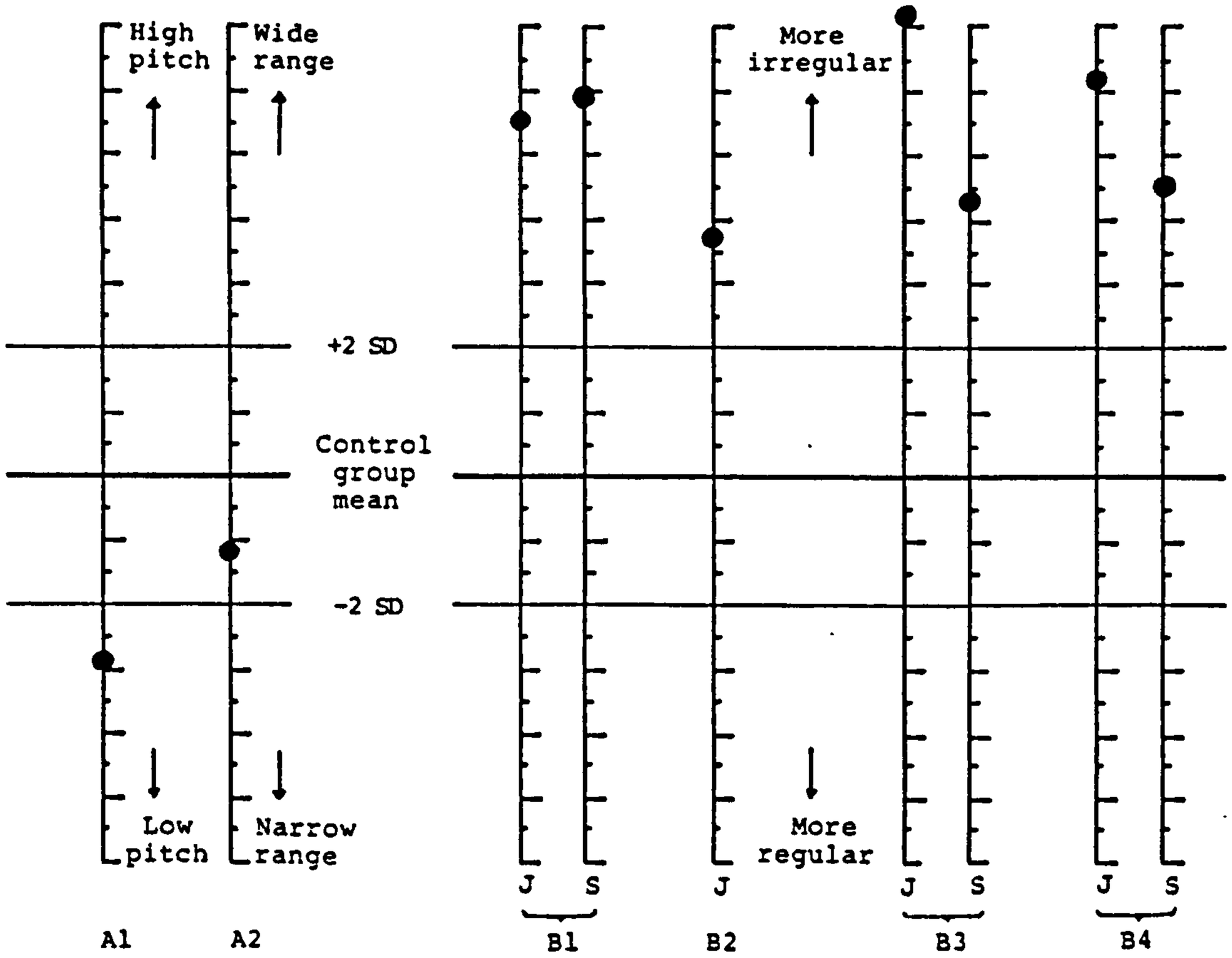
FIGURE 2.5/17: Acoustic profile for a patient with Reinke's oedema

ACOUSTIC PROFILE

Speaker: Sex: Age: Date:

A. PITCH MEASUREMENTS
= smoothed FØ

B. MEASUREMENTS OF PHONATORY IRREGULARITY
J = JITTER (pitch irregularity)
S = SHIMMER (intensity irregularity)



A1 = Pitch mean (mean FØ)

A2 = Pitch variability (SD FØ)

B1 = Average size of irregularities (AVEX)

B2 = Standard deviation of irregularities (DEVEX)

B3 = Percentage of substantial irregularities (RATEX)

B4 = Percentage of substantial reversals in pitch/intensity contour (DPF)

"ACOUSTIC ANALYSIS OF VOICE FEATURES" Research Project.
(MRC Grant No. G8207136) Centre for Speech Technology Research,
Department of Linguistics, University of Edinburgh.

FIGURE 2.5/18: Acoustic profile for a patient with a sessile vocal polyp

again is consistent with the theoretical predictions. Mean F0 is more than 2 SD below the control group mean, and all perturbation scores are greatly increased, being at least 3 SD above the control group means. The low jitter to shimmer scores often seen in speakers with tense phonatory patterns are not evident in this profile, which may indicate that long-term vocal misuse is not implicated in the aetiology of this polyp.

Case 3: A male patient with bilateral keratinization and hyperplasia.

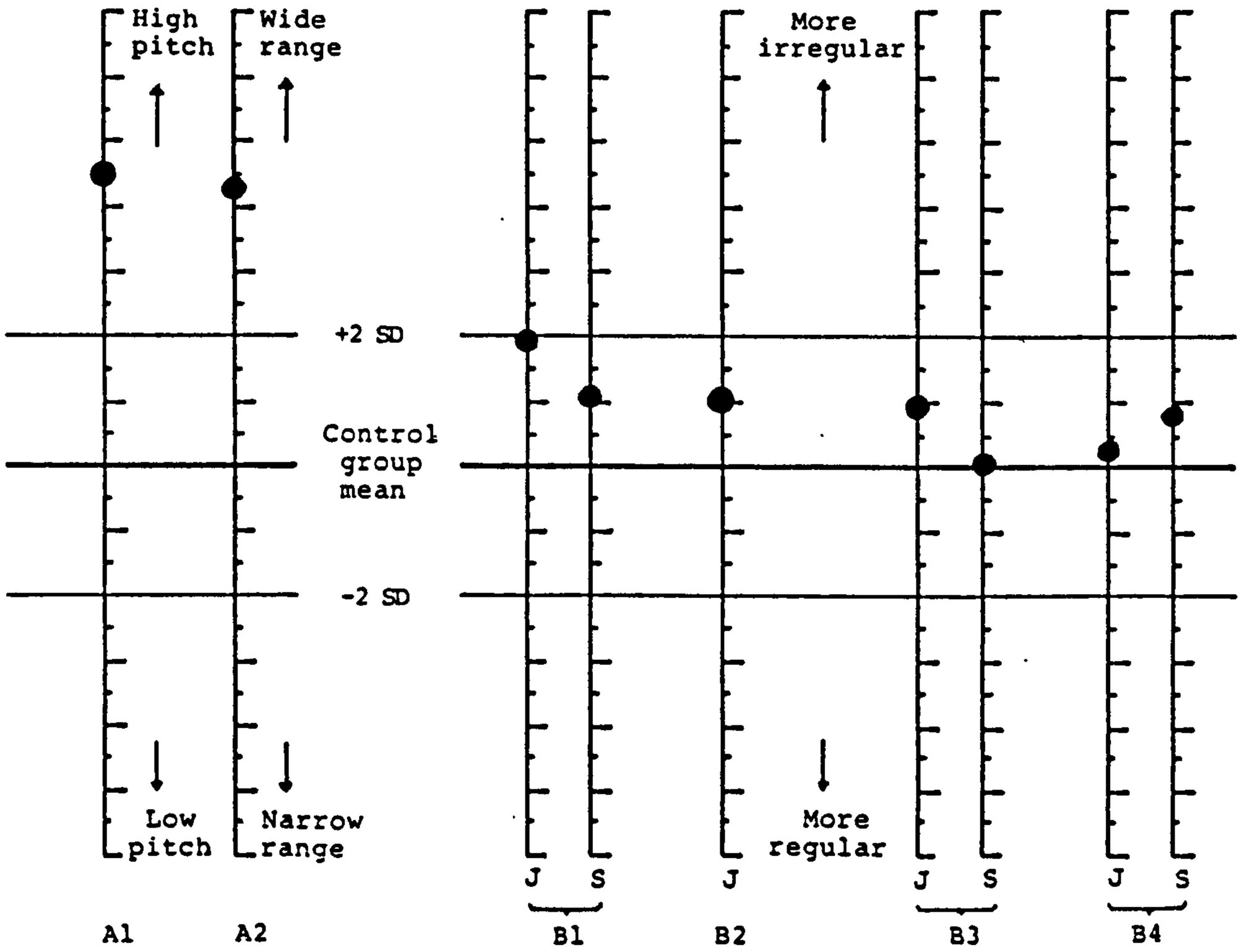
This patient was an 82 year old ex-smoker, who presented following two or three years of progressively increasing hoarseness. Indirect laryngoscopy at the time of recording showed slight thickening of the vocal folds in the middle of the ligamental portion. On the right fold, the mass was whitish, but on the left fold the mass increase was slightly larger and red in colour. Subsequent biopsy showed marked keratinization of both folds, with some hyperplasia, which was more extreme on the left fold. There was some indication of cell abnormality in the hyperplastic tissue of the left fold, but no clear indication of malignancy and no invasion of surrounding tissue. The main mechanical change thus appears to be a stiffness increase, which would theoretically be predicted to cause an increase in F0, although the slight mass increase would tend to reduce this effect. The changes affect both vocal folds, but are not exactly symmetrical, so that some increase in perturbation measurements might also be expected. Figure 2.5/19 shows that mean F0 is unusually high, but that perturbation measures are all within 2 SD of the control group mean, with only J-AVEX approaching a suspicious level. It may be that a greater degree of asymmetry is necessary to produce an increase in perturbation scores.

ACOUSTIC PROFILE

Speaker: Sex: Age: Date:

A. PITCH MEASUREMENTS
= smoothed FØ

B. MEASUREMENTS OF PHONATORY IRREGULARITY
J = JITTER (pitch irregularity)
S = SHIMMER (intensity irregularity)



A1 = Pitch mean (mean FØ)

A2 = Pitch variability (SD FØ)

B1 = Average size of irregularities (AVEX)

B2 = Standard deviation of irregularities (DEVEX)

B3 = Percentage of substantial irregularities (RATEX)

B4 = Percentage of substantial reversals in pitch/intensity contour (DPF)

"ACOUSTIC ANALYSIS OF VOICE FEATURES" Research Project.
(MRC Grant No. G8207136) Centre for Speech Technology Research,
Department of Linguistics, University of Edinburgh.

FIGURE 2.5/19: Acoustic profile for a patient with keratinization and hyperplasia

Case 4: a male patient with squamous carcinoma.

This 57 year old smoker complained of increasing hoarseness. Indirect laryngoscopy showed a tumour extending from the ventricular fold to the ligamental portion of the true vocal fold on the right hand side, and the observation of oedema of the left vocal fold suggested that the tumour might extend to both folds, prompting an inflammatory reaction in the left fold. Biopsy a week later confirmed a diagnosis of invasive squamous carcinoma involving the right ventricular and true vocal folds, but there was no indication of transglottal spread in spite of marked inflammation of the left fold. The malignant tissue had patches of keratinization, which would increase the stiffness of the vocal fold. The mechanical effects of an asymmetrical mass increase, with stiffening of vocal fold tissue are hard to predict. The gross asymmetry is expected to cause perturbed vocal fold vibration, but the mass increase and the stiffness increase act in opposition. Figure 2.5/20 Shows that mean F0 is in fact within the normal range, but that some of the perturbation measures are unusually high. J- AVEX, J-RATEX, S-RATEX and J-DPF are well above the 2SD line, and this patient's scores are typical of the high jitter to shimmer ratios which are common in the epithelial group of disorders.

In summary, the use of acoustic profiles to examine the characteristics of different types of pathology does seem to reflect theoretical predictions about mechanical changes in the vocal folds in at least some cases. It also prompts some suggestions about factors which might be worthy of further investigation, if larger numbers of subjects with well documented organic disorders could be recorded. These are:

i. the relationship between F0 and perturbation scores,

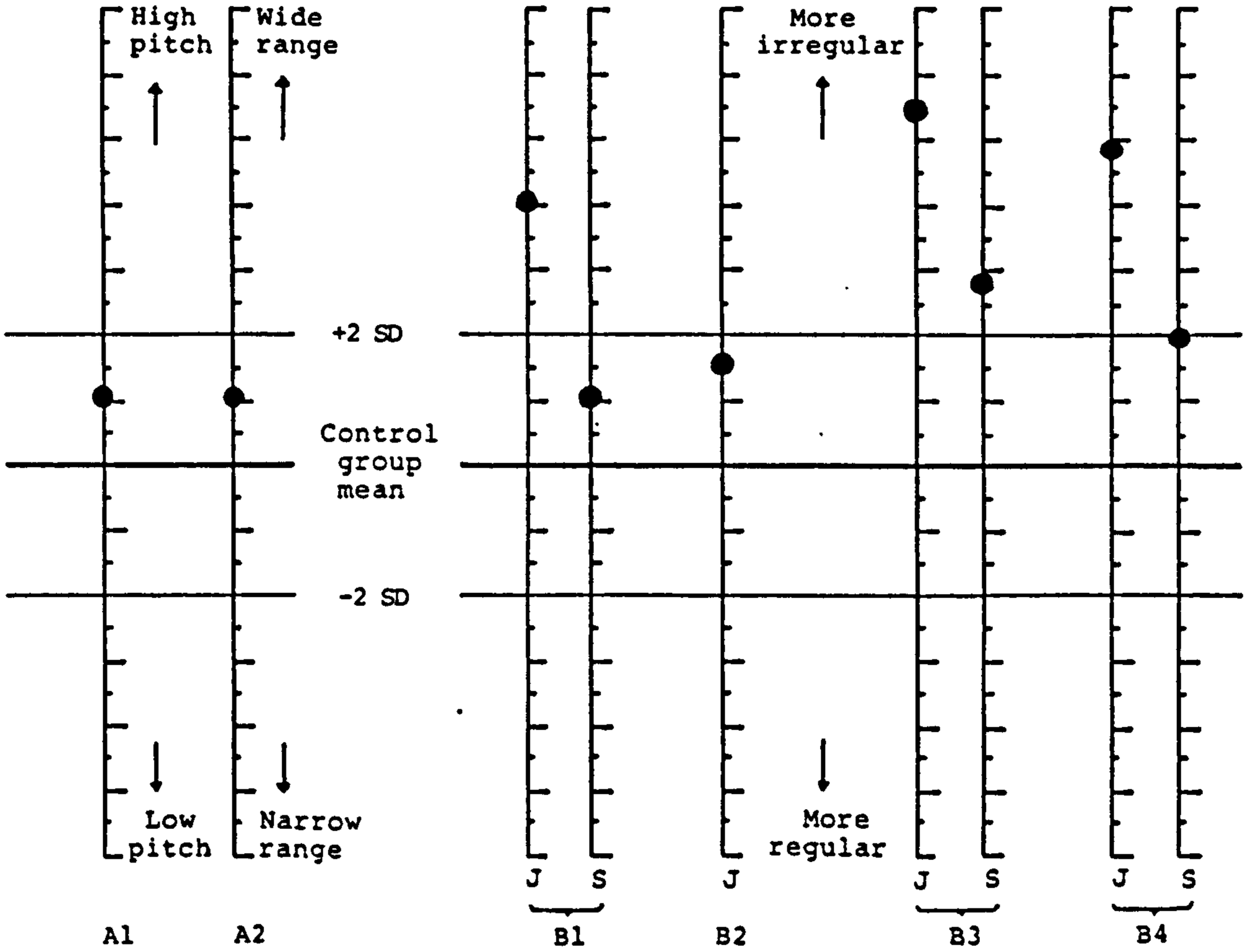
ii. the ratio of shimmer to jitter scores, and

ACOUSTIC PROFILE

Speaker: Sex: Age: Date:

A. PITCH MEASUREMENTS
= smoothed FØ

B. MEASUREMENTS OF PHONATORY IRREGULARITY
J = JITTER (pitch irregularity)
S = SHIMMER (intensity irregularity)



A1 = Pitch mean (mean FØ)

A2 = Pitch variability (SD FØ)

B1 = Average size of irregularities (AVEX)

B2 = Standard deviation of irregularities (DEVEX)

B3 = Percentage of substantial irregularities (RATEX)

B4 = Percentage of substantial reversals in pitch/intensity contour (DPF)

"ACOUSTIC ANALYSIS OF VOICE FEATURES" Research Project.
(MRC Grant No. G8207136) Centre for Speech Technology Research,
Department of Linguistics, University of Edinburgh.

FIGURE 2.5/20: Acoustic profile for a patient with squamous carcinoma

iii. a proper investigation of subjects with low perturbation scores.

In addition, an examination of the relationship between tension and perturbation in normal speakers might go some way towards extricating the effects of vocal misuse from the consequences of organic change. The natural tendency of any speaker whose vocal apparatus undergoes an organic change is to make phonetic adjustments in order to minimise the vocal consequences of the organic change. Any acoustic screening system therefore needs to attempt to separate the acoustic effects of phonetic adjustment from those which have an organic basis. This is especially important in light of the fact that the acoustic system seemed to discriminate between controls and a group of dysphonic patients with no observable vocal fold pathology nearly as well as it did between controls and speakers with known pathology (Laver et al. 1985:11). The acoustic system therefore seems to be sensitive to acoustic abnormalities in phonation regardless of whether they have an organic or a phonetic basis. It may well be that the pattern of acoustic deviation found in the dysphonic group is rather different from that found in the organic disorders, but there is inadequate data to examine this possibility at present.

C. The use of acoustic analysis in tracking longitudinal change

One of the major problems in using phonatory output for screening or diagnosis of laryngeal pathology is that there is such wide interspeaker variation, both in phonetic adjustment of the normal larynx and in phonetic response to the presence of vocal fold pathology. This problem does not arise when acoustic analysis is used to track longitudinal change in individuals. During the span

of the MRC project, 30 patients who were undergoing speech therapy, surgery or radiotherapy were recorded at least twice, so that we collected acoustic data before and after treatment. This was such a diverse group, that no sensible group statistics are possible, but two examples may serve as illustration of the immediate applicability of the acoustic system to the assessment of changes in individual phonatory patterns.

The first example is of a male patient, aged 75 years, with squamous carcinoma affecting the centre of the ligamental portion of the vocal fold. This was a unilateral lesion, with some keratosis, so that there was a significant increase in both mass and stiffness of the epithelium, and a certain amount of disruption of the normal tissue layer relationships. He was first recorded at the time of the first indirect laryngeal examination. He was recorded for a second time three months later, a. month after completion of a course of radiotherapy, which had been preceded by a small biopsy. At this time laryngeal examination showed some reddening and swelling of the vocal folds, which is a normal response to radiotherapy, but no sign of cancerous growth. A third recording was made six months after the first analysis, at which time the vocal folds appeared fairly healthy, and more normal in colour, although there was still minimal oedematous thickening.

The results of these analyses are shown in Figure 2.5/21. At the time of diagnosis, the profile shows that whilst pitch mean and range are within normal limits, five out of seven perturbation measures are radically deviant. The profile demonstrates very clearly the typical epithelial disorder pattern of high jitter/shimmer ratios. After biopsy and radiotherapy, all perturbation measures are within normal limits. Pitch mean and range are now slightly low, although still within the normal range.

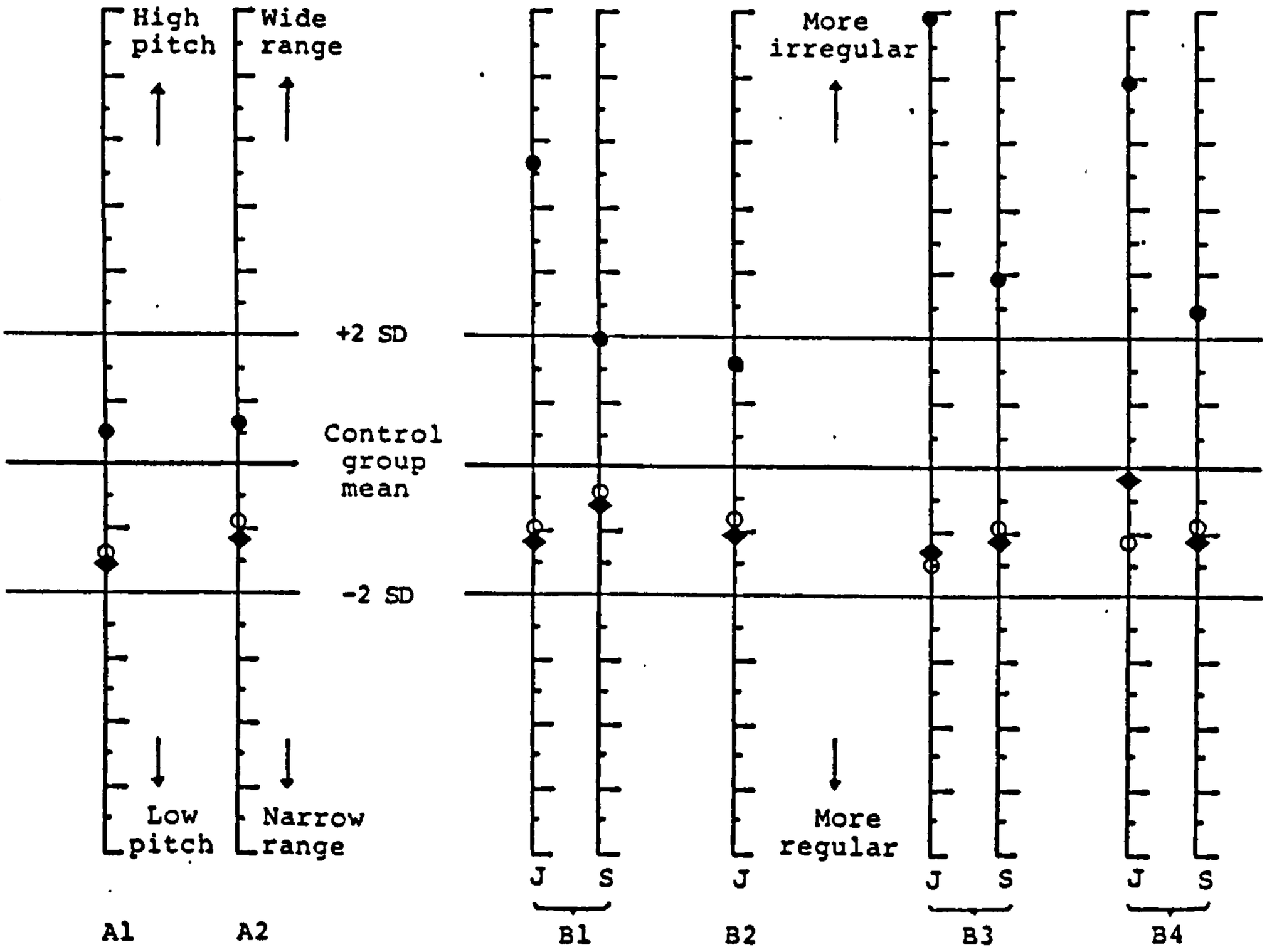
● = at 1st ENT examination
◆ = three months later, following biopsy and radiotherapy
○ = six months after 1st recording

A. PITCH MEASUREMENTS

= smoothed FØ

B. MEASUREMENTS OF PHONATORY IRREGULARITY

J = JITTER (pitch irregularity)
S = SHIMMER (intensity irregularity)



A1 = Pitch mean (mean FØ)

A2 = Pitch variability (SD FØ)

B1 = Average size of irregularities (AVEX)

B2 = Standard deviation of irregularities (DEVEX)

B3 = Percentage of substantial irregularities (RATEX)

B4 = Percentage of substantial reversals in pitch/intensity contour (DPF)

"ACOUSTIC ANALYSIS OF VOICE FEATURES" Research Project.
(MRC Grant No. G8207136) Centre for Speech Technology Research,
Department of Linguistics, University of Edinburgh.

FIGURE 2.5/21: Longitudinal study of a patient with squamous carcinoma

This is a predictable consequence of a bilateral increase in mass due to radiation induced oedema. The results of the third analysis are very similar to the second, which is consistent with the lack of any significant change in laryngeal appearance.

The second example of a longitudinal study concerns a female, aged 43, who presented with small bilateral vocal nodules at the centre of the ligamental portion of the vocal fold. It was felt that three noisy children and a job as a nursery school teacher had encouraged misuse and overuse of her voice, resulting in mechanical trauma to the edges of the vocal folds. This patient was recorded at the first laryngeal examination, and again two months later, following a course of speech therapy aimed primarily at reducing phonatory tension and encouraging the patient to monitor her own phonatory output. At the time of the second recording, the vocal nodules had almost disappeared, and the vocal folds looked generally healthy.

Figure 2.5/22 compares this speaker's acoustic profile before and after therapy. The first profile shows that pitch mean and range are within the normal range, but all four jitter measures are abnormally low. Shimmer measures were also low, but just within the normal range. Two features are of interest here. One is the low jitter/shimmer ratio, which ia very different from the previous speaker, and the other is the "super-regular" phonation, which we assume to be associated with increased laryngeal tension.

Following 10 sessions of therapy, the profile shows some changes. Pitch mean and range are actually further from the control group mean, being lower than previously, but since this may result from a reduction in laryngeal tension this is not necessarily a bad thing. More

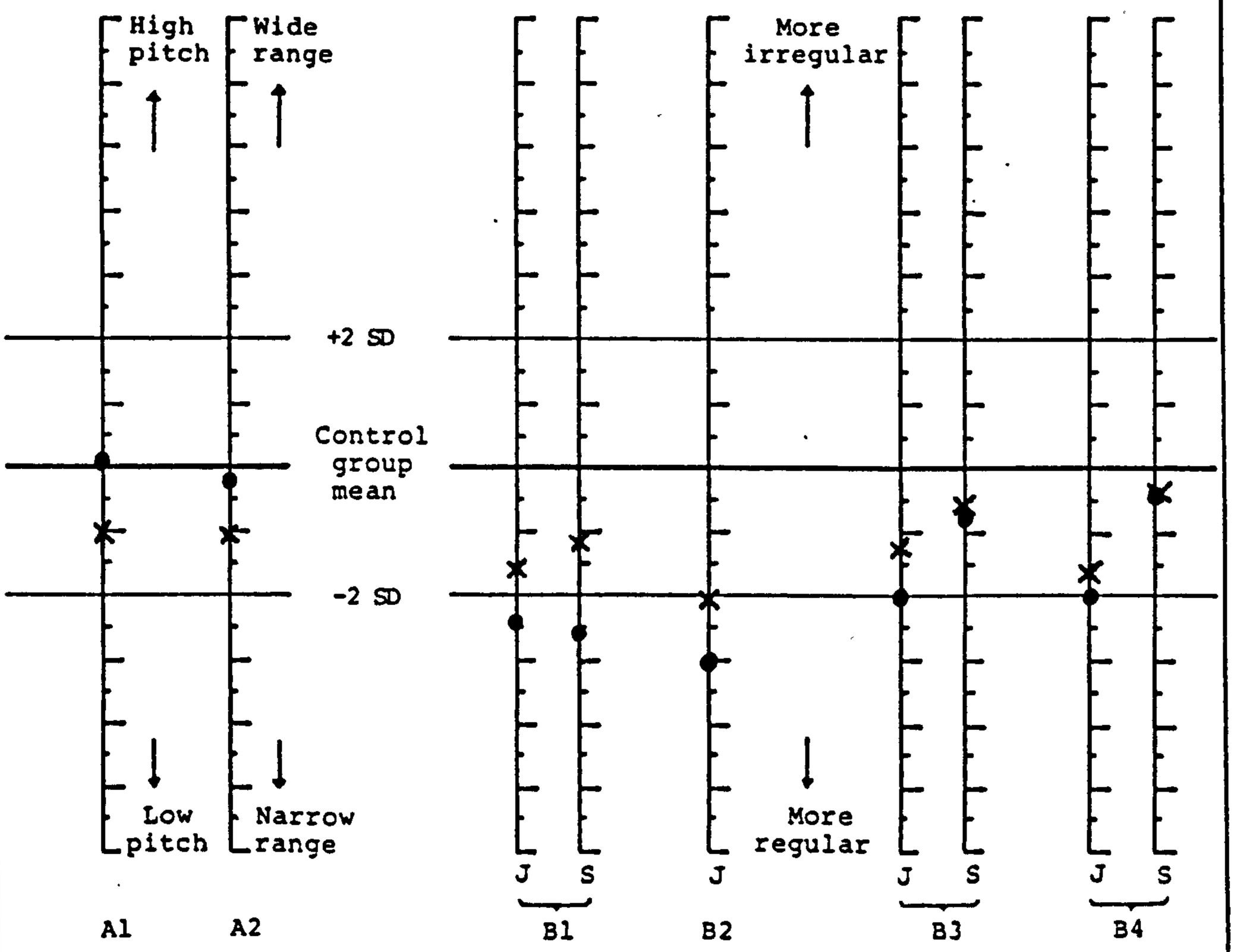
● = 1st. assessment
× = after therapy

ACOUSTIC PROFILE

Speaker: Sex: Age: Date:

A. PITCH MEASUREMENTS
= smoothed FØ

B. MEASUREMENTS OF PHONATORY IRREGULARITY
J = JITTER (pitch irregularity)
S = SHIMMER (intensity irregularity)



A1 = Pitch mean (mean FØ)

A2 = Pitch variability (SD FØ)

B1 = Average size of irregularities (AVEX)

B2 = Standard deviation of irregularities (DEVEX)

B3 = Percentage of substantial irregularities (RATEX)

B4 = Percentage of substantial reversals in pitch/intensity contour (DPF)

"ACOUSTIC ANALYSIS OF VOICE FEATURES" Research Project.
(MRC Grant No. G8207136) Centre for Speech Technology Research,
Department of Linguistics, University of Edinburgh.

FIGURE 2.5/22: Longitudinal study of a patient with vocal nodules

importantly, all perturbation measures are now closer to the control group mean, and only one (J-RATEX) is outwith the normal range. It should be stressed that it is not possible to say how far these improvements are due to organic change and how far they are due to more relaxed and less damaging phonatory patterns. However, it is reassuring for both patient and therapist to have some objective measure showing a phonatory pattern which is closer to a normal baseline. Given the increasing demand for clinicians of all sorts to demonstrate that the therapies they prescribe are cost effective, the value of any technique which allows objective assessment of change during and following therapy is considerable.

2.5.5 DISCUSSION AND CONCLUSIONS

The results of this study are encouraging, but it is not yet possible to say that the value of the acoustic system in screening and diagnosis of laryngeal pathology is proven. Any further studies aiming to fully evaluate the system would need larger subject groups, and detailed information about the laryngeal status of all subjects, including the controls. Exhaustive statistical manipulation of the existing acoustic parameters might need to be supplemented by an examination of other acoustic parameters in an attempt to improve acoustic discrimination between healthy and pathological voices. The relatively low incidence of laryngeal carcinoma in the population as a whole demands that any screening procedure should have a very high degree of accuracy in order to be practicable and economically viable. The implementation of widespread screening could result in ENT clinics being flooded with false alarms, whilst detecting only a few cases of genuine laryngeal cancer, unless the false positive rate can be pared to a much lower rate than was achieved in this study. A full analysis of the economic and human cost of laryngeal

disease would have to be balanced against the cost of screening any given sector of the population, and the efficiency of the screening system.

The demands of longitudinal studies of individual cases are somewhat different, since each case provides its own baseline, and comparisons with proper control populations are less important. Even without further development, the acoustic system seems to show considerable potential as a means of tracking changes in laryngeal function over time.

The acoustic profile form is of particular value in longitudinal studies, but it has also proved to be a useful clinical adjunct to Vocal Profile analysis in the initial assessment of patients (Nieuwenhuis and Mackenzie 1986). The profile form allows an objective record of acoustic output to be kept in patients' files, and is a useful basis for discussion, both with medical colleagues and with patients. In spite of the complexity of the acoustic information, which means that decisions about presentation to patients have to handled with care, many patients do find it useful to see some objective analysis of the "pitch" and "smoothness" of their voices, and are encouraged by any acoustic evidence of improvement following speech therapy or other treatments.

PART THREE: CONCLUSIONS

This thesis had two main objectives. The first of these was to examine the types of organic variation which make each individual's vocal apparatus unique, and to investigate the sources from which such variations arise. In order to fulfill this aim, Part One of the text described first the structure and properties of the cells and tissues which make up the vocal apparatus, and then the ways in which they grow and change during the human life cycle. Tissue responses to trauma, disease and the aging process are also described. The coordinate influence of tissue growth and change on the overall form of the vocal apparatus at each stage of the life cycle is then discussed. Part One concluded with a brief review of the available literature on vocal characteristics of each sex at different ages, and discussed the extent to which these characteristics may be related to organic factors.

It is hoped that this part of the thesis will help to make the most relevant parts of the medical and biological literature available to phoneticians and speech therapists.

The second objective of the thesis was to study the vocal characteristics of some groups of speakers with normal and abnormal vocal anatomy, in order to see if direct links could be drawn between observations of organic state and vocal performance. This necessitated the development of appropriate objective techniques for voice quality analysis. Part Two therefore began with a presentation of two techniques for voice analysis which the author helped to develop.

The first of these, the Vocal Profile Analysis Scheme (VPAS), is now in widespread use in Britain and abroad, but Section 2.1.2 is the first full description which

covers both underlying theory and practical application. This scheme has been instrumental in encouraging speech therapists to look at the whole vocal apparatus when assessing voice quality, and it has helped to highlight the complex interactions of different parts of the vocal apparatus. One of the more valuable ideas to come out of the development of the VPAS has been the idea of auditory equivalence between the vocal output of normal and abnormal vocal apparatuses. The basis of this is that a speaker with an abnormal vocal tract may produce an auditory quality which is equivalent to that produced by a normal vocal apparatus, even though the phonetic adjustments needed to produce that quality must be different from normal. In other words, speakers with dissimilar vocal tracts must achieve the same configuration of the vocal tract by different muscular means. This shifts the emphasis of voice analysis from phonetic adjustment to vocal apparatus configuration. The vocal profile of a speaker with an abnormal vocal apparatus has to be interpreted in terms of vocal tract configuration, and this can be done by knowing the configurational effects of the phonetic adjustments which a normal speaker would make to produce an equivalent voice quality. Trivial though this shift in emphasis may seem, it actually opens the way for a much more sensible interpretation of phonetic assessments of pathological speech than has sometimes been the case hitherto.

The second analysis technique, described in Section 2.1.3, is an acoustic procedure, which focusses more narrowly on aspects of phonation.

Section 2.2 reports the results of Vocal Profile Analysis of a group of 50 normal speakers, and discusses some sex differences which emerged. This serves as background for the study described in Section 2.3, on voice quality in Down's Syndrome (DS). The DS study is prefaced by an

account of the reported organic characteristics of the vocal apparatus in DS, which allows voice quality findings to be related to organic factors.

Section 2.4, describing the acoustic analysis of a control population, is followed by an evaluation of the technique as a means of assessing laryngeal disease (Section 2.5). The results of this study are promising, but an extension of this work is necessary in order to fully evaluate the ability of the acoustic system to detect and identify vocal fold pathologies.

The main conclusion of the experimental work described in this thesis is that objective voice quality analysis may show that organic factors have auditory and acoustic consequences which permeate an individual's vocal communication. This has important implications for many disciplines. The relevance to medicine and speech therapy is perhaps the most obvious point to emerge from this thesis. The ability to use perceptual or acoustic techniques to detect organic abnormality, or to track changes over time, has several major advantages. Firstly, it may, in certain cases, obviate the need for more expensive medical investigations. Whilst no voice analysis technique is likely ever to replace medical examination as the primary means of diagnosis, the use of acoustic analysis as a long term review procedure might eventually prove to be sensitive enough to replace at least some laryngeal examinations. This would be extremely beneficial, in both economic and human terms, in areas which are geographically remote from well equipped ENT clinics. Tape recordings could be made locally, thus removing the need for patients to travel large distances.

The effect of organic variation upon vocal features may also be of great interest at the interface between

phonetics and psychology. The observation that voice quality is closely related to the expression of emotion (Bezooijen 1984) and to other behavioural aspects of interaction (Marwick et al. 1984), coupled with the readiness with which listeners are prepared to attribute personality features to a speaker on the basis of voice quality (Saville 1983), means that any organic abnormalities which affect voice quality may have far reaching consequences for interpersonal relationships. There is a rich seam of research topics to be mined here.

A further motivation for examining the effects of normal organic variation on voice quality stems from the burgeoning area of speech technology. In the forefront of this field at the moment is the development of acoustic systems for speaker recognition and speech analysis.

The implications of organic variation for speaker recognition are obvious. The acoustic parameters which are available for analysis arise from two sources. They arise partly from a speaker's habitual phonetic adjustments of the vocal apparatus, and partly from organic factors. The development of speaker recognition or verification devices must therefore involve careful evaluation of both sources. On one hand, the relatively invariable nature of an individual's organic make-up may be crucial in allowing detection of imposters. An imposter may be able to mimic very closely the acoustic parameters which are susceptible to phonetic adjustment, but may not be able to replicate the acoustic features which result directly from unique organic characteristics. On the other hand, the fact that some organic characteristics are actually prone to change from day to day may pose real problems for a speaker recognition device. For example, the mucosal lining of the larynx is very subject to day to day variation. A cold, or a night spent drinking in a smoky atmosphere,

may cause dramatic inflammation of the vocal folds, and this can have a marked effect on phonation. If the acoustic parameters which reflect vocal fold vibration are too heavily weighted, the recognition device therefore runs the risk of failing to identify individuals every time they catch a cold or overindulge in alcohol or cigarettes.

The implications for speech recognition may be less immediately obvious, but they are nonetheless important. The phonetic strand of speech recognition is largely concerned with the correct identification of individual phonetic segments or groups of segments. A major problem for speech recognition is that no two speakers produce a given set of segments in exactly the same way, so that any system must be trained to respond or adapt to any given speaker. Again, the inter-speaker differences which the system has to deal with are partly phonetic, accent differences, and partly due to organic variation. Since organic features will exert a general influence throughout speech, identification of the acoustic parameters which have the strongest organic basis might allow the most economical approach to "training" a device to cope with this class of inter-speaker differences.

In summary, it is hoped that this thesis has thrown some light on variation in voice quality, and in the human vocal apparatus. The demonstration that at least some of the rich diversity of voice quality has an organic basis may prompt further research in this area, which would be to the mutual benefit of phonetics, medicine and other disciplines.

BIBLIOGRAPHY

Altman, P.L. and Dittner, D.S. (Eds.) (1962) COMMITTEE ON BIOLOGICAL HANDBOOKS. GROWTH, INCLUDING REPRODUCTION AND MORPHOLOGICAL DEVELOPMENT. Federation of American Societies for Experimental Biology, Washington D.C..

Amado,J.H. (1953) Tableau general des problemes poses par l'action des hormones sur le developpement du larynx, le classement d'une voix, la genese des activites rythmogenes encephaliques, et l'exitabilite du sphincter laryngien. Annales de l'otolaryngologie, 70:117-137.

Andria, L.M. and Dias, J.C. (1978) Relation of maxillary and mandibular intercuspid width to bizygomatic and bigonial breadths. Angle Orthodontist, 48:154-162.

Ardran, G.M., Harker, P. and Kemp, F.H. (1972) Tongue size in Down's Syndrome. Journal of Mental Deficiency Research, 16:160-166.

Arnold, G.E. (1962) Vocal nodules and polyps: laryngeal tissue reaction to habitual hyperkinatic dysfunction. Journal of Speech and Hearing Research 27: 205-216.

Aronson, A.E. (1980) CLINICAL VOICE DISORDERS. AN INTERDISCIPLINARY APPROACH. Thieme-Stratton Inc., New York.

Auerbach, O., Hammond, E.C. and Garfinkel, L. (1970) Histological changes in the larynx in relation to smoking habits. Cancer 25: 92-104.

Austin, J.H.M., Preger, L., Siris, E. and Taybi, H. (1969) Short hard palate in newborn: roentgen sign of mongolism. Radiology, 92:775-776.

Baber, W.E. and Meredith, H.V. (1965) Childhood change in depth and height of the upper face, with special reference to Down's A point. American Journal of Orthodontics, 51:913-927.

Bach, A.C., Lederer, F.L. and Dinolt, R. (1941) Senile changes in laryngeal musculature. Archives of Otolaryngology 34: 47-56.

Baer, M.J. & Nanda, S.K. (1975) A commentary on the growth and form of the cranial base. pp. 515-536 in Bosma, J.F. (Ed.) SYMPOSIUM ON DEVELOPMENT OF THE BASICRANIUM. National Institute of Health, Bethesda.

Baer, T. (1973) Measurement of vibration patterns of excised larynxes. Journal of the Acoustical Society of America 54: 318

Bambha, J.K. (1961) Longitudinal cephalometric

roentgenographic study of face and cranium in relation to body height. Journal of the American Dental Association, 63:776-799.

Bauer, W.C. and McGavran, M.H. (1972) Carcinoma in situ and evaluation of epithelial changes in laryngopharyngeal biopsies. Journal of the American Medical Association 221: 72-75.

Benda, C.E. (1960) THE CHILD WITH MONGOLISM. Grune and Stratton, New York.

Benda, C.E. (1969) DOWN'S SYNDROME: MONGOLISM AND ITS MANAGEMENT (Revised Edition). Grune and Stratton, New York.

Benjamin, B.J. (1981) Frequency variability in the aged voice. Journal of Gerontology 36: 722-726.

Berg, van den, J. (1955) On the role of the laryngeal ventricle in voice production. Folia Phoniatrica, 7:57-69.

Berg, van den, J. (1962) Modern research in experimental phoniatrics. Folia Phoniatrica 14: 81-149.

Berg, van den, J (1968) Mechanism of the larynx and laryngeal vibrations. pp.278-308 in Malmberg, B., MANUAL OF PHONETICS. North-Holland, London.

Berg, van den, J., Vennard, W., Berger, D. and Shervanian, C.C. (1960) VOICE PRODUCTION. THE VIBRATING LARYNX (Film) SFW-UNFI, Utrecht.

Bergersen, E.O. (1972) The male adolescent facial growth spurt: its prediction and relation to skeletal maturation. Angle Orthodontist, 42:319-338.

Berry, R.J., Epstein, R., Fourcin, A.J., Freeman, M., McCurtain, F. and Noscoe, N.J. (1982) An objective analysis of voice disorder. British Journal of Disorders of Communication, 17:67-83.

Bevis, R.R., Hayles, A.B., Isaacson, R.J. and Sather, A.H. (1977) Facial growth response to human growth hormone in hypopituitary dwarfs. Angle Orthodontist, 47:193-205.

Bezooijen, R.A.M.G. van (1984) Characteristics and recognizability of vocal expressions of emotion. Doctoral thesis, Catholic University of Nijmegen.

Birrell, J.F. (1977) LOGAN TURNER'S DISEASES OF THE NOSE, THROAT AND EAR (8th. Edition.). John Wright and Sons Ltd., Bristol.

Bjork, A. (1966) Sutural growth of the upper face studied by the implant method. Acta Odontologica Scandinavica 24:109-

127.

Blanchard, I. (1964) Speech pattern and etiology in mental retardation. American Journal of Mental Deficiency 68: 612-617.

Bloom, W. and Fawcett, D.W. (Eds.) (1968) A TEXTBOOK OF HISTOLOGY (9th. Edition). Saunders, Philadelphia.

Boone, D.R. (1977) THE VOICE AND VOICE THERAPY. Prentice-Hall, New Jersey.

Bourne, G.H. (Ed.) (1961) STRUCTURAL ASPECTS OF AGEING. Hafner, New York.

Bosma, J.F. (1963) Maturation of function of the oral and pharyngeal region. American Journal of Orthodontics, 49:94-104.

Brasel, J.A. and Gruen, R.K. (1986) Cellular growth: brain, heart, lung, liver and skeletal muscle. pp. 53-65 in Falkner, F. and Tanner, J.M. (Eds.), HUMAN GROWTH, Vol.1. Plenum Press, New York.

Bristow, G. (1980) A speech training system for the deaf using computer colour graphics. Ph.D. Dissertation, University of Cambridge.

Broad, D.J. (1977) SHORT COURSE IN SPEECH SCIENCE. Speech Communications Research Laboratory, Santa Barbara.

Brooks, D.N., Woolley, H. and Kanjilal, G.C. (1972) Hearing loss and middle-ear disorders in patients with Down's Syndrome. Journal of Mental Deficiency Research, 16:21.

Brousseau, K. and Brainerd, H.G. (1928) MONGOLISM: A STUDY OF THE PHYSICAL AND MENTAL CHARACTERISTICS OF MONGOLOID IMBECILES. Williams and Wilkins, Baltimore.

Brown, R.H. and Cunningham, W.M. (1961) Some dental manifestations of mongolism. Oral Surgery, 14:664-676.

Brushfield, T. (1924) Mongolism. British Journal of Child Disorders, 21:241.

Butterworth, T., Leoni, E.P., Beerman, H., Wood, M.G. and Stern, L.P. (1960) Chelitis of mongolism. Journal of Investigative Dermatology, 35:347.

Catford, J.C. (1977) FUNDAMENTAL PROBLEMS IN PHONETICS. Edinburgh University Press.

Chiba, T. and Kajiyama, M. (1958) THE VOWEL: ITS NATURE AND STRUCTURE. Phonetic Society of Japan, Tokyo.

Clegg, A.G. and Clegg, P.C. (1963) BIOLOGY OF THE MAMMAL

* Davis, S.B. (1976) Computer evaluation of laryngeal pathology based on inverse filtering of speech. Speech Communication Research Lab., Santa Barbara, SCRL Monograph 13.

(2nd. Edition). Heinemann, London.

Cohen, M.M., Arvystas, M.G. and Baum, B.J. (1970) Occlusal dysharmonies in trisomy G (Down's Syndrome, mongolism). American Journal of Orthodontics, 58:367-372.

Cohen, M.M. and Cohen, M.M. (1971) The oral manifestations of trisomy 21 (Down's Syndrome). Birth Defects: Original Article Series, 7:241-251.

Cohen, M.M. and Winer, R.A. (1965) Dental and facial characteristics in Down's Syndrome. Journal of Dental Research, 44:197-207.

Coleman, R.O. (1971) Male and female voice quality and its relationship to vowel formant frequencies. Journal of Speech and Hearing Research, 14:565-77.

Comfort, A. (1965) THE PROCESS OF AGEING. Weidenfeld and Nicolson, London.

Crome, L.C., Cowie, V. and Slater, E. (1966) A statistical note on cerebellar and brain stem weight in mongolism. Journal of Mental Deficiency Research, 10:69-72.

Currie, G. and Currie, A. (1982) CANCER: THE BIOLOGY OF MALIGNANT DISEASE. Edward Arnold, London.

Curry, E.T. (1940) The pitch characteristics of the adolescent male. Speech Monographs, 7:48-62.

Davies, D.V. and Davies, F. (1962) GRAYS ANATOMY (33rd. Edition). Longmans, Green and Co. Ltd., London.

*

Deal, R. and Emanuel, F. (1978) Some waveform and spectral features of vowel roughness. Journal of Speech and Hearing Research, 21:250-264.

Dermaut, L.R. and O'Reilly, M.I.T. (1978) Changes in anterior facial height in girls during puberty. Angle Orthodontist, 48:163-171.

Dickson, D.R. and Maue-Dickson, W. (1982) ANATOMICAL AND PHYSIOLOGICAL BASES OF SPEECH. Little, Brown and Company, Boston.

Down, J.L. (1866) Observations on ethnic classification of idiots. London Hospital Reports, No. 3.

Duffy, R.J. (1958) The vocal pitch characteristics of eleven-, thirteen-, and fifteen-year-old female speakers. Dissertation, State University of Iowa. Dissertation Abstracts, 18:599.

Emery, A.E.H. (1979) ELEMENTS OF MEDICAL GENETICS (5th. Edition). Churchill Livingstone, Edinburgh.

Emery, J.L. (Ed.) (1979) THE ANATOMY OF THE DEVELOPING LUNG. Heinemann; Spastics International Medical Publications, London.

Endres, W., Bambach, W. and Flosser, G. (1971) Voice spectrograms as a function of age, voice disguise, and voice imitation. Journal of the Acoustical Society of America, 49:1842-8.

Engler, M. (1949) Mongolism. J. Wright, Bristol.

Enlow, D.H. and Harris, D.B. (1964) A study fo the postnatal growth of the mandible. American Journal of Orthodontics, 50:25.

Esling, J.H. (1978) Voice quality in Edinburgh: a sociolinguistic and phonetic study. Ph.D. Dissertation, University of Edinburgh.

Espir, M.L.E. and Rose, F.C. (1976) THE BASIC NEUROLOGY OF SPEECH (2nd. Edition). Blackwell Scientific Publications, Oxford.

Eveleth, P.B. and Tanner, J.M. (1976) WORLDWIDE VARIATION IN HUMAN GROWTH. Cambridge University Press, London.

Fairbanks, G. (1942) An acoustical study of the pitch of infant hunger wails. Child Development, 13:227-32.

Fairbanks, G. (1960) VOICE AND ARTICULATION DRILL BOOK (2nd. Edition). Harper and Row, New York.

Fairbanks, G., Herbert, E.S. and Hammond, J.M. (1949) An acoustical study of vocal pitch in seven and eight-year-old girls. Child Development, 20:71-8.

Fairbanks, G., Wiley, J.H. and Lassman, F.M. (1949) An acoustical study of vocal pitch in seven and eight-year-old boys. Child Development, 20:63-9.

Falkner, F. and Tanner, J.M. (Eds.) (1986) HUMAN GROWTH (2nd. Edition). Plenum Press, New York.

Fant, G. (1960) ACOUSTIC THEORY OF SPEECH PRODUCTION. Mouton, The Hague.

Fant, G. (1966) A note on vocal tract size factors and non-uniform f-pattern scalings. Quarterly Progress and Status Report, 4:22-30, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm.

Farnsworth, D.W. (1940) High speed motion pictures of the human vocal cords. Bell Laboratories Record 18: 203-208.

Ferlito, A. (1974) Histological classification of larynx and

hypopharynx cancer. Acta Otolarngologica Supplements 342: 17.

Fields, S. and Dunn, F. (1973) Correlation of echographic visuability of tissue with biological composition and physiological state. Journal of the Acoustical Society of America 54: 809-812.

Fletcher, R.F. (1978) LECTURE NOTES ON ENDOCRINOLOGY. Blackwell Scientific Publications, Oxford.

Fraser, W.I. (1978) Speech and language development of children with Down's Syndrome. Developmental Medicine and Child Neurology, 20:106-109.

Freeman, W.H. and Bracegirdle, B. (1967) AN ATLAS OF HISTOLOGY (2nd. Edition). Heinemann Educational Books Ltd., London.

Friedmann, I. and Osborn, D.A. (1978) The larynx. in W. St. C. Symmers (Ed.) SYSTEMIC PATHOLOGY, Vol. 1: 248-267.

Friend, G.E. and Bransby, E.R. (1947) Physique and growth of schoolboys. Lancet, 2:677.

Fritzell, B., Sundberg, J. and Strange-Ebbesen, A. (1982) Pitch change after stripping oedematous vocal folds. Folia Phoniatrica 34: 29-32.

Frostad, W.A., Cleall, J.F. and Melosky, L.C. (1971) Craniofacial complex in the trisomy 21 syndrome (Down's Syndrome). Archives of Oral Biology, 16:707-722.

Fulton, R.T. and Lloyd, L.L. (1968) Hearing impairment in a population of children with Down Syndrome. American Journal of Mental Deficiency, 73:298.

Garn, S.M. and Clark, D.C. (1975) Nutrition, growth, development and maturation: findings from the ten-state nutrition survey of 1968-1970. Pediatrics, 56:306-319.

Gold, B. and Rabiner, L.R. (1969) Parallel processing techniques for estimating pitch periods of speech in the time domain. Journal of the Acoustical Society of America, 46:442-448.

Goldspink, G. (Ed.) (1974) DIFFERENTIATION AND GROWTH OF CELLS IN VERTEBRATE TISSUES. Chapman and Hall, London.

Goodman, R.M. and Gorlin, R.J. (Eds.) (1970) THE FACE IN GENETIC DISORDERS. Mosby, St. Louis.

Greene, M.C.L. (1972) THE VOICE AND ITS DISORDERS (3rd. Edition). Lippincott, Philadelphia.

Goerttler, K. (1950) Die anordnung, histologie und

histogenese der quergestreiften muskulatur in menschlichen stimmband. Zeitschrift fur Anatomie und Entwickelungsgeschichte 115: 352-401.

Gosman, S.D. (1951) Facial development in mongolism. American Journal of Orthodontics, 37:332-349.

Hall, S.I. and Colman, B.H. (1975) DISEASES OF THE NOSE, THROAT AND EAR: A HANDBOOK FOR STUDENTS AND PRACTITIONERS. Churchill Livingstone: Edinburgh.

Hanley, T.D. (1951) An analysis of vocal frequency and duration characteristics of selected dialect regions. Speech Monographs, 18:78-93.

Hardcastle, W.J. (1977) THE PHYSIOLOGY OF SPEECH PRODUCTION. Academic Press, New York.

Hartlieb, K. (1962) Erbliche Merkmale der menschlichen Stimme. Zeitschrift fur menschliche Vererbung und Konstitutionslehre, 36:413.

Hasek, C.S., Singh, S. and Murry, T. (1980) Acoustic attributes of preadolescent voices. Journal of the Acoustical Society of America, 68:1262-1265.

Helfrich, H. (1979) Age markers in speech. pp. 63-107 in Scherer, K.R. and Giles, H. (Eds.) SOCIAL MARKERS IN SPEECH. Cambridge University Press, Cambridge.

Hiller, S.M. (1985) Automatic acoustic analysis of waveform perturbations. Ph.D. Dissertation, University of Edinburgh.

Hiller, S.M., Laver, J. and Mackenzie, J. (1983) Automatic analysis of waveform perturbations in connected speech. Work in Progress, University of Edinburgh, Department of Linguistics 16: 40-68.

Hiller, S.M., Laver, J. and Mackenzie, J. (1984) Durational aspects of long-term measurements of fundamental frequency perturbations in connected speech. Work in Progress, University of Edinburgh, Department of Linguistics 17:59-76.

Hirano, M. (1974) Morphological structure of the vocal fold as a vibrator and its variations. Folia Phoniatrica 26: 89-94

Hirano, M. (1975) Phonosurgery. Basic and clinical investigations. Otologia Fukuoka, 21:239-440.

Hirano, M. (1981) CLINICAL EXAMINATION OF VOICE. Springer-Verlag, New York.

Hirano, M., Gould, W.J., Lambiase, A. and Kakita, Y. (1981)

Vibratory behaviour of the vocal folds in a case with a unilateral polyp. Folia Phoniatrica 33: 275-284.

Hirano, M., Kurita, S. and Nakashima, T. (1981) The structure of the vocal folds. In K.N. Stevens and M. Hirano (Eds.) VOCAL FOLD PHYSIOLOGY. University of Tokyo Press, Tokyo.

Hirano, M., Kakita, Y., Ohmaru, K. and Kurita, S. (1982) Structure and mechanical properties of the vocal fold. In N. Lass (Ed.) SPEECH AND LANGUAGE: ADVANCES IN BASIC RESEARCH AND PRACTICE. Academic Press, New York: 211-297.

Hiroto, I. (1966) Patho-physiology of the larynx from the stand-point of vocal mechanism. Practica Otologica Kyoto 59: 229-292.

Hollien, H. (1971) Three major vocal registers: a proposal. Proceedings of the 7th. International Congress of Phonetic Sciences, Montreal: 320-331.

Hollien, H. and Copeland, R.H. (1965) Speaking fundamental frequency (SFF) characteristics of mongoloid girls. Journal of Speech and Hearing Disorders, 30:344-349.

Hollien, H., Dew, D. and Philips, P. (1971) Phonational frequency ranges of adults. Journal of Speech and Hearing Research 14: 755-760.

Hollien, H. and Jackson, B. (1967) Normative SSF data on southern male university students. Progress report to NIH, Grant NB-OX397.

Hollien, H. and Malcik, E. (1962) Adolescent voice change in southern Negro males. Speech Monographs 24: 53-58.

Hollien, H., Malcik, E. and Hollien, B. (1965) Adolescent voice changes in southern White males. Speech Monographs, 32:87-90.

Hollien, H. and Michel, J.F. (1968) Vocal fry as a phonational register. Journal of Speech and Hearing Research, 11:600-604.

Hollien, H. and Paul, P. (1969) A second evaluation of the speaking fundamental frequency characteristics of post-adolescent girls. Language and Speech 12: 119-124.

Hollien, H. and Shipp, F.T. (1972) Speaking fundamental frequency and chronological age in males. Journal of Speech and Hearing Research 15: 155-159.

Honikman, B. (1964) Articulatory settings. pp. 73-84 in Abercrombie, D., Fry, D.B., MacCarthy, P.A.D., Scott, N.C. and Trim, J.L.M. (Eds.) IN HONOUR OF DANIEL JONES, Longmans, London.

Honjo, I. and Isshiki, N. (1980) Laryngoscopic and voice characteristics of aged persons. Archives of Otolaryngology 106: 149-150.

Hopkin, G.B. (1967) Neonatal and adult tongue dimensions. Angle Orthodontist, 37:132-133.

Hopkin, G.B. (1978) THE DENTITION AND SPEECH. Leaflet prepared for Speech Therapy students, Edinburgh.

House, A.S. and Stevens, K.N. (1956) Analog studies of the nasalization of vowels. Journal of Speech and Hearing Disorders, 21:218-231.

Hunter, C.J. (1966) The correlation of facial growth with body height and skeletal maturity at adolescence. Angle Orthodontist, 36:44-54.

Ingerslev, C.H. and Solow, B. (1975) Sex differences in craniofacial morphology. Acta Odontologica Scandinavica, 33:85-94.

Ishizaka, K. and Flanagan, J.L. (1972) Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell System Technical Journal 51: 1233-1268.

Iwata, S. and Leden, von H. (1970) Pitch perturbations in normal and pathological voices. Folia Phoniatrica, 22:413-424.

Jackson, J. (1979) Personal communication.

Jensen, G.M., Cleall, J.F. and Yips, A.S.G. (1973) Dentoalveolar morphology and developmental changes in Down's Syndrome. American Journal of Orthodontics, 64:607-618.

Jones, H.B. (1963) An investigation to determine the validity of voice quality as a criterion of mongolism. Unpublished M.A. Thesis, Hunter College.

Joseph, M. and Dawbarn, C. (1970) MEASUREMENT OF THE FACIES: A STUDY IN DOWN'S SYNDROME. Spastics International Medical Publications Research Monograph No. 3. Heinemann, London.

Junqueira, L.C. and Carneiro, J. (1980) BASIC HISTOLOGY (3rd. Edition). Lange Medical Publications, Los Altos.

Kahane, J.C. (1978) Histomorphological study of the aging male larynx. American Speech and Hearing Association, 20:747.

Kane, M. and Wellen, C.J. (1985) Acoustical measurements and clinical judgements of voice quality in children with vocal nodules. Folia Phoniatrica, 37:53-57.

* Kent, R.D.(1981) Sensorimotor aspects of speech development. pp. 161-189 in DEVELOPMENT OF PERCEPTION, Vol. 1. Academic Press, New York.

Kaplan, H.M. (1960) ANATOMY AND PHYSIOLOGY OF SPEECH. McGraw-Hill, New York.

Kasuya, H., Kobayashi, Y. and Kobayashi, T. (1983) Characteristics of pitch period and amplitude perturbations in pathologic voice. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1372-1375.

Keating, P. and Buhr, R. (1978) Fundamental frequency in the speech of infants and children. Journal of the Acoustical Society of America, 63:567-71.

Kenedi, R.M. (Ed.) (1980) A TEXTBOOK OF BIOMEDICAL ENGINEERING. Blackie, Glasgow.

← *
Kisling, E. CRANIAL MORPHOLOGY IN DOWN'S SYNDROME: A COMPARATIVE ROENTGENCEPHALOMETRIC STUDY IN ADULT MALES. Munksgaard, Copenhagen.

Kitajima, K., Tanabe, M. and Isshiki, N. (1975) Pitch perturbations in normal and pathological voices. Studia Phonetica, 9:25-32.

Klecka, W. (1980) DISCRIMINANT ANALYSIS. (Sage University Paper Series on Quantitative Applications in the Social Sciences 17-001). Sage, Beverly Hills, London.

Kleinsasser, O. (1968) MICROLARYNGOSCOPY AND ENDOLARYNGEAL MICROSCOPY. Saunders, London.

Knott, V.B. (1961) Size and form of the dental arches in children with good occlusion studied longitudinally from age 9 years to late adolescence. American Journal of Physical Anthropology, 19:263-284.

Koike, Y. (1973) Application of some acoustic measures for the evaluation of laryngeal dysfunction. Studia Phonetica, 7:17-23.

Koike, Y., Takahashi, H. and Calcaterra, T.C. (1977) Acoustic measures for detecting laryngeal pathology. Acta Otolaryngology, 84: 105-117.

Laitman, J.T. and Crelin, E.S. (1975) Postnatal development of the basicranium and vocal tract region in man. pp. 206-219 in Bosma, J.F. (Ed.) SYMPOSIUM ON DEVELOPMENT OF THE BASICRANIUM. National Institute of Health, Bethesda.

Laver, J. (1968) Voice quality and indexical information. British Journal of Disorders of Communication 3:43-54.

Laver, J. (1974) Labels for voices. Journal of the International Phonetic Association 4:62-75.

Laver, J. (1975) Individual features in voice quality.

Ph.D. dissertation, University of Edinburgh.

Laver, J. (1980) THE PHONETIC DESCRIPTION OF VOICE QUALITY. Cambridge University Press.

Laver, J. and Hanson, R.J (1981) Describing the normal voice. pp. 51-78 in Darby, J. (Ed.) SPEECH EVALUATION IN PSYCHIATRY. Grune and Stratton, New York.

Laver, J., Hiller, S. and Mackenzie, J. (1984) Acoustic analysis of vocal fold pathology. Proceedings of the Institute of Acoustics, 6:425-430.

Laver, J., Hiller, S., Mackenzie, J. and Rooney, E. (1985) An acoustic screening system for the detection of laryngeal pathology. Symposium on Voice Acoustics and Dysphonia, Katthammarsvik, Sweden.

Laver, J., Hiller, S., Mackenzie, J. and Rooney, E. (1986) An acoustic screening system for the detection of laryngeal pathology. Journal of Phonetics, 14:517-524.

Laver, J. and Trudgill, P. (1979) Phonetic and linguistic markers in speech. pp.1-32 in Scherer, K.R. and Giles, H. (Eds.), SOCIAL MARKERS IN SPEECH, Cambridge University Press.

Laver, J., Wirz, S., Mackenzie, J. and Hiller, S. (1981) A perceptual protocol for the analysis of vocal profiles. Work in Progress, University of Edinburgh, Department of Linguistics 14: 139-155.

Laver, J., Wirz, S., Mackenzie, J. and Hiller, S. (1982) Vocal profiles of speech disorders. Final Report, Medical Research Council Project No. G7811925N.

Leeson, C.R. and Leeson, S.L. (1976) HISTOLOGY (3rd. Edition). W.B. Saunders Company, Philadelphia

Lejeune, J., Turpin, R. and Gautier, M. (1959) Mongolism, a chromosomal illness. Bull. Acad. Nat. Med. (Paris), 143:256-265.

Lemperle, G. and Radu, D. (1980) Facial plastic surgery in children with Down's Syndrome. Plastic and Reconstructive Surgery, 66:337-345.

Levinson, A., Friedman, A. and Stamps, F. (1955) Variability of Mongolism. Pediatrics, 16:43-54.

Leudar, I., Fraser, W.I. and Jeeves, M.A. (1981) Social familiarity and communication in Down syndrome. Journal of Mental Deficiency Research 25 (Pt. 2): 133-142.

Lind, J., Vuorenkoski, V., Rosberg, G., Partanen, T.J. and Wasz-Hockert, O. (1970) Spectrographic analysis of vocal

response to pain stimuli in infants with Down's Syndrome. Developmental Medicine and Child Neurology, 12:478-486.

Linke, E. (1953) A study of pitch characteristics and their relationship to vocal effectiveness. Ph.D. Dissertation, State University of Iowa.

Linville, S.E. and Fisher, H.B. (1985) Acoustic characteristics of womens' voices with advancing age. Journal of Gerontology, 40:324-330.

Luchsinger, R. (1962) Voice disorders on an endocrine basis. Chapter 2 in Levin, N.M. (Ed.) VOICE AND SPEECH DISORDERS: MEDICAL ASPECTS. Thomas, Springfield.

Luchsinger, R. (1970) DIE STIMME UND IHRE STORUNGEN (3rd. Edition). Vienna.

Luchsinger, R. and Arnold, G.E. (1965) VOICE-SPEECH-LANGUAGE. CLINICAL COMMUNICOLOGY: ITS PHYSIOLOGY AND PATHOLOGY. Constable, London.

Ludlow, C.L., Coulter, D.C. and Gentges, F.H. (1983a) The differential sensitivity of frequency perturbation to laryngeal neoplasms and neuropathologies. pp.381-392 in Bless, D.M. and Abbs, J.H. (Eds.) VOCAL FOLD PHYSIOLOGY: CONTEMPORARY RESEARCH AND CLINICAL ISSUES. College Hill, San Diego.

Ludlow, C.L., Coulter, D.C. and Gentges, F.H. (1983b) The effects of change in vocal fold morphology on phonation. pp.77-89 in Lawrence, V.L. (Ed.) Transcripts of the 11th. Symposium on Care of the Professional Voice, Part I: Scientific Sessions, Papers. The Voice Foundation, New York.

McIntyre, M.S. and Dutch, S.J. (1964) Mongolism and general hypotonia. Americal Journal of Mental Deficiency, 68:669-670.

Mackenzie, J., Laver, J. and Hiller, S.M.. (1983) Structural pathologies of the vocal folds and phonation. Work in Progress, University of Edinburgh, Department of Linguistics 16: 80-116.

Mackenzie, J., Laver, J. and Hiller, S. (1984) Acoustic screening for vocal pathology: preliminary results. Work in Progress, University of Edinburgh, Department of Linguistics 17: 98-110.

Majewski, W., Hollien, H. and Zalewski, J. (1972) Speaking fundamental frequency of Polish adult males. Phonetica, 25:119-25.

Malina, R.M. (1980) Growth of muscle tissue and muscle mass. pp. 77-99 in Falkner, F. and Tanner, J.M. (Eds.) HUMAN

GROWTH. Plenum Press, New York.

Marshall, W.A. (1981) Geographical and ethnic variations in human growth. British Medical Bulletin, 37:273-279.

Martin, D. (1961) Some facies in the diseases of childhood. Medical and Biological Illustration, 11:76-84.

Matsushita, H. (1969) Vocal cord vibration of excised larynges - study with ultra-high-speed cinematography. Otologia Fukuoka 15: 127-142 (in Japanese).

Maw, A.R., Cullen, R.J. and Bradfield, J.W.B. (1982) Verrucous carcinoma of the larynx. Clinical Otolaryngology 7: 305-311.

McGlone, R.E. and Hollien, H. (1963) Vocal pitch characteristics of aged women. Journal of Speech and Hearing Research, 6:164-70.

Meditch, A. (1975) The development of sex-specific speech patterns in young children. Anthropological Linguistics, 17:421-33.

Michaels, L. (1976) Histopathology of nose and throat. pp. 667-700 in R. Hinchcliffe and D. Hamson (Eds.) SCIENTIFIC FOUNDATIONS OF OTOLARYNGOLOGY. William Heinemann Medical Books Ltd., London.

Michel,J.F., Hollien, H. and Moore, P. (1966) Speaking fundamental frequency characteristics of 15, 16 and 17 year-old girls. Language and speech, 9:46-51.

Montague, J.C. (1976) Perceived age and sex characteristics of voices of institutionalised children with Down's Syndrome. Perceptual and Motor Skills, 42:215-219.

Montague, J.C., Brown, W.S. and Hollien, H. (1974) Vocal fundamental frequency characteristics of institutionalised D.S. children. American Journal of Mental Deficiency, 78:414-418.

Montague, J.C. and Hollien, H. (1973) Perceived voice quality disorders in Down's Syndrome children. Journal of Communication Disorders, :76-87.

Montague, J.C., Hollien, H., Hollien, B. and Wold, D.C. (1978) Perceived pitch and F.F. comparisons of institutionalised D.S. children. Folia Phoniatrica, 30:245-256.

Morris (1953) HUMAN ANATOMY (11th. Edition) Ed. by J.P. Schaeffer. McGraw-Hill, New York.

Mueller, P.B., Sweeney, R.J. and Baribeau, L.J. (1985) Senescence of the voice: morphology of excised male

larynges. Folia Phoniatrica 37: 134-138.

Murray, T. and Doherty, E.T. (1980) Selected acoustic characteristics of pathological and normal speakers. Journal of Speech and Hearing Research, 23:361-369.

Mysak, E.D. (1959) Pitch and duration characteristics of older males. Journal of Speech and Hearing Research, 2:46-54.

Negus, V.E. (1949) THE COMPARATIVE ANATOMY AND PHYSIOLOGY OF THE LARYNX. Heinemann Medical, London.

New, G.B. and Erich, J.B. (1938) Benign tumours of the larynx: a study of 722 cases. Archives of Otolaryngology 28: 841.

Nieuwenhuis, R. and Mackenzie, J. (1986) The use of two voice analysis techniques in clinic. The College of Speech Therapists Bulletin, 412:1-3.

Nolan, M., McCartney, E., McArthur, K. and Rowson, V.J. (1980) A study of the hearing and receptive vocabulary of the trainees of an adult training centre. Journal of Mental Deficiency Research, 24:271-286.

Novak, A. (1972) The voice of children with Down's Syndrome. Folia Phoniatrica, 24:182-194.

O'Reilly, M.T. (1979) A longitudinal growth study: maxillary length at puberty in females. Angle Orthodontist, 49:234-238.

Øster, J. (1953) MONGOLISM. Danish Science Press Ltd., Copenhagen.

Ostwald, P.F. (1963) SOUNDMAKING: THE ACOUSTIC COMMUNICATION OF EMOTION. Springfield, Illinois.

Ostwald, P.F., Phibbs, R. and Fox, S. (1968) Diagnostic use of infant cry. Biology of Neonates, 13:68-82.

Pantoja, E. (1968) The laryngeal cartilages. Physiologic nonmineralization masquerading malignant destruction. Archives of Otolaryngology, 87:416-421.

Pearce, F., Rankine, R. and Ormond, A. (1910) Notes on 28 cases of mongolian imbeciles. British Medical Journal, 2:186.

Penrose, L.S. (1963) THE BIOLOGY OF MENTAL DEFECT (3rd. Edition). Sidgwick and Jackson, London.

Perello, J. (1962) The muco-undulatory theory of phonation. Annals of Otolaryngology 79: 722-725.

Perkins, H. (1977) SPEECH PATHOLOGY, AN APPLIED BEHAVIOURAL SCIENCE. The C.V. Mosby Co., St, Louis.

Pritchard, J.J. (1974) Growth and differentiation of bone and connective tissue. pp. 101-128 in Goldspink, G. DIFFERENTIATION AND GROWTH OF CELLS IN VERTEBRATE TISSUES. Chapman and Hall, London.

Ptacek, P.H. and Sander, E.K. (1966) Age recognition from the voice. Journal of Speech and Hearing Research, 9:273-277.

Ptacek, P.H., Sander, E.K., Maloney, W.H. and Jackson, C.C.R. (1966) Phonatory and related changes with advanced age. Journal of Speech and Hearing Research, 9:353-60.

Rabiner, L.R., Sambur, M.R. and Schmidt, C.E. (1975) Applications of nonlinear smoothing algorithm to speech processing. IEEE Transactions on Acoustics Speech and Signal Processing, 23:552-557.

Ramig, L.A. and Ringel, R.L. (1983) Effects of physiological aging on selected acoustic characteristics of voice. Journal of Speech and Hearing Research 26: 22-30.

Redman, R.S., Shapiro, B.L. and Gorlin, R.J. (1966) Measurement of normal and reportedly malformed palatal vaults. II. Normal juvenile measurements. Journal of Dental Research, 45:266-269.

Ringel, R. and Klungel, D. (1964) Neonatal crying. A normative study. Folia phoniatrica, 16:1-9.

Roche, A.F. (1965) The stature of Mongols. Journal of Mental Deficiency Research 9: 131-145.

Roche, A.F., Roche, P.J. and Lewis, A.B. (1972) The cranial base in trisomy 21. Journal of Mental Deficiency Research, 16:7.

Rolfe, C.R., Montague, J.C., Tirman, R.M. and Vandergrift, J.F. (1979) Pilot perceptual and physiological investigation of hypernasality in Down's Syndrome adults. Folia Phoniatrica, 31:177-187.

Romanes, G.J. (Ed.) (1978) CUNNINGHAM'S MANUAL OF PRACTICAL ANATOMY, Vol. 3, Head, Neck and Brain (14th. Edition). Oxford University Press, Oxford.

Rona, R.J. (1981) Genetic and environmental factors in the control of growth in childhood. British Medical Bulletin, 37:265-272.

Rose, G.J.A. (1953) A quantitative study of the facial areas from the profile roentgenograms and the relationships to body measurements. Abstract in American Journal of

Orthodontics, 39:59.

Salmon, L.F.W. (1979a) Acute laryngitis. pp. 345-380 in J. Ballantyne and J. Groves (Eds.) SCOTT-BROWN'S DISEASES OF THE EAR NOSE AND THROAT (4th. Edition), Volume 4.

Salmon, L.F.W. (1979b) Chronic laryngitis. pp.381-420 in J. Ballantyne and J. Groves (Eds.) SCOTT-BROWN'S DISEASES OF THE EAR NOSE AND THROAT (4th. Edition), Volume 4.

Sandritter, W. and Wartman, W.B. (1969) COLOUR ATLAS AND TEXTBOOK OF TISSUE AND CELLULAR PATHOLOGY 4th. Edition). Year Book Medical Publishers Inc., Chicago.

Saunders, W.H. (1964) THE LARYNX. CIBA Corp., New Jersey.

Saville, D. (1984) Personal communication.

Saxman, J.H. and Burk, K.W. (1967) Speaking fundamental frequency characteristics of middle-aged females. Folia Phoniatrica, 19:167-172.

Scherer, K.R. and Giles, H. (Eds.) (1979) SOCIAL MARKERS IN SPEECH. Cambridge University Press, Cambridge.

Schlanger, B.B. and Gottsleben, K.H. (1957) Analysis of speech defects among institutionalised mentally retarded. Journal of Speech and Hearing Disorders 22: 98-103.

Schwartz, M.F. and Rine, H.E. (1968) Identification of speakers from whispered vowels. Journal of the Acoustical Society of America, 44:1736-7.

Scott, J.H. (1967) DENTO-FACIAL DEVELOPMENT AND GROWTH. Pergamon, Oxford.

Sedlackova, E., Vrticka, K. and Supacek, I. (1966) Das altern der stimme. Proceedings of the 7th. International Congress of Gerontology, Vienna. Clinical Medicine, vol. iv, 7:469-72.

Shah, P.J., Joshi, M.R. and Darnwala, N.R. (1980) The interrelationships between facial areas and other body dimensions. Angle Orthodontist, 50:45-53.

Shapiro, B.L. (1973) Amplified developmental instability in Down's Syndrome. Annals of Human Genetics, 38:429-437.

Shapiro, B.L. (1970) Prenatal dental anomalies in mongolism: comments on the basis and implications of variability. Annals of the New York Academy of Science 171: 562-567.

Shapiro, B.L., Gorlin, R.J., Redman, R.S. and Bruhl, H.H. (1967) The palate and Down's Syndrome. New England Journal of Medicine, 276:1460-1463.

Shapiro, B.L., Redman, R.S. and Gorlin, R.J. (1963) Measurement of normal and reportedly abnormal palatal vaults. I. Normal adult measurements. Journal of Dental Research, 42:1039.

Shaw, H. (1979) Tumours of the larynx. pp. 421-508 in J. Ballantyne and J. Groves (Eds.) SCOTT-BROWN'S DISEASES OF THE EAR, NOSE AND THROAT (4th. Edition), Volume 4.

Shuttleworth, G.E. (1909) Mongolian idiocy. British Medical Journal, 2:661.

Siegel, S. (1956) NONPARAMETRIC STATISTICS FOR THE BEHAVIOURAL SCIENCES. McGraw-Hill Kogakusha, Tokyo.

Sinclair, D. (1978) HUMAN GROWTH AFTER BIRTH (3rd. Edition). Oxford University Press, Oxford.

Smith, G.F. and Berg, J.M. (1976) DOWN'S ANOMALY (2nd. Edition). Churchill Livingstone, Edinburgh.

Smith, P.M. (1979) Sex markers in speech. pp. 109-46 in K.R. Scherer and H. Giles (Eds.), SOCIAL MARKERS IN SPEECH, Cambridge University Press, Cambridge.

Smith, S. (1956) Mouvements des cordes vocales (Film No. 4). Government Film Office, Copenhagan.

Smith, S. (1961) On artificial voice production. PROCEEDINGS OF THE 4TH. INTERNATIONAL CONGRESS OF PHONETIC SCIENCES, Helsinki: 96-110.

SPSSx USER'S GUIDE. (1983) McGraw-Hill, New York.

Stark, R.E., Rose, S.N. and McLagen, M. (1975) Features of infant sounds: the first eight weeks of life. Journal of Child Language, 2:205-21.

Stevens, K.N. and House, A.S. (1961) An acoustical theory of vowel production and some of its implications. Journal of Speech and Hearing Research, 4:303-320.

Strazulla, M. (1953) Speech problems of the mongoloid child. Quarterly Review of Pediatrics 8: 268-273.

Strome, M. (1981) Down's Syndrome: a modern otorhinolaryngological perspective. Laryngoscope, 91:1581-1594.

Swallow, J.N. (1964) Dental disease in children with Down's Syndrome. Journal of Mental Deficiency Research, 8:102-118.

Tanner, J.M. (1978) FOETUS INTO MAN: PHYSICAL GROWTH FROM CONCEPTION TO MATURITY. Open Books, London.

Tanner, J.M. and Whitehouse, R.H. (1976) Clinical longitudinal standards for height, weight, height velocity and weight velocity and the stages of puberty. Archives of Disease in Childhood, 51: 170-179.

Terracol, J., Guerrier, Y. and Camps, F. (1956) Le sphincter glottique; etude anatomo-clinique. Annales d' Otolaryngologie (Paris) 73: 451.

Thelander, H.E. and Pryor, H.B. (1966) Abnormal patterns of growth and development in mongolism. Clinical Pediatrics, 5:493-501.

Titze, I.R. (1973) The human vocal cords: a mathematical model, Part 1. Phonetica 28: 129-170.

Titze, I.R. (1974) The human vocal cords: a mathematical model, Part 2. Phonetica 29: 1-21.

Titze, I.R. and Strong, W.J. (1975) Normal modes in vocal cord tissues. Journal of the Acoustical Society of America 57: 736-744.

Tofani, M.I. (1972) Mandibular growth at puberty. American Journal of Orthodontics, 62:176-195.

Toft, A., Campbell, I. and Seth, J. (1981) DIAGNOSIS AND MANAGEMENT OF ENDICRINE DISEASES. Blackwell Scientific Publications, Oxford.

Van Riper, R.C. and Irwin, J.V. (1958) VOICE AND ARTICULATION. Prentice-Hall, Englewood Cliffs.

Vieregge, W. (1981) Een transcriptie-systeem voor afwijkende spraak (T.C.P.S.). Logopedie en Foniatrie, 53:290-298.

Vuorenkoski, V., Lenko, H.L., Tjernlund, P., Vuorenkoski, L. and Perheentupa, J. (1978) Fundamental voice frequency during normal and abnormal growth, and after androgen treatment. Archives of Disorders in Childhood, 53:201-209.

Waddington, C.H. (1957) THE STRATEGY OF THE GENES. Allen and Unwin, London.

Wahi, P.N., Cohen, B., Luthra, U.K. and Torlini, H. (1971) HISTOLOGICAL TYPING OF ORAL AND OROPHARYNGEAL TUMOURS. International Histological Classification of Tumours No. 4. World Health Organization, Geneva.

Walker, G.W. and Kowalski, C.J. (1972) On the growth of the mandible. Americal Journal of Physical Anthropology, 36:111-118.

Wasz-Hockert, O., Lind, J. Vuorenkoski, I.V., Partanen, T. and Valanne, E. (1968) INFANT CRY. A SPECTROGRAPHIC AND AUDITORY ANALYSIS. Heinemann, London.

Wei, S.H.Y. (1970) Craniofacial width dimensions. Angle Orthodontist, 40:141-147.

Weinberg, B. and Zlatin, M. (1970) Speaking fundamental frequency characteristics of five- and six-year-old children with mongolism. Journal of Speech and Hearing Research, 13:418-425.

West, R., Ansberry, M. and Carr, A. (Eds.) (1947) THE REHABILITATION OF SPEECH (2nd. Edition). Harper, New York.

Westerman, G.H., Johnson, R. and Cohen, M.M. (1975) Variables of palatal dimensions in patients with Down's Syndrome. Journal of Dental Research, 54:767.

Wilcox, K.A. and Horii, Y. (1980) Age and changes in vocal jitter. Journal of Gerontology 35: 194-198.

Widdowson, E.M. (1951) Mental contentment and physical growth. Lancet, 1:1316-1318.

Winer, R.A. and Cohen, M.M. (1962) Dental caries in mongolism. Dental Progress, 2:217-219.

Winer, R.A., Cohen, M.M., Felter, R.F. and Channcey, H.H. (1965) Composition of human saliva, secretory rate, and electrolyte composition in mentally subnormal persons. Journal of Dental Research, 44:632.

Wirz, S.L. (1987) Vocal characteristics of hearing impaired people. Ph.D. Dissertation, University of Edinburgh.

Wynter, H. and Martin, S. (1981) The classification of deviant voice quality through auditory memory. British Journal of Disorders of Communication 16: 204-210.

APPENDIX ONE

APPENDIX ONE

VOCAL PROFILE ANALYSIS SCHEME:

A USER'S MANUAL

Janet Mackenzie

This manual is based on work by John Laver, Sheila Wirz and Janet Mackenzie.

VOCAL PROFILE ANALYSIS SCHEME: A USER'S MANUAL

INTRODUCTION

This manual is intended to act as a back-up to training courses in the VPAS, and as revision material for people already trained in the scheme. It is not sufficient in itself as a training in the scheme, but we hope that it will offer some practical hints about using the VPAS protocol. There are three sections: the first deals with some basic concepts of the scheme, the second is a guide to the protocol form and the third describes each setting, and gives some guidelines for the assignment of scalar degrees. A summary chart of setting characteristics is also included.

A. BASIC CONCEPTS.

There are several features about the VPAS with which all users should be familiar.

a) It considers the whole vocal tract. The lips, jaw and tongue may contribute to voice quality just as much as the larynx or the velopharyngeal system.

b) It analyses the voice in terms of different strands, or components, which may be combined in a variety of ways. These components are known as SETTINGS. A setting can be described as a long-term tendency for some part of the vocal tract to be held in a particular position. This should be thought of as some kind of long-term average position around which any short term movements which are needed for articulation of phonetic segments are made.

c) It is a perceptual scheme. Although each setting can be defined in terms of its normal acoustic and physiological correlates, use of the scheme is based on a knowlege of the perceptual quality associated with each setting. The VPAS relies on phonetic ear training in just the same way as segmental phonetic analysis.

d) All voices are compared with a clearly defined baseline, the NEUTRAL setting. This is defined in terms of acoustic and physiological correlates. Neutral is a convenient reference quality which should not be confused with any idea of normality. Almost all speakers deviate from neutral in some way.

B. USING THE VPA PROTOCOL FORM

The VPA protocol form was developed as the result of close collaboration with both speech therapists and phoneticians, and it is designed to be used for the description of both normal and pathological voices. The lay-out is intended to structure the listening task for the judge, and to give a clear picture of the type and degree of any deviations from the neutral setting. A copy of the protocol form is shown in Figure 1,

Vocal Profile Analysis Protocol

Speaker: ..... **Sex:** ..... **Age:** ..... **Date of Analysis:** ..... **Tape:** ..... **Judge:** .....

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS: Neutral	FIRST PASS: Non-neutral Normal	FIRST PASS: Non-neutral Abnormal	SECOND PASS: SETTING	Scalar Degrees: Normal 1	Normal 2	Normal 3	Abnormal 4	Abnormal 5	Abnormal 6
A. Supralaryngeal Features										
1. Labial				Lip Rounding/Protrusion						
				Lip Spreading						
				Labiodentalization						
				Extensive Range						
				Minimised Range						
2. Mandibular				Close Jaw						
				Open Jaw						
				Protruded Jaw						
				Extensive Range						
				Minimised Range						
3. Lingual Tip/Blade				Advanced						
				Retracted						
4. Lingual Body				Fronted Body						
				Backed Body						
				Raised Body						
				Lowered Body						
				Extensive Range						
				Minimised Range						
5. Velopharyngeal				Nasal						
				Audible Nasal Escape						
				Denasal						
6. Pharyngeal				Pharyngeal Constriction						
7. Supralaryngeal Tension				Tense						
				Lax						
B. Laryngeal Features										
8. Laryngeal Tension				Tense						
				Lax						
9. Larynx Position				Raised						
				Lowered						
10. Phonation Type				Harshness						
				Whisper(y)						
				Breathiness						
				Creak(y)						
				Falsetto						
				Modal Voice						

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

II PROSODIC FEATURES

CATEGORY	FIRST PASS: Neutral	FIRST PASS: Non-neutral Normal	FIRST PASS: Non-neutral Abnormal	SECOND PASS: SETTING	Scalar Degrees: Normal 1	Normal 2	Normal 3	Abnormal 4	Abnormal 5	Abnormal 6
1. Pitch				High Mean						
				Low Mean						
				Wide Range						
				Narrow Range						
				High Variability						
				Low Variability						
2. Consistency				Tremor						
3. Loudness				High Mean						
				Low Mean						
				Wide Range						
				Narrow Range						
				High Variability						
				Low Variability						

III TEMPORAL ORGANIZATION FEATURES

CATEGORY	FIRST PASS: Adequate	FIRST PASS: Inadequate	SECOND PASS: Scalar Degrees, Inadequate 1	2	3	
1. Continuity						Interrupted
2. Rate						Fast
						Slow

IV COMMENTS

	Adequate	Inadequate	1	2	3
Breath Support					
Rhythmicality					

	Present	Absent
Diplophonia		

Other Comments:

FIGURE 1: Vocal Profile Analysis Protocol

and this will be used as the basis of an outline of what a judge must do, when faced with an individual and asked to analyse that speaker's voice quality.

The judge must first decide what speech material to base the analysis upon. Ideally, the analysis should be based on both a face-to-face interview and on a tape-recorded sample of speech. As with segmental phonetic analysis, visual cues may be valuable in confirming auditory impressions, but it is possible to complete the analysis without seeing the speaker. Tape-recording, however, is essential, as it is not often feasible to attempt full VPA analysis in a live interview. Recordings should be of reasonable quality, since some features are particularily prone to distortion by common recording faults. Tape hiss, for example, may mask or mimic whisperiness, and loss of high frequency sound mimics one acoustic effect of increased nasality.

Choice of speech sample (reading, spontaneous speech etc.) will vary according to the aims of the analysis, but in all cases the sample should be of a reasonable length. It is not possible to abstract long-term average tendencies from a sample of much less than 40 seconds, although some features, such as phonation type, may be analysed from shorter samples.

The protocol is divided into three sections, concerned with vocal quality features, prosodic features and comments on breath control, continuity etc.. The first section, on vocal quality, has the soundest theoretical base, and will be the focus of this description. The same general procedure can be applied to the other sections, with some minor differences which will be covered in the last part of this manual.

On the left hand side of the form are listed the major categories within which a speaker may differ from neutral; labial, mandibular, lingual tip/blade, etc.. These are arranged in an order which corresponds approximately to an anatomical progression down the vocal tract from lips to larynx. Supralaryngeal and laryngeal features are clearly separated on the form. This allows judges to see at a glance to what extent any disorder is restricted to laryngeal output or to modification of the tract by the articulators, but it should be stressed that the complex interrelationships between all parts of the vocal tract mean that it is rare to find vocal disorders that are completely localized.

To the right of the category labels the form is divided vertically into two sections, headed 'First Pass' and 'Second Pass'. This allows a two stage process of evaluation, at different levels of subtlety. It is often relatively easy to judge that a given voice deviates from neutral in some category, but much more difficult to specify the exact nature of the deviation. The 'First Pass' section demands only the easier decision between neutral and non-neutral for each category of settings.

Under 'Second Pass' are listed the various settings which fall within each category, and the judge is here required to specify not only the precise nature of the deviation away from neutral, but also the scalar degree of that deviation. There are six non-neutral scalar degrees for most settings, of which degrees 1-3 are classed as 'normal' and 4-6 as

'abnormal'. Scalar degree 1 for any setting is the minimum deviation from neutral which can be auditorily identified. Scalar degree 6 corresponds to the maximum deviation which a normal vocal tract is capable of producing. The remaining scalar degrees are intended to reflect equal auditory steps between these extremes. This will be explained in more detail in the following section. The meaning of the labels 'normal' and 'abnormal', and the reasons for using them on the protocol, need some expansion, however. Firstly, there are some things which the labels do NOT mean. It is not true that a speaker whose vocal profile shows one or two settings within the 'abnormal' range is necessarily pathological, or even dramatically unusual. Similarily, it is not true that the vocal profile of a speaker with a grossly pathological voice will inevitably have many settings within the abnormal range. The interpretation of a vocal profile as normal or pathological will depend on an examination of the combination of settings within the whole profile, and on a knowlege of what non-neutral settings are characteristic of a given speech community. In some speech communities it is not uncommon to find one or two settings which fall in the 'abnormal' range. Many American accents, for example, are typically nasal at scalar degree 4. This underlines the point that neutral is definitely not synonymous with normality.

Having said all that, there are some points which favour the retention of the normal / abnormal distinction. Firstly, it is true to say that, for most settings, scalar degree 3 is the maximum deviation from neutral which is often characteristic of specific accents. There are exceptions to this rule, such as the case of nasality in some American accents mentioned above, but they are relatively uncommon. As result, non-clinical phoneticians, who are unfamiliar with the wide range of voice types which pass through speech therapy clinics, may be tempted to let their judgements drift towards higher scalar degrees than is appropriate. The dividing line between scalar degree 3 and 4 may help to check that tendency.

C. A PERCEPTUAL GUIDE TO VOCAL PROFILE ANALYSIS

It may be useful to preface this section with a few general hints about approaches to listening. The skills required are similar to those used in segmental phonetics, but the emphasis is somewhat different. In segmental analysis much of the emphasis is placed on isolating the features which distinguish each segment from its neighbours. In Vocal Profile analysis the task is instead to identify those features which are common to all, or to some sizeable subset, of the segments in a sample of speech. The analysis of a particular setting is often a two stage process which uses two rather different perceptual strategies. The first involves the abstraction of any long-term average biasing which underlies the rapid movements needed for segmental production. This means cultivating the ability to ignore the linguistic message, and to concentrate on the overall general impression. This strategy is most useful in the initial identification of a setting.

Confirmation of the presence of a setting, and assignment of a scalar degree often demands the analysis of classes of segments. This requires the auditory ability to isolate segments from the stream of continuous

speech, and hold them in memory long enough to analyse their perceptual characteristics. This is an appropriate point to introduce two concepts which can help in the selection of segments for individual attention.

a) Susceptibility.

A central concept of the VPAS is that individual segments differ in their susceptibility to the biasing effect of a given setting. This is most easily shown by example. Phonation type settings, such as whisperiness, will affect only those segments which are phonologically voiced. Voiceless segments will not be susceptible to phonation type settings. Similarly, a spread lip setting will have a major effect on segments which are normally rounded, such as /u/, whilst segments such as /i/, which are normally spread anyway, will be much less susceptible to its effects. When listening for the segmental consequences of a given setting it is therefore useful to know which segments are susceptible to its effects.

b) Key segments.

Following from this idea of differential susceptibility, it is often possible to identify a small set of segments on which the auditory effect of a setting is particularily prominent. These 'key' segments allow an economical listening strategy. Once the listener suspects the presence of a particular setting in a given voice, she can then test her initial impressions by concentrating only on the key segments.

The following descriptions of individual settings will include comments on susceptibility and key segments wherever appropriate. Detailed analysis of key segments is often necessary when making decisions about scalar degree.

This guide will assume that the language of the speakers whose voices are being analysed is English. The general principles of the scheme apply to all languages, but the phonological details discussed below are specific to English.

VOCAL QUALITY FEATURES

Vocal quality settings can be divided into three main types. Firstly, there is a group which can be called CONFIGURATIONAL settings. These are the settings which effect the long-term-average configuration of the vocal tract. A lip-rounded setting would be an example of this. Here, although the lips would be constantly moving to differentiate individual segments, there would be a continuous bias towards a lip-rounded position. The majority of settings fall into this category.

A second class of settings concerns the range of articulatory movement. The habitual range of lip, jaw or tongue movement may be just as characteristic of a speaker as the long-term configuration of her vocal tract, so that ARTICULATORY RANGE settings are thought to be a necessary part of vocal profile analysis.

The third pair of settings which are different again are OVERALL TENSION settings. Any change in muscle tension which is generalized throughout the vocal tract will cause a constellation of local setting changes, and

it would be possible to identify changes in tension by analysing their local consequences. It is, however, often valuable to comment on overall tension as a single underlying factor, so the form allows comment on two broad categories of tension: overall tension of the larynx and of the supralaryngeal vocal tract.

Configurational settings will be covered first, with a preliminary summary of the neutral setting. The neutral setting will also be discussed in more detail in relation to other settings.

The neutral setting

For the supralaryngeal portion of the vocal tract, the neutral setting is the one where the the vocal tract, in terms of its long-term average configuration, is as nearly as possible in the shape of a tube with equal cross-section along its whole length. To achieve this the following factors must be true.

- The lips must not be protruded, spread or rounded.

- The jaw must not be protruded, unduly open, or closed.

- Segments which are conventionally alveolar should have an alveolar place of articulation.

- The tongue body should be neither advanced nor retracted, and neither raised nor lowered.

- Audible nasality should be present only where it is phonologically required.

- There should be no constriction of the pharynx.

- The larynx must be neither raised nor lowered.

It is also possible to specify a neutral phonation type. This is what Hollein has called 'modal voice', and it involves the following features.

- Only the true vocal folds are involved in phonation.

- Vibration must be regularly periodic.

- Vibration must be efficient in air use, with full glottal adduction and without audible friction.

It is possible to specify the acoustic correlates of neutral voice quality, and of the vocal quality settings outlined below, but since this is meant as a perceptual guide, they will be omitted.

Scalar degree conventions

Detailed guidelines about scalar degree conventions will be given wherever possible as they relate to individual settings, but it may be useful to offer some general hints here.

- Scalar degree 1 should be used where the presence of a setting is just noticeable.
- Scalar degree 2 suggests that the judge is fairly confident that the setting is present, but at no more than moderate strength.
- Scalar degree 3 is the strongest setting which could reasonably be exlected to act as a regional or sociolinguistic marker for some hypothetical community. There are rare exceptions to this rule (see section B).
- Scalar degree 4 should be used if the judge has no doubts at all about the presence of a setting, and feels that it is beyond the limits of the normal population.
- Scalar degree 5 represents almost the maximum strength.
- Scalar degree 6 is reserved for the auditory effect which corresponds to the most extreme adjustment of which the normal, non-pathological vocal tract is capable.

It should be stressed again that the relationship between ideas of 'normality' and the boundary between scalar degrees 3 and 4 should be treated with extreme caution. Since any given accent is likely to be characterized by the common occurence of some settings at the scalar degree 2 or 3 level, it follows that judges who are familiar with that accent may not feel slightly increased presence, at scalar degree 4, to be abnormal. On the other hand, the presence of settings which are uncommon in that speech community may seem abnormal at lower scalar degrees.

Intermittent presence of settings

Another useful scoring convention which should be mentioned here is the use of the letter 'i' to indicate the intermittent presence of a setting. Many speakers are characterized by the regular, but intermittent, adoption of a setting. In these cases 'i' can be used to indicate the appropriate scalar degree of the setting. The scalar degree used should reflect the strength of the setting when it is present, rather than the frequency of occurence. As a general rule, 'i' is used whenever a setting is heard on less than 90% but more than 10% of susceptible segments. Where the judge feels that it is neccessary to indicate the proportion of susceptible segments which are affected by an intermittent setting, a percentage may be written alongside the scalar degree judgement. This is useful in monitoring progress of some dysphonic patients, for example, where the aim of therapy is to reduce the incidence of intermittent harshness associated with peaks of laryngeal tension.

SUPRALARYNGEAL FEATURES - Configurational settings

1. Labial Features

In judging labial settings from tape-recorded voices, try to visualize the 'set' of the speaker's face, copy it, and then imitate a few phrases for auditory comparison. It may be helpful to use a mirror to check details of your own production while doing this. Many people find this kind of non-analytical approach surprisingly accurate, and these first impressions can then be checked using the information below.

Neutral

The neutral setting for the labial category is where the long-term average lip posture is as it would be for [ə], i.e. the lips are neither spread, nor rounded, nor protruded.

Lip rounding/protrusion

Lip rounding and protrusion are physiologically separable, but since lip-rounding most commonly occurs with a comparible degree of protrusion, and vice versa, they have been collapsed into a single setting.

Key segments

- front oral fricatives [s] and [θ] have a lower apparent 'pitch' in a lip rounded/protruded setting.
- /i/ and other vowels which are conventionally spread or unrounded will tend to become more rounded. If you can isolate the actual phonetic realization of /i/ in a word like 'heed', for example, a speaker with a lip-rounded setting will often use a rather rounded vowel, saying [hyd] rather than [hid].
- /r/, /ʃ/, /t͡ʃ/ and /d͡ʒ/, where lip-rounding is optional in English, will tend to be produced with lip-rounding.

Scalar degrees

Scalar degrees 1-3 are used for long-term average (LTA) lip positions of open rounding, and scalar degrees 4-6 are used for close rounding. Scalar degree 3 is where the LTA position is equivalent to that used for cardinal vowel 6 [ɔ]. In scalar degrees 4-6 the lip aperture becomes prgressively smaller, until scalar degree 6 has a LTA lip position comparible to cardinal vowel 8 [u].

Lip-spreading

Key segments

-front oral fricatives [s] and [θ] have a higher apparent 'pitch' in lip-spreading.
-/r/, /ʃ/, /t͡ʃ/, and /d͡ʒ/ tend to be pronounced without lip-rounding. This is most easily heard in the transitions to and from these segments.
-/w/ and vowels which are normally rounded, such as /u/ and /ɔ/, will tend to lose their rounding. Again, it is useful to concentrate on the exact phonetic realization of words with these vowels. The word 'two', for example, may be produced as [tɯ] rather than [tu].

Scalar degrees
Scalar degree 4 is used to mark the point where the LTA lip position is as spread as it would be for cardinal vowel 2 [e]. Scalar degree 6 corresponds to the position for an overspread [i].

Lip-rounding/protrusion and lip-spreading can be thought of as diametrically opposed deviations from neutral. Together they form a 13-point scale with neutral forming the central point. Although lip-protrusion affects the length of the vocal tract, the focus of attention is on the cross section of the labial opening. The next setting is rather different.

Labiodentalization

This setting is produced by bringing the lower lip closer to the upper teeth, thus shortening the vocal tract. Labiodentalization may co-exist with lip-rounding/protrusion or lip-spreading. Many people produce some degree of labiodentalization with the kind of short-term lip-spreading that results from talking whilst smiling or laughing.

Key segments
-bilabial stops /p/, /b/, and /m/ are most susceptible to labiodentalization. There may be audible labiodentalization at onset and offset of these segments, or they may actually be produced as labiodental stops.
-front oral fricatives, especially [s], may have a lower 'pitch'. This is a rather variable feature, however, because of the possible interaction with lip-rounding or spreading.
-/r/, /w/ and /u/ often have audible labiodentalization.

Scalar degrees
Scalar degrees 1-3 add an audible labiodental factor to onset and offset of /p/, /b/ and /m/. In scalar degrees 4-6 there is a progressive increase in the realization of these segments as labiodental stops, so that by scalar degree 6 they are all produced as fully labiodental segments.

2. Mandibular features.

As with labial features, a useful first step in the analysis of jaw settings is often to try to visualize the speaker's face, and to imitate the 'set' of the jaw.

Neutral

In the neutral configuration, there is a very small vertical gap between the upper and lower incisors for most speakers. In the horizontal plane, the lower incisors lie just inside the upper ones.

Open and Close Jaw

The LTA position of the jaw may be more open or more more close than the specified neutral position. Unlike lip spreading and rounding, neutral does not form the midpoint of a 13-point scale. The physical and

auditory distance between neutral and scalar degree 6 open jaw is much greater than the distance between neutral and a maximally close jaw. For this reason, only three scalar degrees are used for close jaw settings. These correspond to a collapsing of scalar degrees 1 and 2, scalar degrees 3 and 4, and scalar degrees 5 and 6.

Key segments

The degree of jaw opening used by a speaker may have rather general effects, since in the absence of compensatory adjustments it will have inevitable consequences for labial opening and for the carriage of the tongue relative to the roof of the mouth. The amount of 'travel' heard during the articulation of front consonants and close front vowels is often a useful clue.

Scalar degrees

Scalar degree 1/2 of close jaw corresponds to a position in which there is no longer any vertical gap between the upper and lower incisors. Scalar degree 5/6 corresponds to totally clenched teeth. For open jaw, scalar degree 4 corresponds to the jaw position which just allows the upper surface of the tongue to be clearly visible. Scalar degree 6 is the maximum possible opening achievable with a normal anatomy.

Protruded Jaw

Protrusion of the jaw is associated with a change in the horizontal relationship between the upper and lower incisors, and between the tongue and the roof of the mouth.

Key segments

- /s/ and /ʃ/ have a 'darker', low-pitched quality, which becomes obvious at scalar degrees of 4 or more.
- Since the protruded jaw carries the tongue forward relative to the upper teeth and the palate, all lingual articulations will tend to be fronted unless compensatory adjustments of the tongue are made. Where compensatory adjustments are made, a slightly retroflex quality is often heard on front oral consonants.

Scalar Degrees

In scalar degree 4 the lower incisors are held just in front of the upper incisors. In scalar degree 6, the lower teeth are level with the upper lip, as long the lip itself is not protruded.

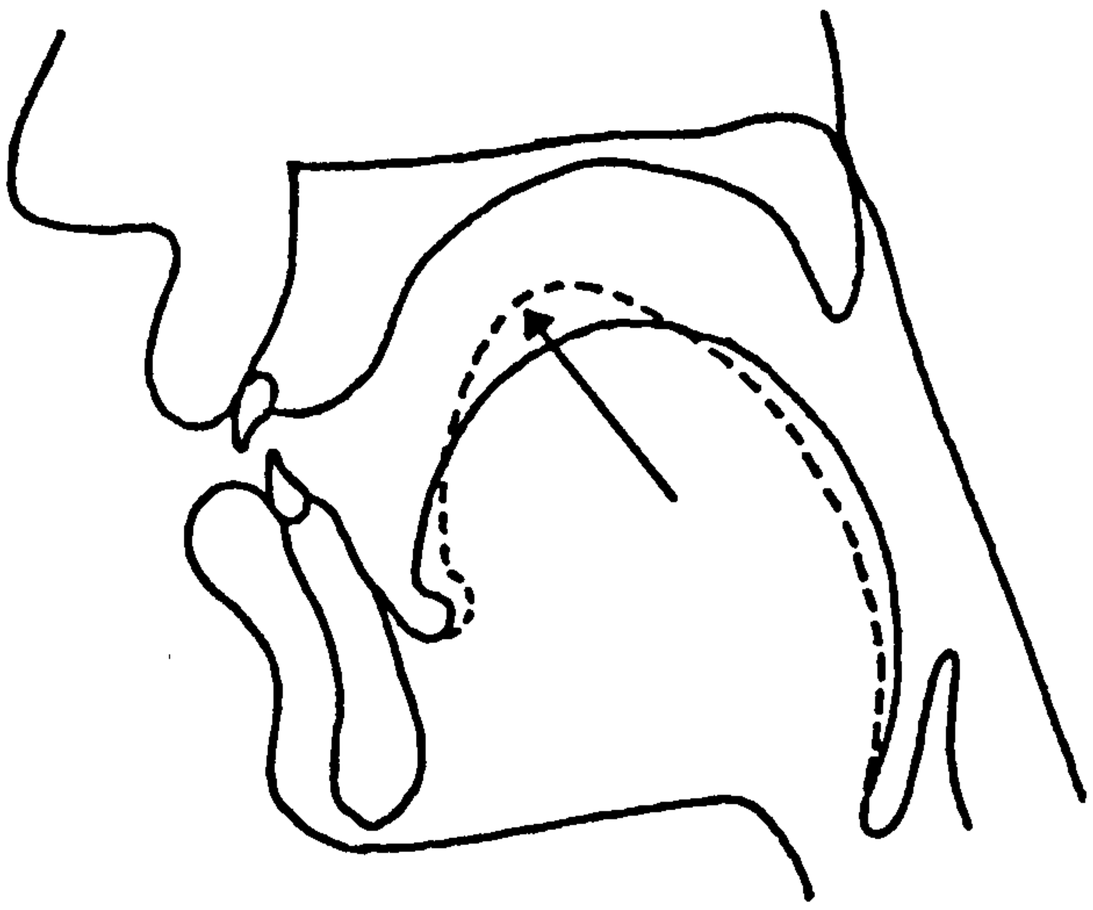
3. Lingual Tip/blade settings

The first category of lingual settings is specifically concerned with the actual place of articulation of the set of segments which are conventionally described as 'alveolars', i.e. /t, d, s, z, n, l/.

Neutral

In a neutral tip/blade setting all the so-called 'alveolar' segments are produced with a truly alveolar place of articulation. The active articulator may be either the tip or the blade of the tongue.

A.



B.

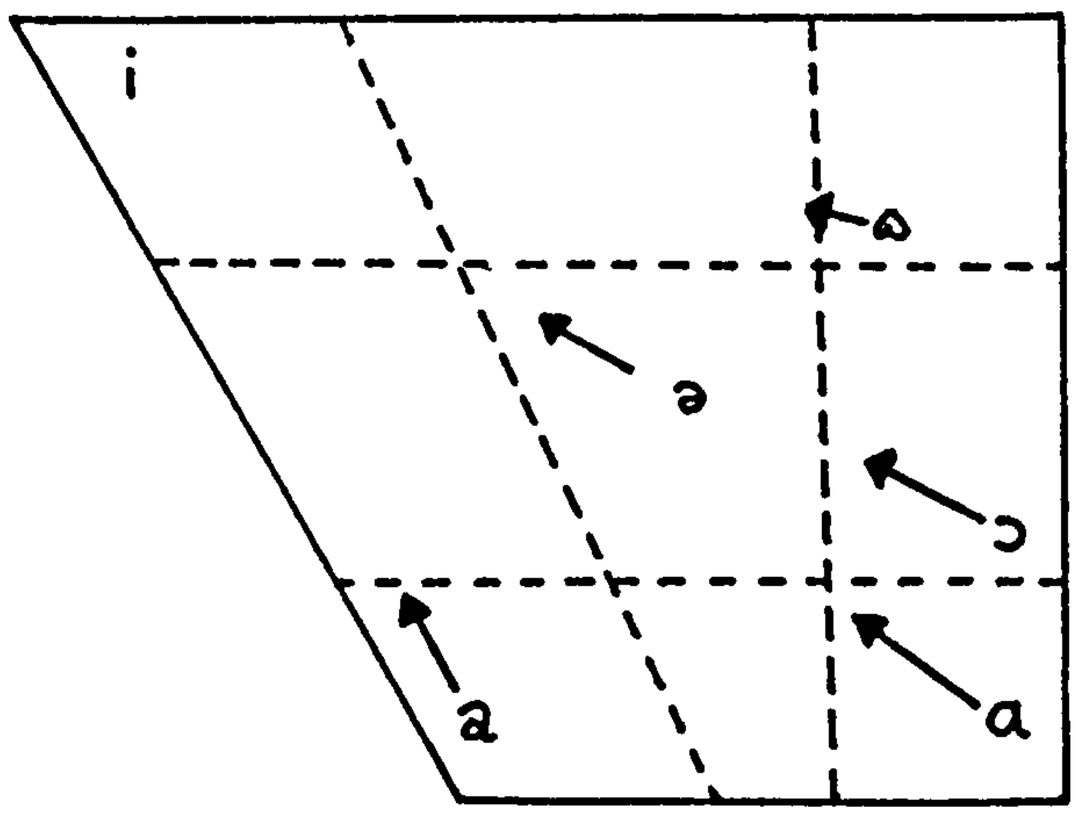


FIGURE 2: Diagram of changes in A. vocal tract configuration and B. vowel distribution in neutral (solid line) and fronted and raised tongue body setting (broken line)

Advanced and retracted tip/blade

It is obviously possible to produce the above set of 'alveolar' segments with a place of articulation which is either in front of the alveolar ridge (advanced) or behind the alveolar ridge (retracted). It is usual in speakers of English for retraction to be associated with increasing degrees of retroflection, so that extreme degrees of retraction involve retroflex articulation of the so-called alveolar segments.

Key segments

All the susceptible segments, i.e. /t, d, s, z, n, l/, should be used as key segments. The effect of advanced or retracted tongue tip/blade is often most prominent on /s/, but the judge must check that any deviation from the alveolar position in /s/ production is generalized throughout the whole set of segments. It is not uncommon for an accent, or an individual, to be characterized by non-alveolar pronounciation of only one of the set, often /s/. In this case it is more appropriate to view this deviation from neutral as a segmental characteristic than as a vocal quality characteristic.

Scalar degrees

For an advanced tongue setting, scalar degree 1 is the point where the tongue tip or blade begins to make contact with the back surface of the teeth as well as with the front of the alveolar ridge. Scalar degree 4 corresponds to fully dental articulation, with no alveolar contact. Scalar degree 6, being the maximum possible for normal speakers, corresponds to extreme interdentalization.

In retracted settings, the place of articulation moves progressively back, so that scalar degree 3 involves a post-alveolar place of articulation. In scalar degree 4 the tongue tip is begining to move towards a retroflex position, with the tongue tip pointing directly up just behind the post-alveolar place of articulation. In this degree of retraction /s/ may have a very distinctive 'whistling' quality. Scalar degree 6 has the underside of the tongue tip making contact with the roof of the mouth in fully retroflex articulation.

4. Lingual body settings

The second category of lingual settings is concerned with the LTA position of the central mass of the tongue. From the neutral position, the tongue body may move up or down, and backwards or forwards. Several listening strategies may be useful. The first is to try to abstract a LTA vowel quality from the continuous stream of speech. If this can be done, it follows that the LTA tongue position must correspond to the position needed to produce the abstracted vowel. A second technique is to concentrate on specific vowel segments, and to judge where they fall in a traditional vowel area diagram. In a neutral setting the vowels will be evenly distributed around the centre of the vowel area, but in non-neutral settings the distribution will be skewed away from the centre (See Figure 2). A third approach is to concentrate on secondary articulation of consonants such as /l/ and /w/. On the protocol form there are two pairs of diametrically opposed setting scales; fronted/backed and raised/lowered, but in practice tongue body settings

are often combinations of these, such as fronted + raised, or backed + lowered.

Neutral

In neutral, the LTA position of the tongue body is the position used to produce the vowel /ə/ (see Figure 2).

Fronted/backed tongue body

Key segments
- Vowels are the segments most susceptible to change by tongue body settings. In fronted tongue body, back vowels will be most affected, becoming progressively more fronted, so that in extreme degrees of fronted tongue body there will be no vowels in the right hand half of the vowel area. Tongue backing, in contrast, affects front vowels most, pushing all vowels backwards, towards the right of the vowel area.
- /l/ and /w/ may vary in terms of secondary articulation. Palatalization is likely to be more extreme in speakers with fronted tongue body, whilst velarization or pharyngealization are likely to be more marked in speakers with backed tongue body.

Scalar degrees
Assignment of scalar degree depends on a judgement of how far the vowel area is limited to left or right (front or back). Scalar degree 4 of fronted tongue body brings the furthest back vowels forward to a central position. /u/, for example, would tend to be realized as a close central vowel. In a backed tongue body setting, scalar degree 4 shifts all vowels back, so that the 'frontest' vowels are in the centre of the vowel area. /i/ would in this case be realized as a close central vowel.

Raised/lowered tongue body

The principles of judging these settings are the same as for fronted and backed tongue body. Raised tongue body makes all vowels closer, and lowered tongue body makes all vowels opener. Tongue body lowering will also affect semi-vowels /j/ and /w/, so that they may be realized as half-close variants.

Scalar degrees
Scalar degree 4 of raised tongue body will bring the most open vowels up to a borderline position between half-close and half-open. Scalar degree 4 of lowered tongue body will bring the closest vowels down to a similar position. In scalar degree 4 and beyond, /j/ and /w/ will become half-close.

5. Velopharyngeal features

Velopharyngeal settings pose some of the most complex problems for phonetics. This scheme forces a decision between nasal and denasal resonance, but we recognise that this two way distinction may not always allow an adequate description of velopharyngeal features.

Neutral

The neutral velopharyngeal setting is where audible nasality is present only where it is necessary to maintain phonological identity. For English that means that only /m/, /n/ and /ŋ/ will have audible nasality, and anticipatory nasality will be cut to the minimum which is physiologically necessary. In practice, neutral is virtually never heard in English.

Nasal

Key segments

- Vowels and continuant consonants may be heard to have nasal resonance. Nasality is heard most easily on open vowels, but close vowels and eventually some consonants (e.g. voiced fricatives) will have audible nasality at higher scalar degrees.

Scalar degrees

Up to scalar degree 2 nasality will be easily heard only on open vowels. At scalar degree 3 some closer vowels will show audible nasality. By scalar degree 4 all vowels will have clearly audible nasality. Nasality begins to affect consonants at scalar degree 5, increasing at scalar degree 6 so that nasality will be clearly heard on voiced fricatives, for example.

Denasal

Key segments

- /m/, /n/ and /ŋ/ progressively lose nasal resonance.
- vowels have a 'cold-in-the-head' quality.

Scalar degrees

In scalar degrees 1-3 the most prominent feature is the 'cold-in-the-head' effect on some vowels. In scalar degree 4 the so-called nasal stops will be clearly losing nasality. At scalar degree 6, they will have lost all nasality. The distinction between /m/, /n/ and /ŋ/ and their oral counterparts will be maintained only by having different amounts of voicing, so that severe problems of intelligibility may arise.

Audible nasal escape

Audible nasal escape is audible, fricative airflow from the nose. Since it is considered to be abnormal in all accents of English, the protocol shows only scalar degrees 4-6. Audible nasal escape will tend to occur first on segments which require the maintenance of high oral air pressure, e.g. /s/, /ʃ/. At scalar degree 4 only these segments will have fricative nasal airflow, whilst at grade 6 it will be present on virtually all segments. It should be stressed that whilst audible nasal escape occurs most commonly with high degrees of nasal resonance, this is not an invariable association. In rare instances it may even occur with a denasal setting.

6. Pharyngeal Constriction

This setting is used to describe constriction of the pharynx which results not from retraction of the body of the tongue into the pharynx, but from sphincteric contraction of the pharyngeal constrictor muscles. It lends a 'strangulated' quality to the voice, so that at high scalar degrees the empathetic listener is aware of considerable discomfort and obstruction of the pharynx.

SUPRALARYNGEAL SETTINGS - Articulatory range settings

Articulatory range settings specify the maximum span of movement which lips, jaw and tongue cover during speech. This should not be confused with rate of articulatory movement, although there is an obvious interaction between the two. It is, however possible to have a wide overall range of, say, jaw movement, but for the rate of jaw movement to nevertheless be rather slow.

Key segments
- Dipthongs: these will show a long travel from from start to end point in extensive range settings, and very little or no travel in minimised range settings.

Scalar degrees
For range of lips, jaw and tongue, the end points of the scales are easily defined. Scalar degree 6 of extensive range means that the articulator must reach the most extreme positions of which it is capable, in all directions. Scalar degree 6 of minimised range means that the articulator is totally immobile. Neutral refers to the range of movement which will just maintain clear intelligibility without the need for some other articulator to compensate.

SUPRALARYNGEAL FEATURES: Overall Tension Settings

Alterations in overall tension of the vocal tract tend to cause constellations of changes in configurational and range settings. Judgement of overall tension is therefore based largely on a knowlege of these constellations, which are outlined below. Problems may, however, arise in cases where physiological anomalies mean that a change in tension is not associated with the usual changes in other settings. In these cases, the listener may have to rely on an empathetic judgement of muscular tension.

Lax

Generalised laxness is often associated with the following local changes:
- Open jaw setting
- Nasal setting
- Minimised ranges of lip, jaw and tongue.

In addition, acoustic clues to laxness, which presumably contribute to its auditory characteristics, include damping of high frequency noise, and broad formant peaks.

Tense

Generalised tension is associated with a different set of local changes:
- Reduced degrees of nasality
- Extensive ranges of lips, jaw and tongue
- Pharyngeal constriction.

Acoustically, there is less absorption of high frequencies by the vocal tract walls, and formant peaks are sharper.

LARYNGEAL FEATURES: Configurational settings

9. Larynx Position

The potential range of larynx positions is quite wide, as evidenced by the displacement of the larynx which occurs during swallowing. The complex of muscles from which the larynx is slung means that alterations in larynx position may be accompanied by a wide range of other changes, and this sometimes makes it difficult to isolate the auditory effect of larynx position settings. The judge needs to concentrate on the auditory effects of lengthening or shortening the vocal tract, and try to dissociate these from features such as changes in pitch or pharyngeal constriction, which often, but not always, accompany changes in larynx position.

Neutral

Neutral corresponds to the auditory quality associated with a larynx position approximately in the mid-point of its potential range.

Raised and Lowered Larynx

The effects of larynx position settings are most clearly audible on vowels, as a result of changes in formant ratios associated with vocal tract length. It is not possible to give specific guidelines for scalar degrees, so the general conventions should be followed.

10. Phonation type settings

Neutral

Neutral phonation is very rarely heard in normal continuous speech, but it has very clear acoustic and physiological correlates. Neutral phonation, or to give it its alternative label, modal voice, involves very regular and efficient vocal fold vibration. Only the true vocal folds are involved in phonation, and the pattern of vibration is perfectly regular; each cycle of vibration has the same duration and magnitude as its neighbours. Acoustically, it is possible to see this regularity in terms of pitch (fundamental frequency) and intensity.

Phonation may deviate from neutral either by the addition of audible turbulence of the airflow, or by an alteration in the pattern of vocal fold vibration. When modal voice occurs in combination with other phonation types in non-neutral phonation, the term 'voice' is used to describe this component.

Scalar degree conventions in non-neutral phonation

Modal voice is marked simply as being present, intermittently present or absent on the protocol form. Where it occurs as a component of complex phonation types, the auditory balance between it and other components is indicated by the scalar degrees assigned to the other components. Where any phonation type is combined with voice, scalar degrees 1-3 are used where the voice component is perceptually more prominent and scalar degrees 4-6 are used if the other phonation type is perceptually more prominent. A similar rule applies when falsetto is combined with other phonation types (see below).

Falsetto

Falsetto cannot occur at the same time as modal voice, although it can be combined with all other phonation types. Like modal voice, it is marked only as present, intermittently present or absent.

Harshness

Harshness is a disturbance of the vibratory pattern associated with either voice or falsetto, and can therefore only occur in combination with one or other of these basic phonation types.

Whisper or whisperiness

The whisper(y) setting is used whenever there is audible friction due to turbulent airflow through the glottis. Whisper can occur alone, or in combination with any other setting.

Creak or creakiness

The creak setting is reserved for voices in which discrete pulses can be perceived in the phonation. Like whisper, it can occur alone, or combined with other settings.

LARYNGEAL FEATURES: Overall Tension Settings

The same general comments apply as for supralaryngeal tension settings. Lax settings often result in lowered larynx, low pitch, and moderate degrees of whisperiness. Tense settings tend to be more often associated with raised larynx, high pitch, and harshness.

PROSODIC FEATURES

It is harder to offer objective guidelines for the judgement of prosodic features. Pitch is taken to be the perceptual correlate of fundamental frequency, but the perception of pitch is complex, and seems to relate also to spectral acoustic features. In addition, expectations are affected by the sex, age and physique of the speaker in a way which is not always easy to quantify. Loudness is the perceptual correlate of acoustic intensity, but is very hard to judge from tape-recorded material. It is therefore impossible to give clear definitions of neutral for pitch and loudness settings. It is, however, possible to give general definitions for the prosodic features, and these are

summarised below. For most voices these seem to allow a reasonable level of agreement between judges, but the VPAS cannot pretend to be properly objective in this area. Various sorts of acoustic instrumentation are available which can give objective measures of fundamental frequency and intensity, and it is recommended that these should be used wherever possible.

Pitch

Pitch Mean: this refers to the average perceived pitch for the whole speech sample. It may be judged to be neutral, high or low.

Pitch Range: this is a comment on the span between the highest and the lowest pitch used by the speaker. It may be judged to be neutral, wide or narrow.

Pitch Variability: this refers to the frequency with which a speaker moves around within his or her pitch range.

Consistency

This relates to consistency and coordination of respiratory and phonatory processes. When these break down, the audible result is often tremor. Tremor can be defined as the occurence of audible fluctuations in pitch and/or loudness, which typically occur at a rate of 1-3 per syllable.

Loudness

The definitions of loudness settings are exactly parallel to those for pitch settings, i.e. loudness mean refers to the long-term average loudness, loudness range refers to the span between greatest and least loudness, and loudness variability refers to the amount of movement within that loudness range.

TEMPORAL ORGANIZATION FEATURES

This section is similar to the previous one, in that it is difficult to specify a neutral baseline, so judges should use this simply to make comments about the adequacy or otherwise of a speaker's continuity and rate.

Continuity in this context concerns the incidence of pauses within a speech sample. Marking a speaker as having an interrupted setting implies the presence of inappropriate silent pauses between words or syllables.

Rate is used to describe the actual speed of utterance at the segment or syllable level. This need not neccessarily equate with a measure of words or syllables per minute, since a low number of words per minute could be due to a high incidence of pauses rather than a slow rate of syllable production.

It should be clear that these categories of the VPAS are inadequate to allow full description of speakers, such as stammerers or dysarthrics, where disrupted temporal organization is a major feature. They do, however, act as place holders, signalling the need for further specialized investigation.

COMMENTS

The VPA protocol also allows comments on breath support, rhythmicality and diplophonia. Breath support may be marked as adequate or inadequate for normal speech production. Rhythmicality is similarily scored as adequate or inadequate, although this may seem a slightly odd concept. The acceptability of the rhythm used by a speaker will obviously depend both on linguistic content, and on language or accent. Syllable timing, for example, would be appropriate, and therefore adequate, in French, but be undoubtedly inappropriate in most British speech communities.

Diplophonia is obviously closely related to phonation type, but until there is clearer agreement about the physiological and acoustic bases for diplophonia it cannot properly be placed within a phonetic theory. The perceptual definition for diplophonia used here is that two fundamental pitches should be audible simultaneously. This excludes some voices which are often described as diplophonic, where there is rapid fluctuation of pitch, often associated with an alternation between modal voice and falsetto. Diplophonia is scored simply as being present, intermittently present (by the use of the 'i' convention), or absent.

APPENDIX TWO

TEXT CUT OFF IN ORIGINAL

A PERCEPTUAL PROTOCOL FOR THE ANALYSIS OF VOCAL PROFILES

John Laver : Sheila Wirz
Janet Mackenzie : Steven Hiller

INTRODUCTION

A vocal profile will be taken here to consist of a statement of the speaker-characterising, long-term features of a person's overall vocal performance. It includes comment on laryngeal and supralaryngeal aspects of voice quality, on means, ranges and variability of prosodic aspects such as pitch and loudness, and on factors of temporal organization such as rate and continuity. In lay terms, a vocal profile summarizes the phonetic features of a speaker's habitual 'voice'.

It is reasonable to describe a statement of these features as a 'profile', rather than merely as a listing, to the extent that a theoretical relationship exists between the items within the profile. A descriptive model, set in the framework of general phonetic theory, has recently been put forward which analyses a speaker's voice as the product of perceptually distinguishable components, each specified in terms of acoustic, articulatory and physiological correlates (Laver 1968, 1974, 1975, 1979, 1980; Laver and Hanson 1981; Laver and Trudgill 1979). The basic unit of this scheme is an auditory component correlated with an articulatory 'setting' (Honikman 1964), which is a long-term muscular bias on articulation. Examples are habitual tendencies to lip-rounding, to nasality, or to a whispery mode of phonation. Each such setting/component is defined in terms of its deviation from an acoustically and articulatorily specified neutral reference configuration of the vocal apparatus (Laver 1980:14; Laver and Hanson 1981:59). There are certain constraints of mutual compatibility between individual setting/components, on both physiological and acoustic grounds, and it is this necessary theoretical relationship between the elements of the descriptive model that justifies the use of the term 'profile'. The perceptual product of constellations of such setting/components in the speech of a given speaker makes up his vocal profile.

The analytic model in the references listed above is (largely) confined to a description of phonetic aspects of the normal voice. A three-year research project[1] which started in 1979, employing two speech therapists, two speech scientists and a clerical assistant, is currently extending the descriptive technique beyond normal voices to include abnormal voices found in speech disorders. A major objective of the project is to test the medical applicability of the descriptive system, and to make it available as a standard descriptive tool in speech therapy and speech pathology clinics.

A working hypothesis of the project is that particular speech disorders have characteristic vocal profiles associated with them. To create a data-base on which this hypothesis can be tested, tape recordings are being made of not less than 30 young male and female adult speakers, mostly from Scottish hospitals, day-centres and clinics[2], for each of eight types of disorder. A comparable control group is also being recorded. The task of making the recordings is nearly complete, and to date some 200 recordings have been made.

Vocal Profile Analysis Protocol

Judge: ..... Tape: ..... Sex: .....

Date of Analysis: ..... Speaker: ..... Age: .....

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial				Lip Rounding/Protrusion						
				Lip Spreading						
				Labiodentalization						
				Extensive Range						
				Minimised Range						
2. Mandibular				Close Jaw						
				Open Jaw						
				Protruded Jaw						
				Extensive Range						
				Minimised Range						
3. Lingual Tip/Blade				Advanced						
				Retracted						
4. Lingual Body				Fronted Body						
				Backed Body						
				Raised Body						
				Lowered Body						
				Extensive Range						
				Minimised Range						
5. Velopharyngeal				Nasal						
				Audible Nasal Escape						
				Denasal						
6. Pharyngeal				Pharyngeal Constriction						
7. Supralaryngeal Tension				Tense						
				Lax						
B. Laryngeal Features										
8. Laryngeal Tension				Tense						
				Lax						
9. Larynx Position				Raised						
				Lowered						
10. Phonation Type				Harshness						
				Whisper(y)						
				Breathiness						
				Creak(y)						
				Falsetto						
				Modal Voice						

II PROSODIC FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
1. Pitch				High Mean						
				Low Mean						
				Wide Range						
				Narrow Range						
				High Variability						
				Low Variability						
2. Consistency				Tremor						
3. Loudness				High Mean						
				Low Mean						
				Wide Range						
				Narrow Range						
				High Variability						
				Low Variability						

III TEMPORAL ORGANIZATION FEATURES

CATEGORY	FIRST PASS		SECOND PASS			
	Adequate	Inadequate	Scalar Degrees			
			Inadequate			
			1	2	3	
1. Continuity						Interrupted
2. Rate						Fast
						Slow

IV COMMENTS

	Adequate	Inadequate	1	2	3
Breath Support					
Rhythmicality					

	Present	Absent
Diplophonia		

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh. © June 1981.

FIGURE 1

The eight types of disorder, some containing sub-types, are:

1. Profound hearing loss[3]
2. Cerebral palsy[4]
3. Down's Syndrome[5]
4. Sex-chromosome anomalies[6]
5. Parkinson's Disease[7]
6. Thyroid disorders[8]
7. Dysphonia[9]
8. Cleft palate[10]

It will be noted that the above list contains abnormalities both of anatomy and of neurological control. An important conclusion of the research will concern the extent to which such abnormalities constrain any attempt to unify the description of normal and pathological vocal performance.

The primary analytic task of the project has been to construct perceptual and acoustic profiles of the voice of each subject in the above groups. The purpose of this progress report is to give an account of the work to date on the development of a protocol form for the analysis of the perceptual profiles of the subjects. An account of the construction of the corresponding acoustic profiles, derived from computer analysis of the recordings, largely using LPC signal-processing techniques, will be published elsewhere.

DEVELOPMENT OF THE VOCAL PROTOCOL

Given the objective of clinical applicability, it was decided from the outset of the project that the protocol form would be designed in collaboration with experienced speech therapists, during sessions of training in the use of the descriptive system. The present version, shown in Figure 1, is the tenth generation of the protocol. The content and rationale of the protocol have been the product of collaborative experience with four successive training panels, two from Lothian Region, one from Glasgow, and one from Newcastle, comprising over 50 individual therapists. Two further panels have been arranged for the immediate future, one in Nottingham and one in London, and further minor evolution of the protocol is not ruled out, though development seems now to have reached a relatively stable plateau.

USING THE VOCAL PROFILE ANALYSIS PROTOCOL

It may be helpful from the outset to distinguish sharply between the terms 'profile' and 'protocol'. The VPA protocol is the form shown in Figure 1; the profile is represented by the data entered on the protocol. A therapist uses a protocol to record a patient's vocal profile; changes in a patient's vocal profile during a course of remedial therapy can be quantified by noting changes in the data entered on the corresponding VPA protocols; and the completed protocol constitutes a permanent, written record that can be stored in a patient's case-notes and interpreted by anyone trained in the descriptive system (and readily explained to medical personnel not trained in the system).

The ultimate goal of the project is to link the perceptual and the acoustic analysis approaches. However, important though it is for automatic acoustic analysis to be made available for clinical use, the major value of the scheme for clinicians is likely to lie initially in the perceptual technique. The immediate accessibility of perceptual judgments allows a therapist to make direct assessments of vocal factors independently of complex, expensive and often physically remote technology; and provided that the perceptual system has a demonstrable correlation with objective acoustic measurement, then the main function of the acoustic technology will for some time to come, until powerful computers become standard equipment in speech clinics, be a back-up, confirmatory function.

1. The Speech Sample

Normally, the perceptual analysis is performed on tape recordings. Ideally, this should be supplemented by visual observation of the patient. This is not essential, but as in segmental analysis, visual clues may be valuable in confirming auditory judgments. Labial and mandibular settings are obviously associated with visible factors, but this is also true, to a smaller extent, of lingual and larynx position settings.

Good quality audio recording is advisable for the accurate analysis of vocal features, as some setting components are extremely prone to distortion by poor recording. Attenuation of high frequency energy, for example, mimics the acoustic damping correlated with nasality, and so tends to bias perceptual judgments towards higher ratings on the nasal setting. Tape hiss interferes with the assessment of the fricative qualities attributable to whisperiness, breathiness, or audible nasal escape.

The speech sample must be long enough to allow long-term-average setting effects to be perceptually abstracted from the shorter-term segmental performance. The time needed for accurate assessment varies from setting to setting, and depends in part on the proportion of segments which are susceptible to the influence of each setting. Phonation type, audible in all phonetically voiced segments, can be judged over samples of only a few syllables, but settings which exert their influence on a more limited number of susceptible segments, such as advanced or retracted articulation of the tip/blade of the tongue, will require much longer samples. Laver and Hanson (1981:53) review evidence suggesting that 45-70 seconds of speech is necessary for the automatic abstraction of long-term features by computer, but human judges may need a sample of a longer duration.

2. Completing the VPA Protocol

The protocol shown in Figure 1 is made up of four sections: vocal quality features, prosodic features, temporal organization features and comments. The procedure to be followed when completing the VPA protocol will be described as it applies to the vocal quality section, as a model for the other sections.

On the left hand side of the vocal quality section are listed the major categories within which adjustments away from neutral may occur: labial, mandibular, lingual tip/blade, lingual body,

velopharyngeal, pharyngeal, supralaryngeal tension, laryngeal tension, larynx position, and phonation type.

Supralaryngeal and laryngeal features are separated on the form, but this is to some extent a pragmatic division. The interdependence of supralaryngeal and laryngeal settings is very close, both at the level of perceptual analysis, and at the level of the underlying muscular systems. Laryngeal settings may mask or enhance the perception of supralaryngeal settings quite markedly, and vice versa. Velopharyngeal factors, for example, seem to be prone to masking by the presence of whisper as a component of phonation type. At the muscular level, the interactivity of muscle groups responsible for the production of different categories of setting leads to the common co-occurrence of particular constellations of laryngeal and supralaryngeal settings. Raised larynx and pharyngeal constriction are good examples of this, showing a closely overlapping distribution. There is, however a traditional tendency to treat laryngeal output rather differently from articulatory modifications of the supralaryngeal vocal tract. Laryngeal output (and often nasality also) has generally been considered as a long term feature, whilst supralaryngeal adjustments have more commonly been analysed at a shorter term, segmental level. The division also accommodates itself readily to a source-filter type of acoustic model.

A major factor in preferring to distinguish laryngeal from supralaryngeal settings is the fact that in some clinical populations, such as speakers with dysphonia, or certain groups of speakers with articulation disorders, there is an obvious tendency for severe deviations from neutral to occur solely in either the laryngeal or the supralaryngeal section. There are, of course, many other types of speech disorder where severe deviations occur throughout the vocal apparatus.

The balance of these arguments has been to favour the separation of laryngeal and supralaryngeal factors on the protocol, but the close relationship between them is important enough that its implications should be stressed.

The layout of the protocol allows a two-stage process of evaluation of different levels of decision-taking. There is a vertical division into two sections headed 'First Pass' and 'Second Pass'.

(i) First Pass: On the first pass, which might correspond to the first listening to the speech sample, the judge is required to make only a rather broad decision regarding each category of setting, by marking each as neutral or non-neutral. If the voice is thought to be non-neutral with respect to any category, the judge can then decide whether it falls within the normal or the abnormal range.

The inclusion of a 'First Pass' is a response to the experience that it is often a much easier perceptual task to judge that a given voice deviates from neutral within some category than it is to specify the exact direction of that deviation. It seems to be true, for example, that people learning the scheme find it relatively easy to discern an adjustment of larynx position away from neutral, but find it considerably more difficult to differentiate between the qualities associated with raising and lowering of the

larynx. This is in spite of the clearly differentiable acoustic correlates of raised and lowered larynx. The first pass, then, allows the judge to comment on a deviation away from neutral without specifying the polarity of the deviation, and it also leaves the judge free to ignore all neutral categories when making a second pass through the material.

It deserves emphasis that, in nearly all circumstances, it is important to fill in the whole protocol, even when interest might be thought to focus on sub-sections of the form. It has been a repeated experience of the therapists in the project that the settings relevant to decisions about treatment have been grouped in constellations, rather than as single settings.

(ii) Second Pass: Under 'Second Pass' are listed all the settings within each category and the judge is here required to specify not only the precise direction of any deviation away from neutral, but also the scalar degree of deviation.

There are six scalar degrees for each setting[11], with three exceptions. Falsetto and modal voice are scored simply for presence or absence, and audible nasal escape has only three scalar degrees, in the abnormal range. For all other settings, the scale from 1 to 3 is considered to be normal, and the scale from 4 to 6 is considered to be abnormal.

Taking the lingual body category as an illustration, a judge who had decided on the First Pass that there was a normal but non-neutral setting of both fronting-backing and raising-lowering components of tongue body and an abnormal disturbance of range of movement, might be able on a second listening to fill in the detail of the settings to show that there was, say, grade 2 fronting, grade 3 raising, and grade 5 minimised range of tongue body movement.

The completion of the 'Second Pass' thus provides a detailed graphic representation of the speaker's vocal profile. In other words, it specifies the complex of long term components which characterise the speaker's voice.

(iii) Normal/Abnormal: The normal/abnormal division is somewhat problematic. There is insufficient information about the distribution of vocal settings in the population for the term "normal" to have a rigorous statistical sense, and it is difficult to formulate strict criteria for placing a given setting judgment on either side of the dividing line. A rough rule of thumb might be that settings judged as being in the abnormal range are those which require treatment. The suggestion does not stand up well to examination, however. The decision about treatment will obviously never be based on a protocol in isolation. Even when the vocal profile is taken into account alongside other factors such as diagnosis of pathology, the patient's own assessment of voice, etc., it is seldom single settings, but rather, as mentioned above, constellations of settings which cause the vocal profile as a whole to indicate the need for treatment. It is also the case that a particular speaker may have a profile for which his protocol shows no single setting as abnormal, but that nevertheless he is judged as in need of treatment because of unusual combinations of settings all within the normal range.

It does seem that speech therapists trained in the scheme agree about the 3/4 boundary rather more closely than they agree about other scalar degree boundaries, and it is tempting to assert that the normal/abnormal boundary must therefore be, in some as yet ill-defined way, a valid one. Given that the training programme presents this boundary as being important, and concentrates discussion upon it, the argument becomes very circular. It might be interesting to see if non-clinical phoneticians showed different tendencies.

In spite of the inherent theoretical problems, the normal/abnormal distinction does seem to be helpful, and serves as an anchor for perceptual judgments by emphasising an appropriate mid-point in the scale.

A further danger in the normal/abnormal area is that the protocol implies a continuum from grade 1 (normal) to grade 6 (abnormal). At a perceptual level this is acceptable, but it is necessary to differentiate quite clearly between a continuum of auditory quality and a continuum of underlying physiology. The auditory qualities we are concerned with can all be produced by anyone with a normal vocal apparatus, and most have relatively well-defined physiological correlates. The relationship between auditory quality and physiology is not yet completely understood, however, even in anatomically and physiologically normal speakers. Perceptually equivalent qualities may be produced by physiologically different mechanisms, and in pathological speech the auditory quality-physiology relationship may become very unclear. The continuum implicit in the form is therefore an auditory one only, and the evidence regarding the extent to which there is an underlying physiological continuum is not yet available.

A prerequisite of any clinical assessment is that the time expended in making the assessment should be in sensible proportion to the information gained. On first exposure, the task of completing a protocol form may seem somewhat formidable, and one which is out of proportion to the information gained. In practice, trained judges take between only five and fifteen minutes to evaluate each voice. Given the substantial amount of information contained in a completed protocol, this does not appear an excessive expenditure of time.

Other methods of transcription might have been chosen, but the graphic 'profile' approach has the two clear advantages of ease of completion, and ease of assimilation. Long hand verbal labels, following the tradition of the three-part labels used in segmental phonetics, would be very cumbersome with a scheme of this complexity as the need would be for twenty (or more) part labels. Not only would transcription be laborious, but reading and interpretation would be tedious. Phonetic symbols (available in Laver (1980, p.163)) are slightly faster to write, but both transcription and interpretation require considerable familiarity with the symbols.

Complex systems are generally most easily assimilated if presented in graphic form, and the additional ability of a semi-diagrammatic form to structure the process of auditory evaluation favours the use of the protocol for most purposes.

NATURE AND EFFECTIVENESS OF TRAINING IN THE USE OF THE PERCEPTUAL SYSTEM

An early assumption in the project had been that the perceptual system would be learnable from a package of taped material, with support from a manual. It soon became clear that a small amount of face-to-face training was desirable. A standard pattern of training has now emerged which seems economical and effective; the training programme starts with a preliminary half-day of theoretical presentation, with practical demonstrations. The members of the panel are then asked to read Laver (1980) and Laver and Hanson (1981), and to listen to the cassette provided with Laver (1980). In addition, they are given a 60-minute cassette (the 'Graded Reference Tape'), which contains patients' voices exemplifying nearly all the scalar degrees for all the setting categories, in ascending order of scalar degree. Some weeks later, a 2-day course is held, of intensive practical training in small groups, on both perception and production of all the settings. The ability to manipulate one's own vocal apparatus to produce a given setting is not essential, but it serves pedagogically as a useful focussing device, economically demonstrating the trainee's successful perception, and of course the ability is a potential asset in remedial work with patients. A tape of six test voices is then judged by the panel, and a final follow-up session is held some weeks later, both to communicate the statistical results of the test tape and to discuss the experiences of members of the panel in using the protocols in their own clinics.

The descriptive statistics on the most recent panel we have trained show that the 2½-day pattern is broadly satisfactory. Our method of assessment was as follows: three fully-trained judges in the MRC project team listened to the six voices on the test tape, and determined the notionally 'correct' perceptual judgments for each voice (on a slightly earlier version of the protocol). The performance of each panel-member was then quantified in terms of errors relative to the 'correct' protocols.

The 'correct' results were reached by each of the MRC judges listening independently, and then agreeing a consensus. This was reached under the following criteria: where three judgments occupied two adjacent scalar degrees, the majority judgment was taken as 'correct'; where three judgments occupied three adjacent scalar degrees, the middle judgment was taken as 'correct'; if the 3/4 boundary was involved in either of these cases, relistening was carried out until consensus; in all other cases, the voices were relistened to until consensus was reached under the above criteria. In determining these 'correct' results, 'neutral' was included as a scalar degree; some settings had a scalar range of seven degrees, therefore; and some, where polarity allowed a plausible continuum, as in the tension factors, had a scalar range of thirteen degrees. Modal voice and falsetto had ranges of only two degrees each, and audible nasal escape a range of four degrees. On this basis, considering only the vocal quality features, the average initial disagreement between any two MRC judges was 16.33 errors per voice - i.e. a discrepancy of slightly less than one scalar degree for each of the 21 parameters.

Before discussing the panel results, it is perhaps necessary to state what, under acceptably severe criteria, would constitute adequate performance on the part of the panel. Three 'classes'

of performance were decided. The first (Class 1) would be where a judge scored an average, over the six test voices, of not more than 1 error on a given parameter. This would be a standard broadly comparable to that of the experienced judges of the MRC group, and would therefore show a need for no further training on that parameter.

Class 2 would represent a performance scoring an average error on a given parameter of between 1 and 2. This would constitute the minimum acceptable performance allowing the descriptive system to be used practically in clinics. To maximise the usefulness of the parameter involved, a slight amount of further work with taped materials would be necessary.

Class 3 would represent an unacceptable performance on a given parameter, where the average error score over the six test voices was 2 or more. Substantial further training would be necessary before that judge could reliably use that parameter in clinical situations.

The overall results for the 10 panel judges were that the average number of errors per voice ranged from 18.67 (comparable to the standard of the MRC judges) to 25.67 (still very competent). The acceptability of the performance of the whole panel, averaging the error scores across all judges and listing the averages per parameter, is shown in Figure 2. It will be noted that 18 parameters out of 21 were scored at Class 2 or less. It should be said that none of the test voices had non-neutral values of labiodentalization, breathiness or falsetto. Good scores on these parameters are somewhat misleading, therefore. But a positive judgment of 'neutral' is still necessary in such cases to score a correct result, so their relative success should not be entirely discounted.

Figure 3 shows a comparable set of results for 10 judges trained over a period of eight weeks, at eight 90-minute sessions. (Tip/blade factors were not tested.) The differences in the scores are virtually negligible.

Using non-parametric statistics, it is possible to penetrate the performance shown by the judges in Figure 2 a little more deeply. Taking the average error for a given setting-category, we can ask the question 'Does this degree of agreement with the MRC group's result reliably indicate that the panel judges were using a standard criterion of judgment in listening to this setting in the six voices, or could their judgments have arisen by chance?'. Taking two results, Harshness (Average Error 1.37) in Class 2, and Tongue Body Raising-Lowering (Average Error 2.43) in Class 3, Kendall's Coefficient of Concordance (W)(Siegel 1956: 229) was computed[12]. In the case of Harshness, where W = 0.45 and s = 647.5, we can reject the hypothesis that the panel reached their judgments by chance, at a level of significance greater than .01. In the case of Tongue Body Raising-Lowering, the association between W(.182) and s(171.5) is not significant at the .05 level, so we cannot confidently reject the hypothesis that the error score did not reflect a merely random choice.

The results of the two panels illustrated suggest that the Class 3 performances need considerable further work - either conceptually in the descriptive system or in the training of panel

Vocal Quality Setting	Average Error	Acceptability Class
Breathiness	.03	1
Labiodentalization	.05	1
Modal Voice	.07	1
Falsetto	.12	1
Retroflexion	.17	1
Audible Nasal Escape	.20	1
Tongue Body Range	.48	1
Tip Articulation	.53	1
Nasality-Denasality	.78	1
Whisperiness	.85	1
Creakiness	.93	1
Mandibular Range	1.15	2
Open-Close Jaw	1.30	2
Harshness	1.37	2
Lip Rounding-Spreading	1.37	2
Labial Range	1.72	2
Laryngeal Tension	1.77	2
Supralaryngeal Tension	1.90	2
Larynx Position	2.23	3
Fronted-Backed Tongue Body	2.35	3
Raised-Lowered Tongue Body	2.43	3

FIGURE 2 Acceptability of average error scores per vocal parameter for ten judges trained on a two and a half day programme (Class 1 = 'good performance'; Class 2 = 'acceptable'; Class 3 = 'needs substantial further training')

Vocal Quality Setting	Average Error	Acceptability Class
Falsetto	0.03	1
Labiodentalization	0.13	1
Audible Nasal Escape	0.17	1
Modal Voice	0.48	1
Tongue Body Range	0.58	1
Breathiness	0.85	1
Creakiness	0.92	1
Mandibular Range	1.05	2
Harshness	1.30	2
Whisperiness	1.58	2
Labial Range	1.60	2
Laryngeal Tension	1.60	2
Open-Close Jaw	1.63	2
Rounded-Spread Lips	1.70	2
Nasal-Denasal	1.75	2
Larynx Position	2.20	3
Supralaryngeal Tension	2.20	3
Fronted-Backed Tongue Body	2.25	3
Raised-Lowered Tongue Body	2.35	3

FIGURE 3 Acceptability of average error scores per vocal parameter for ten judges trained on an eight week, one and a half hour session per week programme (Class 1 = 'good performance'; Class 2 = 'acceptable'; Class 3 = 'needs substantial further training')

judges, but that the remaining Class 2 and Class 1 performances, on the very large majority of parameters, reflect the acceptable effectiveness of a 2½-day training method, and the basic fact of the trainability of experienced speech therapists in the use of the descriptive system.

APPLICATIONS OF THE DESCRIPTIVE SYSTEM

The descriptive system can readily find applications in any discipline where a quantified, written record of long-term vocal features is of interest. Some of these applications, in phonology, sociolinguistics, paralinguistics, anthropology, ethnomusicology, social psychology, psychiatry, and communications engineering, are discussed in Laver (1980:10-11). In the immediate context of this project, the most central applications are those in speech therapy and speech pathology. Within the project, use of the VPA protocol on the eight groups of disorders is showing interesting early results. Figure 4 is a 'summated' protocol, amalgamating the protocols for the first 14 of the Down's Syndrome subjects. The numbers in the cells represent numbers of individual subjects judged as showing the scalar degree of the setting concerned. The group protocol was prepared from a computer printout: all the individual protocols are being stored on computer disks, using interactive programs written by Steven Hiller for the Phonetics Laboratory DEC PDP 11/40 computer, and this store can be explored in a variety of ways. PROSUM is the program which amalgamates specified protocols, and is proving an extremely convenient way of showing group trends. As Figure 4 demonstrates, our working hypothesis, that particular speech disorders have characteristic vocal profiles associated with them, is only a little too strong. Very clear trends are visible in the Down's Syndrome group towards characterization by a constellation of central features, with some other features of the profile playing a weaker role.

Collaborative applications of the VPA protocol system with members of the speech therapy panels, using the protocols in their own clinics, have focussed on use in three functions: as an instrument to record a quantified judgment of a vocal profile on one occasion; as an aid in planning strategies and goals of remedial therapy; and as a device for measuring the detail and scale of progress under rehabilitative treatment.

The detailed findings of the project's perceptual and acoustic analyses of the eight groups of patients, together with a manual for the use of the descriptive system, accompanied by cassette material, will be published by Cambridge University Press in 1983 (Laver & Wirz). But we hope that the most important applications will be those by speech therapists and pathologists who become trained in the system and apply it in their own clinics.

Vocal Profile Analysis Protocol

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial	8	6	0	Lip Rounding/Protrusion	1	2				
				Lip Spreading	1	2				
				Labiodentalization						
	3	10	1	Extensive Range						
				Minimised Range	1	3	6	1		
2. Mandibular	2	12	0	Close Jaw						
				Open Jaw	4	2	1			
				Protruded Jaw	5	6	1			
	9	10	1	Extensive Range		1				
				Minimised Range		6	4	1		
3. Lingual Tip/Blade	2	11	1	Advanced	3	1	3	1		
				Retracted	1	3				
4. Lingual Body	0	10	3	Fronted Body		5	3	3		
				Backed Body		1	1			
		10	0	Raised Body	1	6	3			
				Lowered Body						
	0	9	5	Extensive Range						
				Minimised Range		3	6	5		
5. Velopharyngeal	0	7	6	Nasal			7	7		
				Audible Nasal Escape				1		
				Denasal						
6. Pharyngeal	9	5	0	Pharyngeal Constriction	4	1				
7. Supralaryngeal Tension	3	11	0	Tense	1	2	1			
				Lax	3	3	1			
B. Laryngeal Features										
8. Laryngeal Tension	0	14	0	Tense	1	7	4			
				Lax		1	1			
9. Larynx Position	6	8	0	Raised						
				Lowered	4	3	1			
10. Phonation Type	0	11	2	Harshness		7	4	1	1	
		6	8	Whisper(y)		1	5	7	1	
				Breathiness						
		7	0	Creak(y)	3	1	3			
		0		Falsetto	0					
		14		Modal Voice	14					

II PROSODIC FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
1. Pitch	3	10	1	High Mean	2					
				Low Mean	2	4	2	1		
	1	13	0	Wide Range						
				Narrow Range	5	5	3			
	10	4	0	High Variability						
				Low Variability	1	1	2			
2. Consistency	13	1	0	Tremor		1				
3. Loudness	1	12	1	High Mean	5	2	1			
				Low Mean	1		3	1		
	7	7	0	Wide Range						
				Narrow Range	2	3	2			
	10	4	0	High Variability						
				Low Variability		2	2			

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh. © June 1981.

FIGURE 4 (above and on facing page) Summated protocol for 14 Down's Syndrome subjects

References

Honikman, B. (1964) 'Articulatory settings'. In Abercrombie, D. et al. (eds.) In Honour of Daniel Jones. Longman: London, 73-84.

Laver, J. (1968) 'Voice quality and indexical information' British Journal of Disorders of Communication 3: 43-54.

---------- (1974) 'Labels for voices' Journal of the International Phonetic Association 4: 62-75.

---------- (1975) Individual features in voice quality. Ph.D. dissertation, University of Edinburgh.

---------- (1979) Voice Quality: A Classified Bibliography. John Benjamins: Amsterdam.

---------- (1980) The Phonetic Description of Voice Quality. Cambridge University Press.

Laver, J. and Hanson, R. (1981) 'Describing the normal voice' In Darby, J. (ed.) Speech Evaluation in Psychiatry. Grune and Stratton: New York, 51-78.

Laver, J. and Trudgill, P. (1979) 'Phonetic and linguistic markers in speech' In Scherer, K. R. and Giles, H. (eds.) Social Markers in Speech. Cambridge University Press, 1-32.

Laver, J. and Wirz, S. (1983) Vocal Profiles. Cambridge University Press (in preparation)

Pashayan, H. M. (1975) 'The basic concepts of medical genetics' Journal of Speech and Hearing Disorders 40: 147-163.

Siegel, S. (1956) Non-parametric Statistics. McGraw-Hill Book Company: New York.

Footnotes

1. The project is funded by the Medical Research Council (Grant No. G978/1192, 'Vocal Profiles of Speech Disorders'), and is under the direction of John Laver and Sheila Wirz.

2. We are very grateful to our two project consultants, Dr. W. I. Fraser, MD, DPM, FRCPsych, consultant psychiatrist at the Royal Edinburgh and Gogarburn Hospitals, and Senior Lecturer in Psychiatry and Rehabilitation Studies, University of Edinburgh, until recently Physician Superintendent at Lynebank Hospital, Fife, and Dr. Shirley Ratcliffe, MB, BS, FRCP, consultant pediatrician in the MRC Clinical Population and Cytogenetics Unit at the Western General Hospital in Edinburgh, for their advice and help in arranging access to many of our patient subjects. We are also very grateful to our cooperative subjects.

3. From the National Technical Institute for the Deaf, Rochester, New York, through the kind assistance of Professor Joan Subtelney.

4. From the Scottish Council for Spastics, through the collaboration of Mrs. Alison McDonald, the Council's Chief Therapist.

5. Mostly from Gogarburn and Lynebank Hospitals, through Dr. Fraser. Each subject had already been included in a prior MRC genetic survey, so that cases of Trisomy-21 were distinguishable from mosaic cases.

6. This group was made up mostly of patients with Klinefelter's Syndrome, with access through Dr. Ratcliffe and her colleagues, notably Dr. W. H. Price, BSc., MB, BCh, FRCPE, in the MRC Clinical Population and Cytogenetics Unit.

The rationale for including both Down's Syndrome (as an example of an autosomal defect) and Klinefelter's Syndrome (as a sex-chromosome defect) is that 'Patients with chromosomal aberrations usually have characteristic phenotypes, closely resembling those of other patients with the same abnormality' (Pashayan 1975: 154). We hypothesise an extension of this organic resemblance to include vocal features.

7. This group of recordings was kindly made available by Professor F.I. Caird, MA, DM, FRCP, of the Department of Geriatric Medicine, Southern General Hospital, University of Glasgow, and is being analysed collaboratively with Ms Sheila Scott, the speech therapist on Professor Caird's project on Parkinson's Disease.

8. The thyroid group was made up of two sub-groups: an edematous hypothyroid group on chemotherapy, and a hyperthyroid group undergoing chemotherapy or surgery. Access was kindly arranged by Dr. A. D. Toft, BSc, MD, MRCP, consultant endocrinologist in the Department of Medicine, Royal Infirmary, University of Edinburgh.

9. Dysphonic patients' recordings were provided by a number of collaborating therapists. The large majority came from Mrs. Marion Mackintosh, in charge of the Voice Clinic, Royal Infirmary, Edinburgh.

10. The recordings of cleft palate speakers were kindly provided by the members of a Scottish Home and Health Department research project on cleft palate, Dr. A. C. H. Watson, MB, ChB, FRCSE, of the Department of Clinical Surgery, University of Edinburgh, Mr. J. K. Anthony, CEng, MIEE, of this Department, and Ms R. Razzell of the Speech Therapy Department, Royal Hospital for Sick Children, Edinburgh. The material is being analysed collaboratively with Ms Razzell.

11. The therapists in the training panels (to all of whom we express our thanks) seemed comfortable with six scalar degrees, usually. It may be, however, that for particular applications, a smaller number might be more suitable (or in special circumstances say in the judgment of nasality in cleft palate speech, a larger number of degrees for the velopharyngeal settings might be desirable). Six degrees seems a practical number, to allow the assessment of progress in therapy, and to facilitate training in the system.

12. We are grateful to Mrs. Anne Anderson of this Department for statistical advice in this connection.

APPENDIX THREE

Edinburgh University, Dept of Linguistics, Work in Progress
80 16, 80-116, 1983.

STRUCTURAL PATHOLOGIES OF THE VOCAL FOLDS AND PHONATION

Janet Mackenzie, John Laver and Steven M. Hiller

ABSTRACT

The vocal fold is considered as a multi-layered structure. Pathologies of this structure give rise to perturbations of the laryngeal waveform that may be diagnostic of the type of pathology. An account is offered of the layered anatomy of the vocal fold, and of the histological and mechanical characteristics of the individual layers. A typology of structural pathologies is advanced, and initial suggestions are made about the consequences of these pathologies for the detailed mode of vibration of the vocal folds.

A current research project in the Phonetics Laboratory ('Acoustic Analysis of Voice Features' Medical Research Council Grant No. 8207136N, 1982-85) is exploring an automatic acoustic method for characterizing pathological voices. It has three broad objectives. These are, in order:

1. the development of an automatic acoustic system for screening voices for potential laryngeal pathology

2. the acoustic differentiation of various pathologies of the larynx

3. the acoustic evaluation of the degree of progressive deterioration of a laryngeal pathology, or of the degree of rehabilitative progress being made

The project brings together two strands of research. One is research into acoustics and computing. A progress report on this work is available in a companion article in this volume (Hiller, Laver and Mackenzie 1983). The other strand, which is the topic of the present article, concerns normal and pathological aspects of laryngeal anatomy and physiology.

The plan of this article will be to consider first the concept of the true vocal fold as a multi-layered structure made up of a body (the vocalis muscle) and a cover (the epithelium and the underlying lamina propria). Then there is a discussion of the mechanical properties of the tissue-types in each of these layers. The effect of disruptions of inter-layer relationships is then examined, and a typology of structural pathologies of the vocal folds is suggested, based on the type of disruption and changes in mechanical properties. Hypotheses are framed about the possible consequences of different types of pathology for the detailed mode of laryngeal vibration. Finally, summary descriptions of each major vocal pathology are given in an Appendix.

The pre-occupation of the project is the potential for acoustic measurement of vocal disorders. To be susceptible of acoustic registration, a vocal disorder must show either a structural or a functional change from the characteristics of the healthy, normal larynx. This article will concentrate on 'structural' pathologies only, where the disorder involves a structural alteration of the

vocal fold. Further, we shall survey only the more commonly encountered structural pathologies of the vocal fold. Phonatory problems that arise in the absence of any structural alteration will not be considered in any detail. These include neuromuscular disorders, such as paralyses of the vocal folds, as well as a range of psychogenically induced voice disorders where there is no organic change.

An examination of the literature on vocal fold pathology reveals that classification of disorders usually uses criteria related either to the underlying pathology, or to the presumed aetiology. The term 'pathology' is used here to describe processes acting within the tissues in the development of a disorder, such as inflammation or neoplastic change ('neoplastic' refers to altered patterns of tissue growth in tumour formation). The term 'aetiology' can then be reserved for factors which arise externally to the tissues, as in infection, or mechanical abuse of the tissues.

The overriding concern of the medical profession, properly, is to identify the pathological processes involved in a given disorder, since these play a large part in determining the most appropriate treatment. The medical literature is therefore typified by classifications based on the underlying pathology, such as that shown below:

1. Inflammatory conditions
   - i. acute
   - ii. chronic
2. Neoplasms (tumours)
   - i. benign
   - ii. malignant
3. Congenital malformations
4. Traumatic injury

(e.g. Hall and Colman 1975, Ballantyne and Groves 1977, Birrell 1977).

There are some demarcation difficulties with this approach, in that there is no clear agreement about the borderline between chronic inflammatory conditions and some benign tumours. Vocal polyps, for example, are considered by some authors to be inflammatory in origin (New and Erich 1938, Arnold 1962, Aronson 1977, Friedmann and Osborn 1978), and by others to be instances of benign tumours (Birrell 1977).

The speech therapy literature is understandably more concerned with the extent to which poor phonatory habits may be involved in the aetiology of a vocal fold disorder. Hence a distinction is often drawn between those disorders which arise apparently independently of any vocal misuse, versus those which are considered to be the sequel of faulty habitual phonation. The latter type are often called 'functional' or 'psychogenic' disorders (Luchsinger and Arnold 1965, Greene 1972, Aronson 1977, Perkins 1977), in contrast to the former group of 'organic' disorders.

This approach also has a demarcation problem. There seems to be general agreement that vocal nodules, for example, are 'functional' in that they arise most often in speakers who habitually misuse their vocal folds. They may therefore be classed with disorders like conversion aphonia (hysterical loss of voice) or spastic dysphonia (extreme adductive compression of the vocal folds), which exhibit no structural abnormality. Vocal nodules are, however, clearly 'organic', in the sense that there is a structural abnormality of the vocal folds. When fully developed, they may even be indistinguishable, both macroscopically and histologically, from certain types of tumour (Shaw 1979). It is also very difficult to disentangle the relative contributions of 'organic' predisposition and 'functional' misuse in the causation of a disorder. Arnold (1962) considers the role of various predisposing factors in vocal nodule formation, and even in this most 'functional' of vocal fold lesions it seems likely that factors such as general bodily health and infection may play an important part.

The focus of this project is the potential effect of vocal fold disorders on vibratory patterns of the folds, and hence on the acoustic signal. Alterations in aerodynamic and mechanical properties of the larynx thus become of no less importance than pathological and aetiological factors. This paper aims to draw together some of the available information on structural disorders of the vocal fold, in such a way that it may be possible to develop preliminary hypotheses about their differential effects on phonatory output.

A. NORMAL VOCAL FOLD STRUCTURE

It is not possible to predict the mechanical consequences of alterations in vocal fold structure without having some acquaintance with the structure and mechanical properties of the normal vocal fold.

The anatomy of the cartilages, muscles and other tissues which make up the larynx has been extensively described elsewhere (Kaplan 1960, Saunders 1964, Hardcastle 1976, Romanes 1978, Laver 1980). We shall concentrate only upon the tissues of the vocal folds themselves, and the cartilages with which they are intimately associated. This is not meant to imply that structural alterations elsewhere in the larynx are expected to have no phonatory consequences, since this is clearly not the case. Growths in the areas above and below the glottal zone may indeed have quite dramatic effects on phonation if they physically impede vocal fold movement or cause significant airway obstruction. More subtle effects can also be expected from any structural anomaly that disturbs the rate or direction of airflow through the glottis itself. These can be thought of as external constraints on vocal fold vibration, however, and as such they will not be considered in this article.

The anatomical focus of attention will be the region bordered anteriorly and laterally by the thyroid cartilage, and extending as far back as the posterior edges of the arytenoid cartilages. In the vertical dimension, the region includes only the true vocal folds, and so the inferior border can be drawn at the level of the upper edge of the cricoid cartilage.

A convenient distinction can be made between the anterior two thirds of each vocal fold, which is bordered at the glottal edge by the vocal ligament, and the posterior one third, where the inner edge of the arytenoid cartilage, from the vocal process to the inner 'heel' of the cartilage, forms the glottal border. We can then refer to the 'ligamental' part of the fold and the 'cartilaginous' part. This follows the convention initiated by Morris (1953) of distinguishing between the intermembranecus (or ligamental) glottis and the cartilaginous glottis.

A schematic plan of the vocal fold region is shown in Figure 1. The following account offers a brief description of the tissues which make up the vocal folds, together with some comment on the mechanical properties of each tissue type. Implications for pathological alterations within the folds are then discussed.

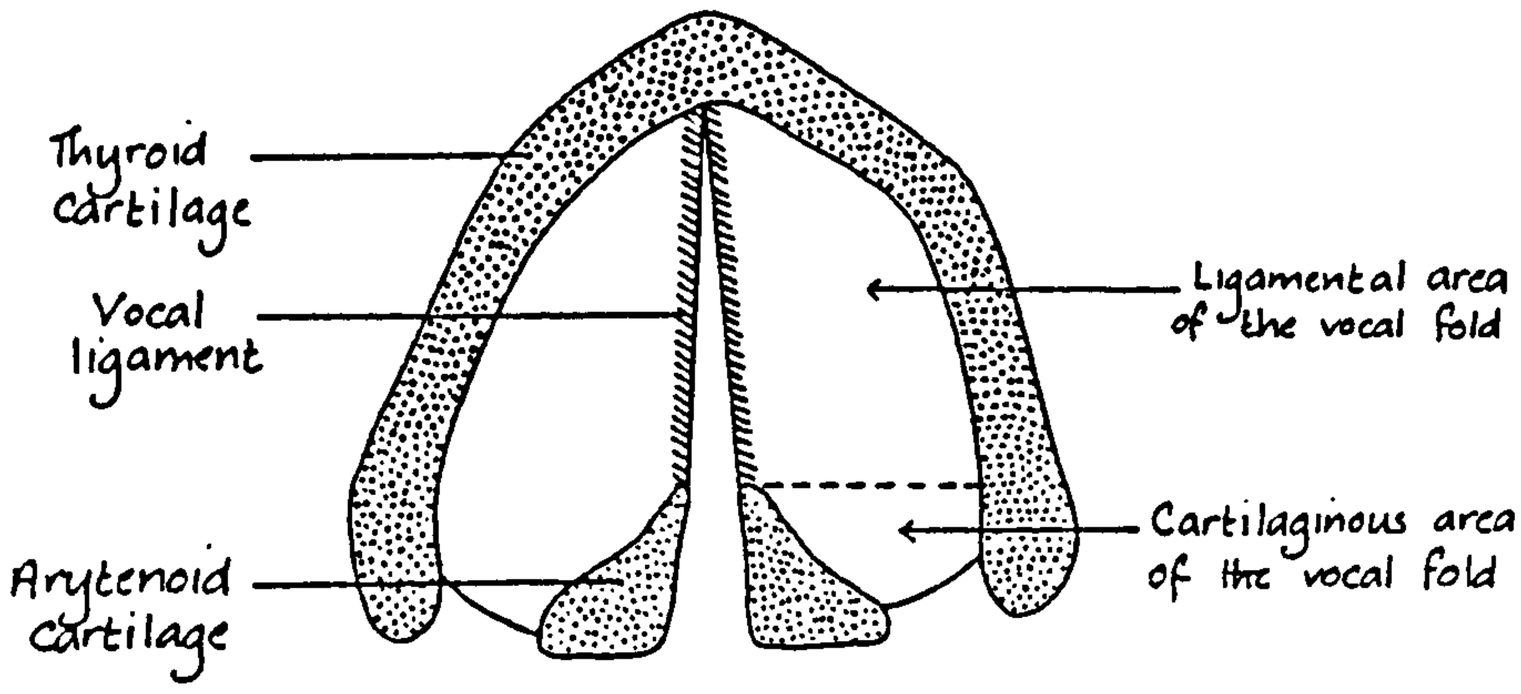


Figure 1. A schematic view of the vocal folds, seen from above.

1.0. The ligamental area of the vocal fold

The ligamental area of the vocal fold is the one most freely involved in vibration during phonation, and it has therefore attracted the most attention from researchers concerned with vocal fold mechanics. Hirano and his associates have recently built up a considerable body of information about the histological structure of the vocal fold, and their work necessarily forms a base for the account that follows (Hirano et al. 1980, Hirano 1981, Hirano et al. 1982). Background sources also include standard texts on anatomy and histology (Davies and Davies 1962, Freeman and Bracegirdle 1967, Romanes 1978).

1.1. Tissue types

The vocal fold is a layered structure, which in the ligamental area consists of the vocalis muscle and a covering of mucous membrane. The importance of these two layers in determining the

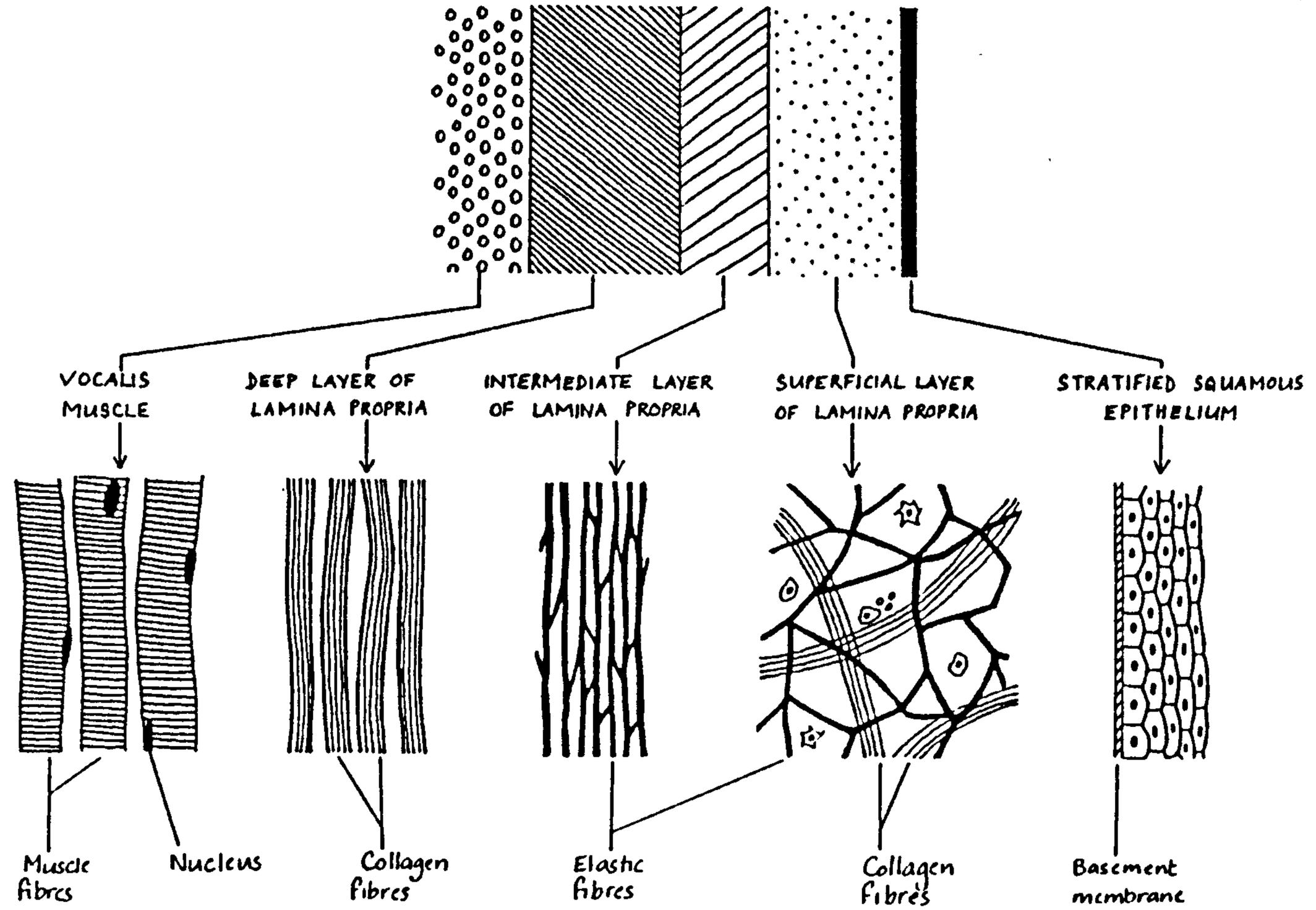


Figure 2. A diagrammatic representation of the tissue layers of the vocal folds.

fine detail of phonation has long been accepted (Smith 1961, Perello 1962, Baer 1973), but Hirano's work focuses attention on yet further tissue distinctions within the mucous membrane. This is divisible into four layers: an outer layer of epithelium, and three layers of underlying connective tissue. These inner three layers together make up the lamina propria (see Figures 2 and 3).

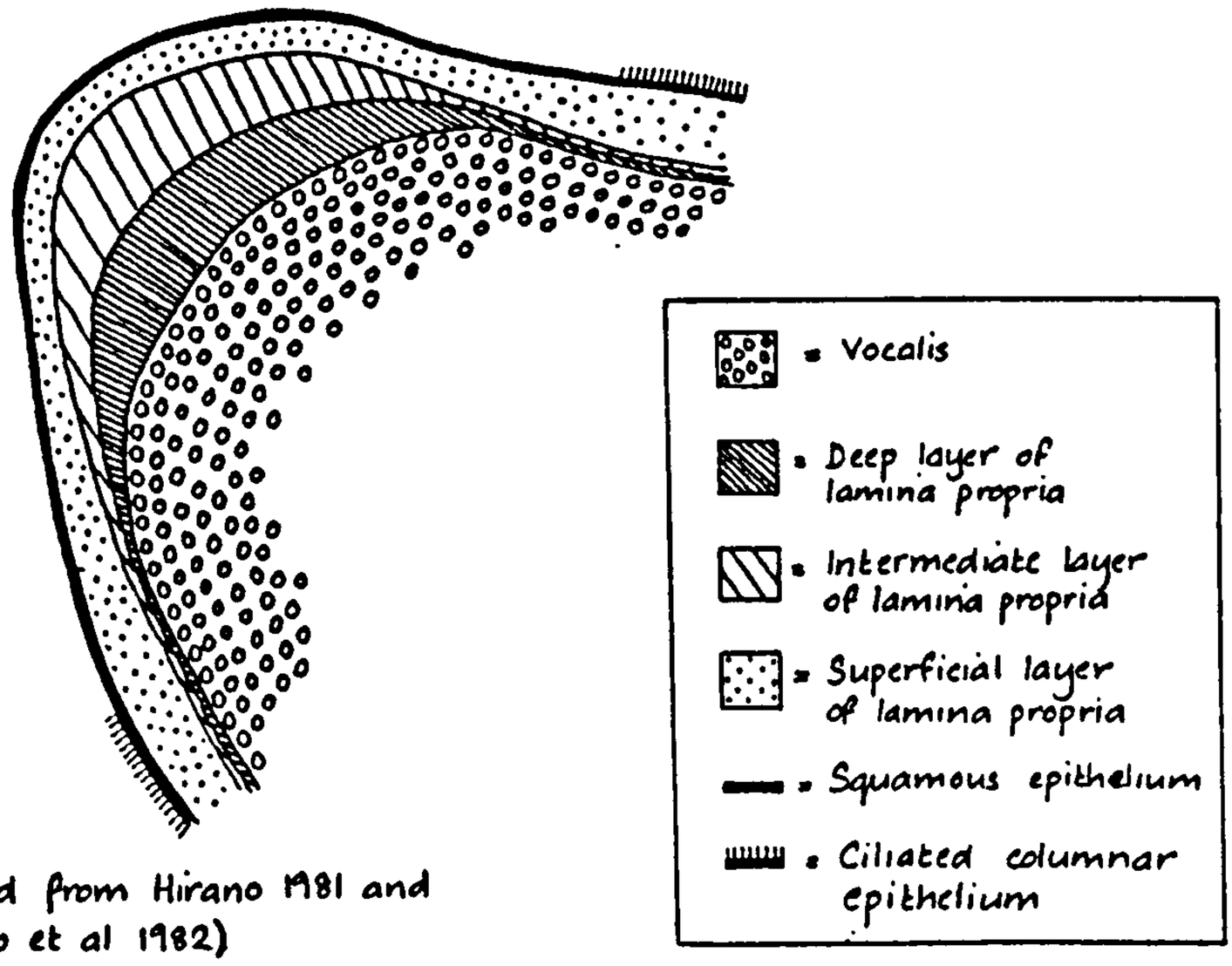


Figure 3. A schematic representation of the ligamental portion of the vocal fold, seen in cross section.

1.1.1. Epithelium

Epithelium is the generic name for all the tissues which line the internal and external surfaces of the body. It occurs in various forms, but all are characterized by a pattern of closely packed cells, cemented together by a minimal amount of inter-cellular matrix. The epithelial covering of the free border of the vocal fold is of a type known as non-keratinizing stratified squamous epithelium (see Figure 2). These three descriptive labels relate simply to the detailed structure of this area. It is non-keratinizing because it does not produce keratin. Keratin is the substance which forms the horny layer in the skin covering the external surface of the body. The term 'stratified' describes the arrangement of the cells, which are here arranged in orderly layers, with the deepest layer resting on a basement membrane. The basement membrane is a zone where substances similar to those found between the cells and fibres of the underlying lamina propria are highly condensed, to form a thin sheet dividing the two tissue types.

Epithelium undergoes constant regeneration by replication of the basal cell layer (i.e. the layer of cells lying closest to the basement membrane), and in normal tissue this process is sufficiently organized to give a clearly stratified structure. The most mature cells are those on the surface of the fold. The number of cell layers in the epithelium probably varies considerably, but in a large post-mortem study of 942 adult male larynges, Auerbach et al. (1970) found most samples of vocal fold epithelium to be between 5 and 10 cells thick. Hirano et al. (1982:278) report that there is no systematic relationship between epithelial thickness and age.

The term 'squamous' refers to the shape of the cells, which are commonly likened to paving stones. In surface view they are usually polygonal, but cross sections show flattening, especially in the surface layers.

On the upper and lower surfaces of the vocal fold there is a transition to ciliated columnar epithelium (see Figure 3). The cells here are taller than the squamous cells, and carry cilia (microscopic hair-like projections) protruding from their surfaces.

The epithelium of the canine vocal folds, which appears histologically to be very similar to that in humans, has been tested mechanically by Hirano and his colleagues (Hirano et al. 1982), and it seems to be a relatively stiff, non-elastic tissue. In other words, compared with the underlying lamina propria, it requires greater stress to stretch it by a given amount. It is assumed, because the cells do not show any directionality in their arrangement, that the tissue will be isotropic. That is, it will be equally easy (or difficult) to stretch it in longitudinal or transverse directions.

1.1.2. Lamina propria

The lamina propria consists of three layers of connective tissue. Some types of connective tissue (bone, cartilage) form the skeletal framework of the body, whilst others act as structural coordinators, binding organs, muscles and nerves to each other and to the skeleton. The nature of any given connective tissue is determined less by the cells than by the non-cellular matrix within which they are contained. This matrix may contain fibres of various kinds which can also be important in determining the mechanical properties of the tissues.

The vocal ligament derives from a thickening of the intermediate and deep layers of the lamina propria, but it will be discussed in more detail in a later section.

a. superficial layer of the lamina propria

The layer of the lamina propria lying immediately below the epithelium consists of areolar tissue (see Figure 2). Cells are embedded in a soft, semi-fluid matrix, which contains a loose network of haphazardly arranged elastic and collagen fibres. These fibres will be discussed further in relation to the intermediate and deep layers of the lamina propria.

Hirano (1981:5) likens this layer to soft gelatin, and it is probably the most pliable of the vocal fold tissues. Titze (1973) in his mathematical model of vocal fold vibration assumes that it

acts like a fluid. Unfortunately, the experiments by Hirano et al. (1982) on canine lamina propria do not allow extrapolation to the human tissue, because the lamina propria of the dog does not exhibit a comparable three-layered structure..

An alternative name for this layer, Reinke's Space, signals that this is a potential site for loss of the normally tight attachment of the mucous membrane to the vocalis muscle.

b. intermediate layer of the lamina propria

The next layer of connective tissue has a much higher fibre content. These are mostly elastic fibres, formed from a protein called elastin, and they are arranged in an orderly fashion running parallel to the free border of the vocal fold (i.e. anterior to posterior). Elastic fibres are quite fine, and they form branches and cross links with adjacent fibres (see Figure 2). Hirano's analogy between elastic fibres and rubber bands (1981:5) highlights, as does the name, their marked elastic properties. Freeman and Bracegirdle (1967:20) describe them as having 'considerable' elasticity, and Fields and Dunn (1973) report that they are three times easier to stretch than collagen fibres (see next section). Freeman and Bracegirdle (ibid.) also state that they have 'little tensile strength'. The parallel arrangement of the fibres in the tissue is assumed to cause considerable anisotropy. That is, elasticity, as judged by the stress required to stretch the tissue by a given amount, will be different when the stress is applied in a direction parallel to the fibres from that measured with the stress at right angles to the course of the fibres. The tissue is further assumed to be incompressible (Titze 1973).

c. deep layer of the lamina propria

The deep layer of the lamina propria is similar in structure to the intermediate layer, in that it is rich in fibres which are arranged parallel to the edge of the vocal fold. In this layer, however, the fibres are mostly formed from the protein collagen. This forms rather coarser fibres than elastin, and collagen fibres are unbranched. Hirano's analogy here is with cotton thread, emphasising the relative non-elasticity of collagen when compared with elastin (Freeman and Bracegirdle 1967, Fields and Dunn 1973). Like the intermediate layer, the deep layer is assumed to be anisotropic and incompressible.

1.1.3. The vocalis muscle

The body of the vocal fold is composed of part of the thyroarytenoid muscle, the vocalis, which is voluntary striated muscle tissue (ordinary skeletal muscle). In spite of controversial suggestions by Goerttler (1950) that the vocalis muscle fibres run at an angle to the edge of the vocal fold, it is now generally accepted that they in fact run parallel to the edge of the fold.

The mechanical properties of muscle vary dramatically, depending on its state of contraction. Hill (1970), cited by Hirano et al. (1982), suggests as much as a tenfold difference in elasticity between resting and contracted muscle. Resting muscle from the canine vocal fold is easier to stretch than either the lamina propria or the epithelium, but like them it is assumed to be incompressible. Anisotropy is also expected, because of the parallel fibre arrangement.

1.2. The vocal ligament

Figure 3, which represents a cross section of the vocal fold at the midpoint of the ligamental area, shows the uneven distribution of these tissue layers. Over the upper and lower surfaces of the vocal fold the intermediate and deep layers of the lamina propria are very thin, but at the glottal edge they become greatly thickened, and constitute the part known as the vocal ligament.

The relative thicknesses of the layers of the lamina propria vary along the length of the vocal ligament. The superficial layer is thinner at the ends than in the middle, whilst the intermediate, elastic layer is thicker at the ends (Hirano 1981:7, Hirano et al. 1982:276). Figure 4a shows our calculations for longitudinal variations in tissue thickness from data presented by Hirano et al. (1982:275) for five females and five males. This represents a rather small sample, but the figures can probably be accepted as being illustrative of general tendencies.

Figure 5 shows that the intermediate layer of the lamina propria is greatly thickened in a small area at each end of the vocal ligament. These thickened areas, the anterior and posterior maculae flavae, act as cushions of elastic material, and probably afford some protection against impact during vocal fold vibration. The reduced depth of elastic and collagen fibres at the centre of the ligamental portion increases pliability in this area.

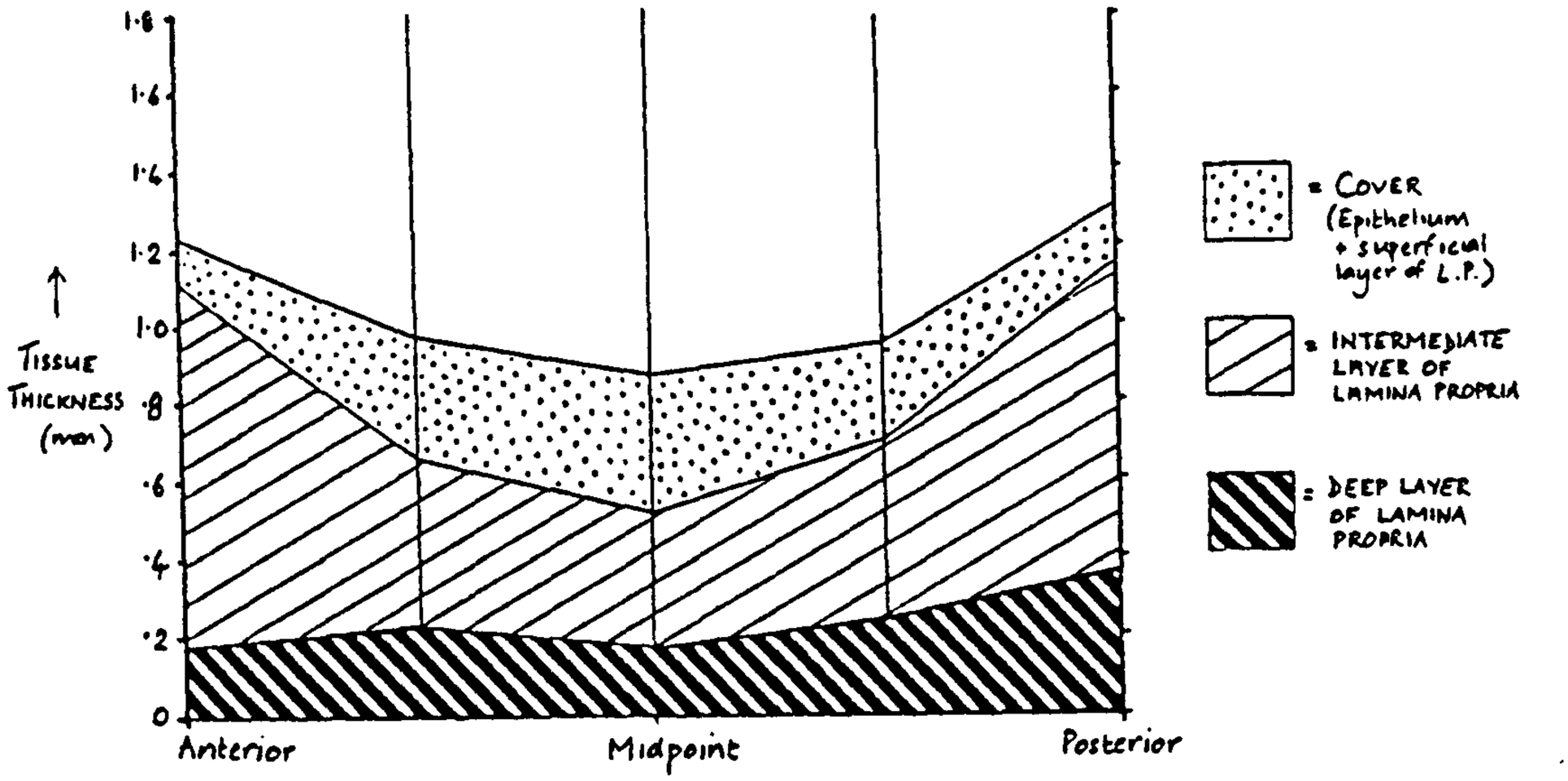
1.3. Age-related changes in tissue thickness

The measurements used in Figure 4a reflect the state of the laryngeal tissues in young adults, and cannot be taken as representative of all age groups. Young children seem to exhibit only a rudimentary vocal ligament, and the adult tissue layer relationships are not seen until after puberty. After histological examination of 48 male vocal folds of subjects between 0 and 70 years of age, Hirano, Kurita and Nakashima (1981:39) wrote:

> 'In a newborn, no vocal ligament is observed. The entire lamina propria looks rather uniform and pliable in structure. The fibrous components are slightly dense only at the ends of the vocal fold. In a four year old child, a thin and immature vocal ligament is observed. The vocal ligament is still immature at the ages of 12 and 16. It is only after puberty that a mature layer structure forms.'

After reaching maturity, too, there may be continuing changes in tissue thickness, and these are indicated in Figure 4b, which represents measurements of larynges from subjects in their 50's. A comparison between Figures 4a and 4b suggests that in the older larynx there is an increase in the thickness of the cover relative to the intermediate and deep layers of the lamina propria. Hirano et al. (1982:278) found no systematic age-related changes in epithelial thickness, so the increased depth of the cover is attributed to changes in the superficial layer of the lamina propria. A decrease in fibre density in this layer is also reported. It would be interesting to know if this trend could be confirmed by

i. Females



ii. Males

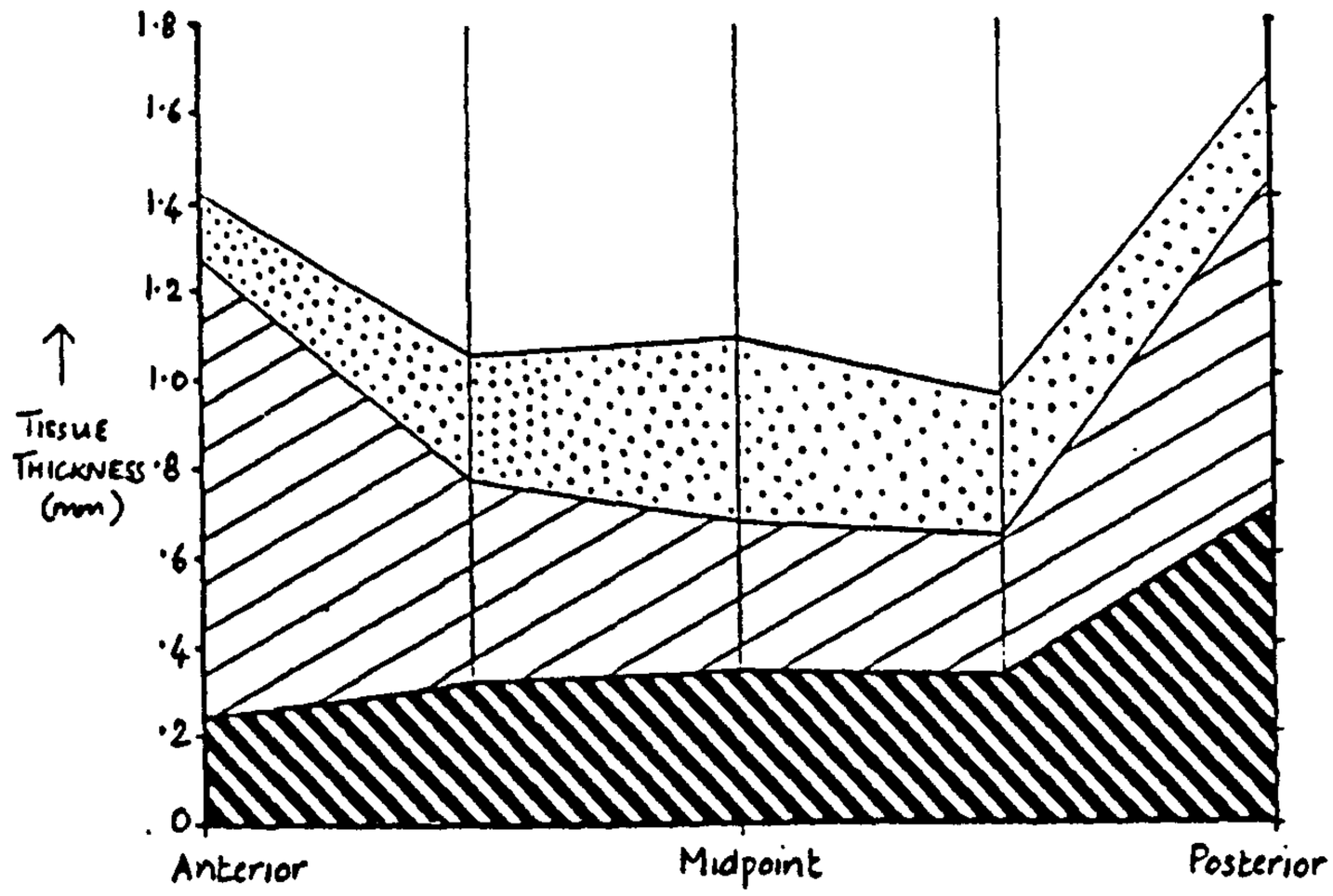
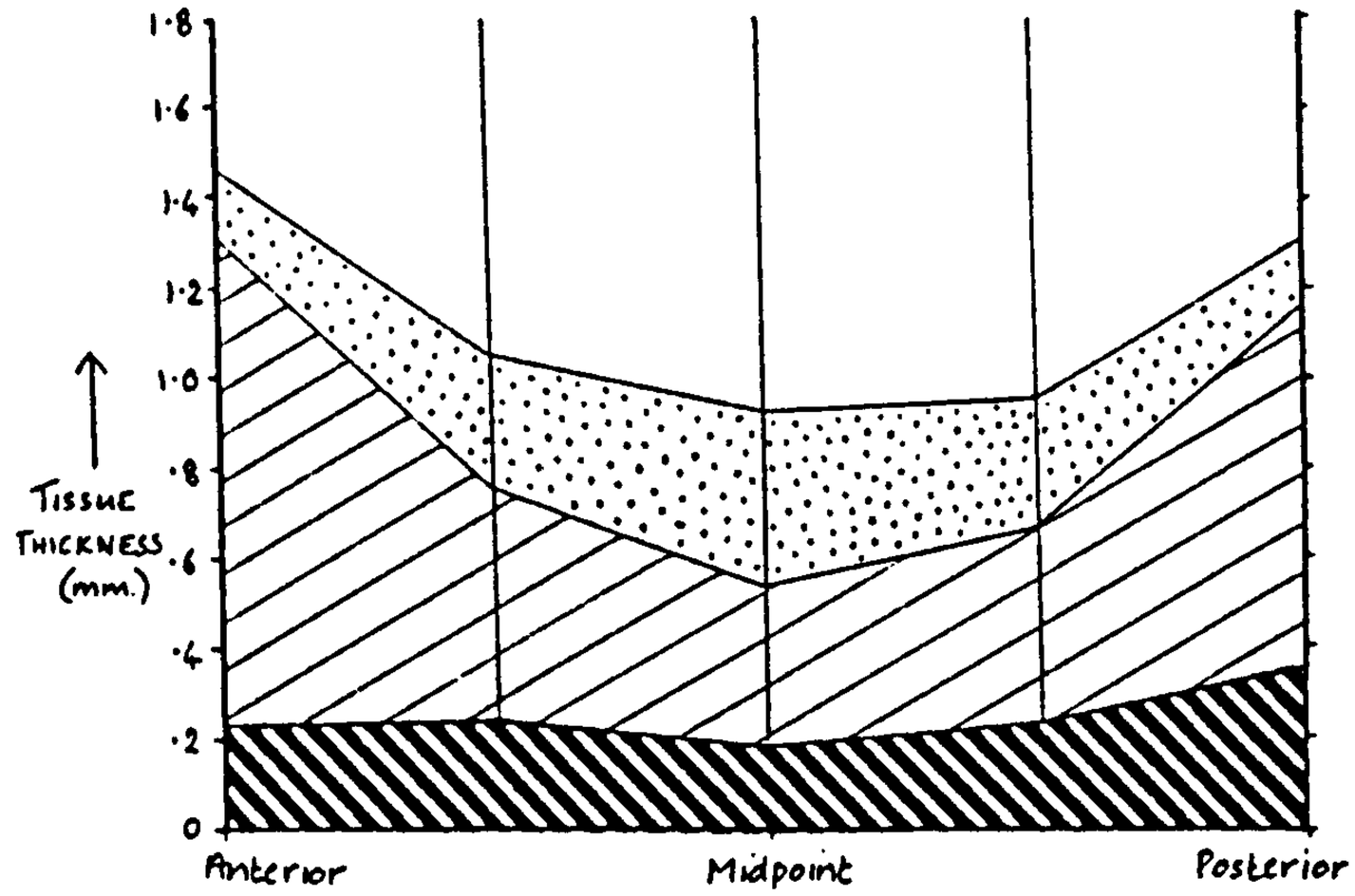


Figure 4a. A graphic representation of tissue thickness variation along the glottal edge of the ligamental portion of the vocal fold; subjects of 20-29 years (- using data given in Hirano et al. 1982:274)



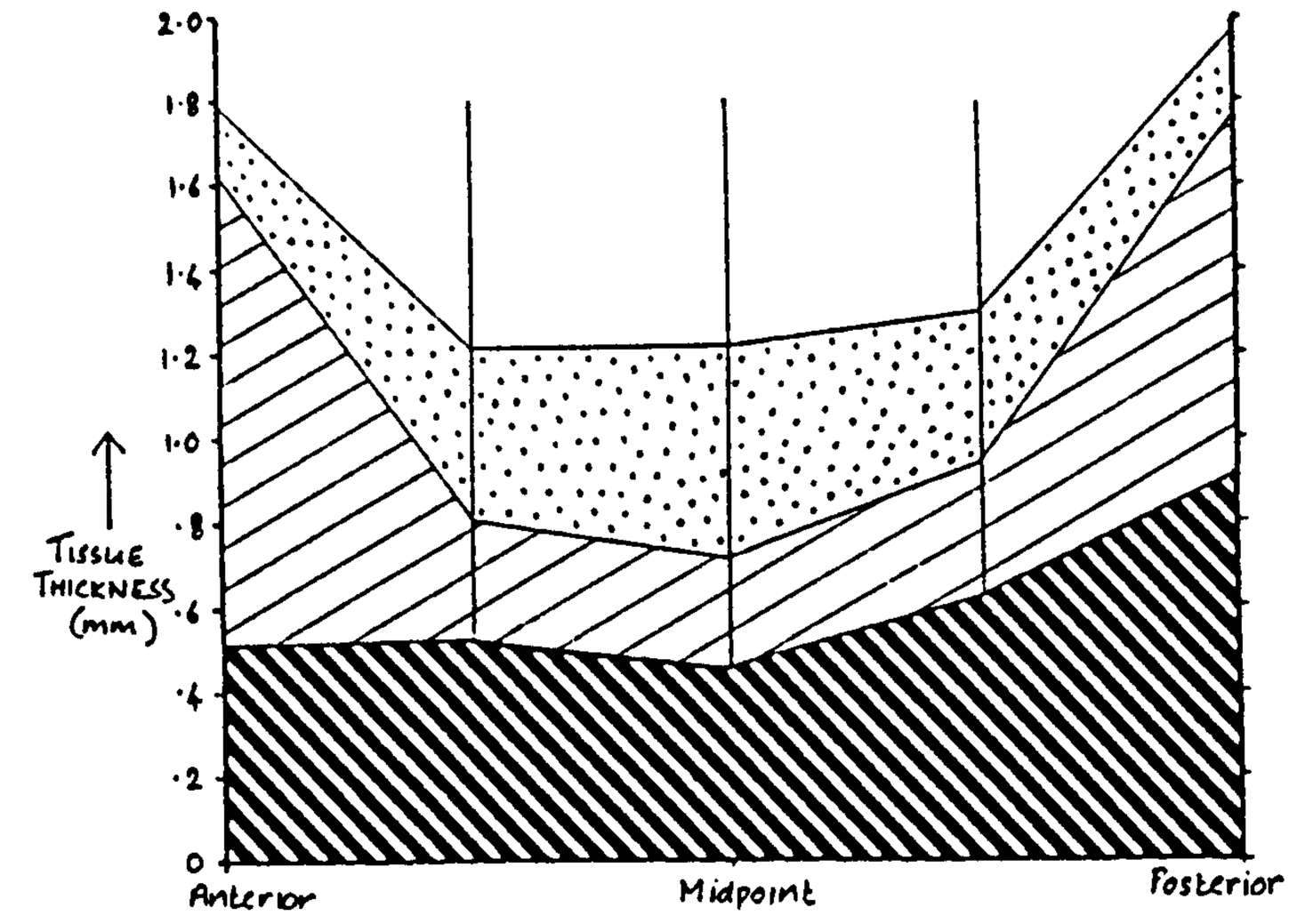


Figure 4b. A graphic representation of tissue thickness variation along the glottal edge of the ligamental portion of the vocal fold; subjects of 50-59 years (- using data given in Hirano et al. 1982:274)

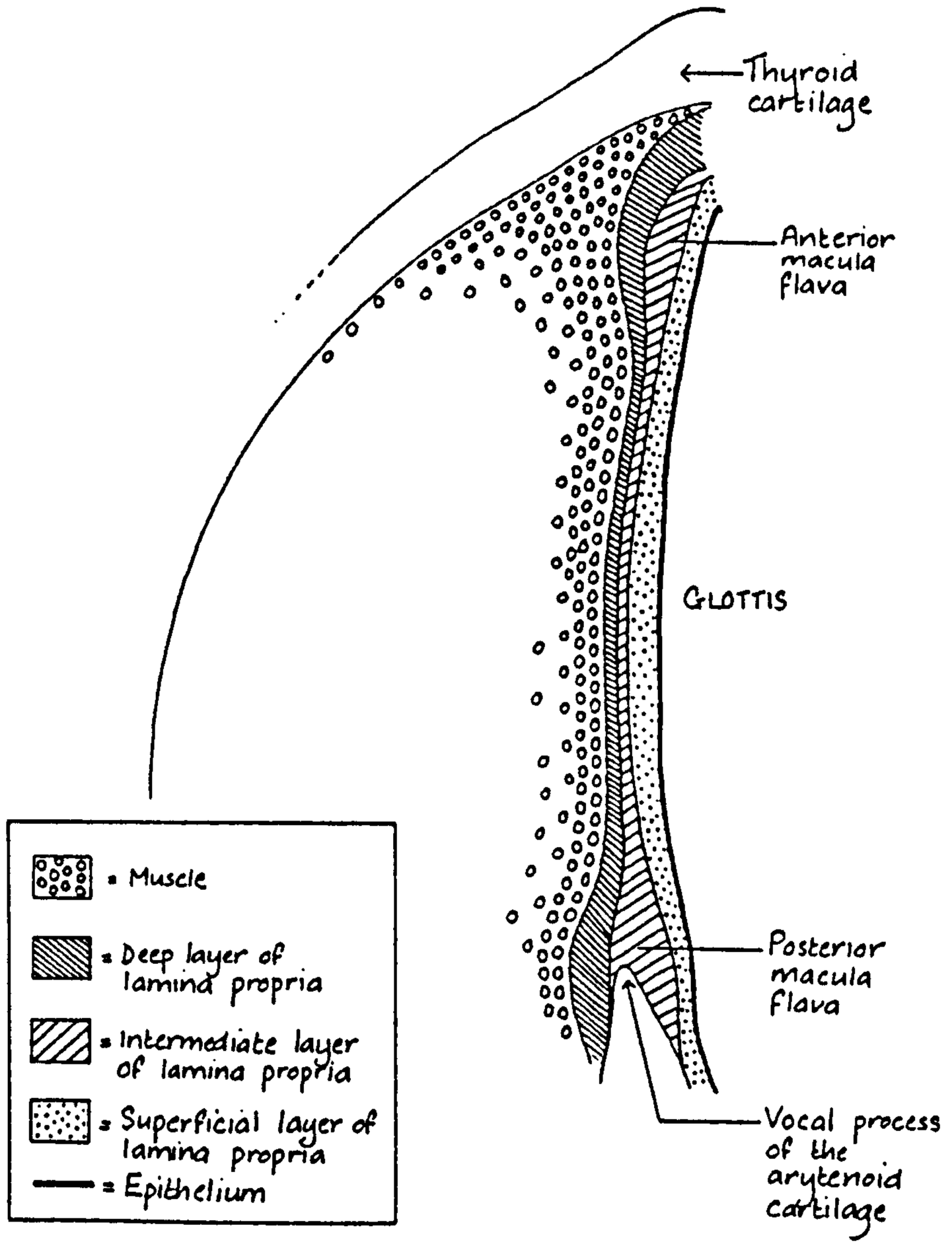


Figure 5. A diagram of the vocal fold in horizontal section, to show the maculae flavae.

examination of a larger sample of larynges, because this pattern of generalized thickening of the cover corresponds very closely to clinical descriptions of Reinke's Oedema (see Appendix). If Figure 4 reflects a widespread trend, then it may be that some degree of Reinke's Oedema is a not uncommon feature of aging, especially in males.

1.4. A summary of the mechanical properties of vocal fold tissue

Given the preceding description of vocal fold structure, it is now possible to summarize the mechanical properties of each tissue type, and to consider how they might interact during vibration.

Figure 6 summarises the tissue properties which have already been discussed. "Tensile stiffness" is used in this context as an indication of elasticity - i.e. tensile stiffness means the stress required to stretch a tissue sample of given cross section by a given amount. All tissues are assumed to be incompressible.

TISSUE LAYER		ANISOTROPY		TENSILE STIFFNESS	
		Canine	Human	Canine	Human
EPITHELIUM		-	-	high*	high
LAMINA PROPRIA	SUPERFICIAL	+*	-	moderate*	(fluid)
	INTERMEDIATE		+		low
	DEEP		+		high
VOCALIS MUSCLE		+*	+	LOW *(relaxed) → HIGH	LOW (relaxed) → HIGH

*Indicates an entry based on experimental evidence from canine tissue. Remaining entries are based on information about histological structure of the tissues, or on reports of tissue behaviour during vocal fold vibration.

Figure 6. A summary of the mechanical properties of vocal fold tissue.

1.5. Independence of tissue layers

The picture so far is of a structure with clearly defined layers, separated from each other by well marked boundaries, but this is something of an oversimplification. The extent to which tissue layers are actually differentiated and kept separate from one another has important implications of two kinds. Firstly it is relevant to the mechanical independence of each layer. Secondly it is relevant to the ease of spread of pathological change from one layer to another.

1.5.1. Mechanical implications

It is a reasonable assumption that two tissue layers are more likely to behave independently of one another during vocal fold

vibration if they fulfil two basic criteria:-

a. They should exhibit clearly different mechanical properties.

b. There should be a rapid transition of mechanical properties at the border between the two tissues.

The mechanical properties of each tissue have already been outlined, and it can be seen that each of the five tissue layers differs from its neighbours in at least one mechanical parameter. The question of transitions between the tissue types now needs to be addressed.

i) Epithelium / lamina propria

The basement membrane of the epithelium forms a well defined boundary between the tightly packed cells of the epithelium and the gelatinous superficial layer of the lamina propria, so that both the suggested criteria for mechanical independence are fulfilled. The fluid nature of this layer of the lamina propria has already been mentioned. Titze (1973) suggests that, because the epithelium is relatively thin, these two layers do in fact act in concert, with the epithelium mimicking the effect of a high surface tension.

ii) Superficial / intermediate layers of the lamina propria

Hirano et al. (1982:274) report that there is generally a clearly marked and rapid transition between these two layers. There is a very dramatic difference in mechanical properties between the fluid or semi-fluid areolar tissue of the superficial layer and the much denser, anisotropic elastic tissue of the intermediate layer. A fairly high degree of independence may therefore be expected.

iii) Intermediate / deep layers of the lamina propria

In the same study Hirano et al. found that the transition from elastic to collagen tissue is not so well defined. There is a gradual transition, with an intervening area where collagen and elastic fibres occur in equal numbers. In spite of their very different mechanical properties these two layers are not, therefore, likely to act truly independently.

iv) Deep layer of the lamina propria / vocalis muscle

Skeletal muscles are typically contained within connective tissue sheaths (epimysia)(Freeman and Bracegirdle 1967), and the muscle tissue is thus clearly delimited and separated from the lamina propria. The degree of disparity in mechanical properties of collagen and muscle tissue depends on the contractile state of the muscle. The mechanical properties of the collagen tissue are relatively invariable, but the tensile stiffness of the muscle may vary as much as tenfold. It is probable that under at least some conditions of muscular contraction these two tissue layers are sufficiently different to act with a degree of independence.

Many researchers have noted that a travelling wave can be observed on the surface of the vibrating vocal fold (Farnsworth 1940,

Smith 1956, van den Berg, Vennard, Berger and Shervanian 1960, Perello 1962, Hiroto 1966, Matsushita 1969, Baer 1973, Hirano 1975, Titze and Strong 1975, Broad 1977). This ripple-like mucosal wave can be taken as illustration of the fact that at least the outer two layers of the vocal fold (the fluid-like superficial layer of the lamina propria and the epithelium) are acting relatively independently of the deeper tissues.

It may be useful to examine some approaches to mathematical modelling of vocal fold vibration in the light of the above comments on tissue mechanics. Workers in this field have been conscious for some time of the need to consider at least two semi-independent masses when modelling cross sectional movement of the fold (Ishizaka and Flanagan 1972, Titze 1973, 1974). Ishizaka and Flanagan (1972: 1235) comment that 'a two-mass approximation can account for most of the relevant glottal detail, including phase differences of upper and lower edges'. Titze's model further subdivides the mass of each vocal fold into eight individual sections (see Figure 7). One of the suggested virtues of the sixteen mass model is that it allows simulation of longitudinal variations in mass and stiffness, and so can simulate some of the effects of vocal fold pathologies. The shortcoming of both the Ishizaka and Flanagan and the Titze models it that they are not capable of separating abnormalities arising in different layers of the mucosa, because all the different mucosal tissue layers are represented within a single mass. In a later paper Titze and Strong (1975) do, indeed, conclude that a more accurate model would require at least three masses in cross section.

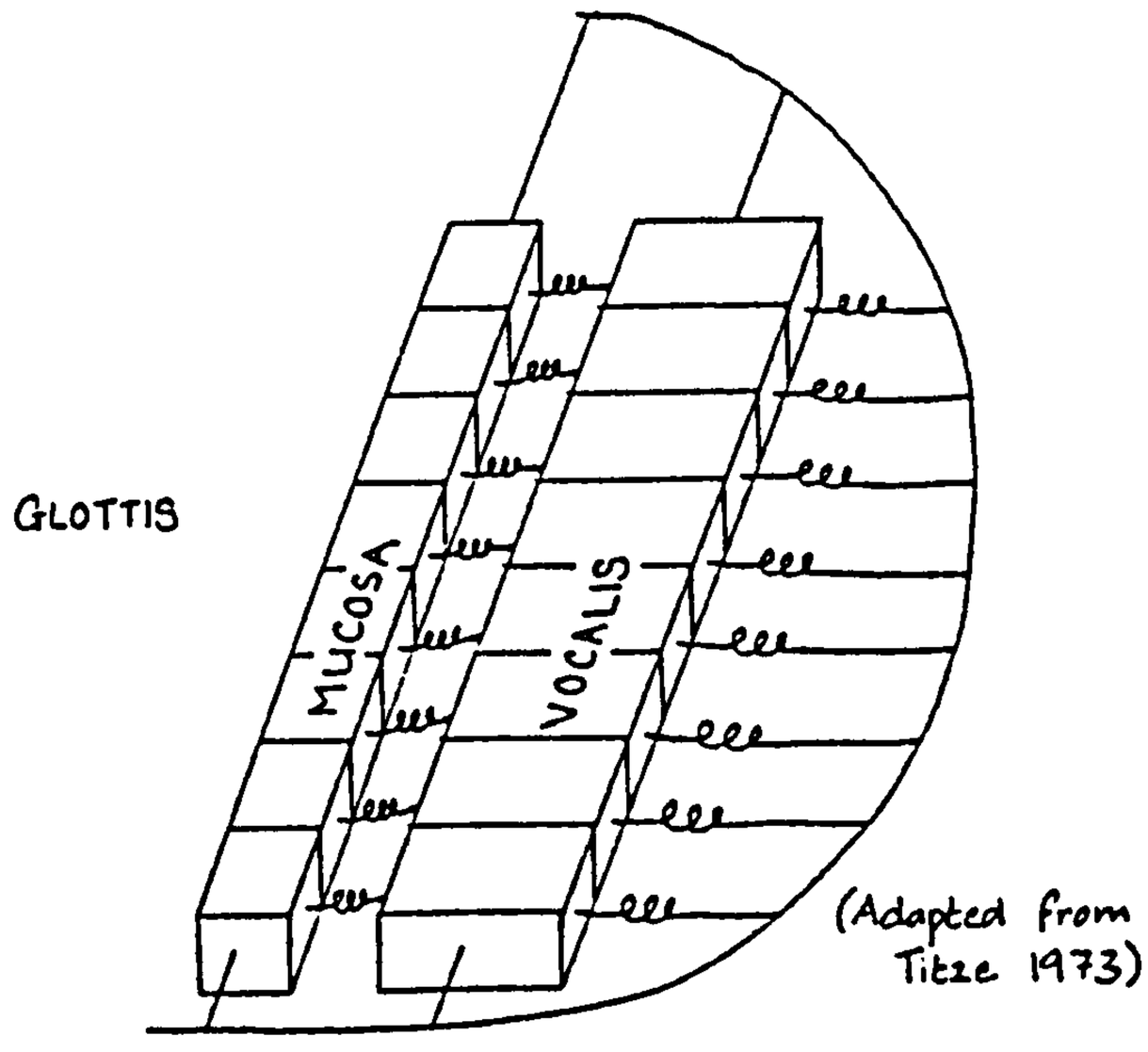


Figure 7. A diagram of Titze's (1973,1974) sixteen-mass model of vocal fold vibration.

PAGE

MISSING

IN

ORIGINAL

B. STRUCTURAL FACTORS LIKELY TO BE IMPORTANT IN PREDICTING VIBRATORY AND ACOUSTIC EFFECTS OF PATHOLOGIES

Given the structural framework outlined above, we can begin to suggest factors that are likely to be influential in determining the effect of structural change on vibration. We shall look first at the different types of change of tissue consistency and distribution that can occur within a tissue layer, and then at changes of tissue geometry that affect the spatial relationships between different tissue layers. We shall then discuss changes in the physical parameters of rigidity/flexibility, tensile stiffness/elasticity, mass and symmetry, and their consequences for acoustic parameters.

1. Changes of tissue consistency and internal distribution within a layer

The consistency of a tissue layer can change in a number of ways. One particular instance is inflammation. This is described in more detail in the Appendix, but, in brief, inflammation can involve capillary dilation, an infusion of white blood cells, collection of oedematous fluid in the interceullar space, a proliferation of collagen fibres and granulation tissue, and the deposition of hyaline. Another instance is keratinization (described above and in the Appendix), where, in the skin-forming process, the epithelium becomes stiffened by the deposition of keratin.

Changes in the distribution of cells within a tissue layer include processes such as hyperplasia (see Appendix), where a multiplication of cell numbers results in a thickening of the layer, often with a folding, buckling consequence for the overall layer. The density of cell distribution can also change, in that oedematous fluid collection in the intercellular space can cause an effective decrease in both cell and fibre density. Fibre density can also increase, in fibrosis.

2. Changes of the geometrical relationship between tissue layers

Three kinds of disruption of the geometrical relationship between two tissue layers can be described. The first is one involving the intrusion of one layer into another, where invasion is achieved by displacement. This is characteristic of disorders such as verucous carcinoma and sessile polyps (see Appendix for further descriptions of these). The second involves invasion by infiltration, where cells of the first tissue intermingle with those of the second. This happens in squamous cell carcinoma (see Appendix). The third is a disruption of the geometrical relationship between the two layers by material from one layer penetrating the frontier of the other to form a narrow-necked extrusion. This is found in disorders such as papilloma and pedunculated polyps (see Appendix).

3. Changes in physical parameters, and their acoustic consequences

A survey of various models of vocal fold vibration (Ishizaka and Flanagan 1972, Titze 1973, 1974, Hirano et al. 1982) suggest various factors which should be considered. These are:

i) Rigidity/flexibility
ii) Tensile stiffness/elasticity
iii) Mass
iv) Symmetry

Rigidity (i.e. resistance to bending) and tensile stiffness (i.e. resistance to stretching) can both for convenience be included under the general concept of 'stiffness'. This seems to follow Hirano's (1981:52) undefined usage of the term 'stiffness', when referring to visual examination of the vocal folds.

A further factor which can influence the acoustic output is the degree of approximation of the vocal folds, since under certain conditions of airflow inadequate approximation may induce turbulence. This will be seen in the acoustic signal as interharmonic energy (Laver 1980:121).

Boone (1977:47) organizes voice disorders according to mass/size changes and approximation changes, but these two factors alone allow only a rather vague prediction of phonatory quality. We will try to expand this approach to classifying organic vocal fold pathologies by taking the following criteria into account:

i) In which tissue layers are there structural alterations?
ii) Do these alterations involve a significant change in mass?
iii) Do they involve a significant change in stiffness?
iv) Is there a protrusion of any mass into the glottal space, so as to interfere with vocal fold approximation, or to cause turbulent airflow?
v) Is the structural alteration symmetrical, affecting both folds equally?
vi) Are the normal geometric relationships between the different tissue layers maintained?

Structural changes of the above sort will have a number of consequences for phonatory parameters. Hirano (1981:52-53) mentions some of these in his comments on the interpretation of strobolaryngoscopic examination. His guidelines can be briefly summarized:

a. Increased mass tends to decrease fundamental frequency and amplitude.
b. Increased stiffness tends to increase fundamental frequency, decrease amplitude, and prevent full approximation of the vocal folds. It also inhibits the action of the mucosal wave.
c. Localized protrusion of any mass into the glottal space will interfere with approximation of the vocal folds.
d. Asymmetry of mass, configuration or consistency will cause dysperiodic vibration, as will any localized mass or stiffness change.

The rationale underlying these guidelines deserves some consideration.

a. Mass

An increase in mass adds inertial force to the vocal fold, which will tend to decrease the speed of oscillation. It may be expected to exert its effect most strongly at the onset of phonation, when the vocal fold is accelerating from a relatively stationary position. The influence of mass on amplitude of vibration is less straightforward, and it should be noted that Hirano, Gould, Lambiase and Kakita (1981) contradict the above guideline, where they suggest that a larger mass should increase amplitude and speed of vocal fold excursion. Oedematous increases in mass, as associated for example with chronic laryngeal inflammation, should actually be expected to show a lower fundamental frequency. Fritzell et al. (1982) demonstrate that this is in fact the case.

The detailed location of any increase in mass needs also to be taken into account. A local increase in mass will have the greatest inertial contribution to vocal displacement when it is close to the point of maximum excursion - i.e. close to the longitudinal midpoint, and near the surface of the fold.

b. Stiffness

It is reasonable to expect that increasing the stiffness of a vibrating body should inhibit the vibratory movement, causing a decrease in amplitude of excursion. The mucosal wave, which is visible during normal vocal fold vibration, is a travelling wave in the mucosal layer. This presumably depends on having a semi-fluid superficial layer of the lamina propria behaving relatively independently of the deeper tissues. Increased stiffness of this layer, or of the epithelial layer (as in keratosis), should therefore limit the mucosal wave. Changes in stiffness of the underlying tissues would not necessarily have the same effect.

c. Protrusion

Protrusion of a mass into the glottal space will only interfere with vocal fold approximation if it is relatively localized. A uniform swelling along the full length of a vocal fold may actually improve approximation, as seems to be the case in some speakers with mild inflammation of the vocal folds during upper respiratory tract infections. A distinction must therefore be drawn between localized and non-localized protrusions. An example of localized protrusion is a vocal polyp, which may become wedged between the vocal folds, thus preventing the folds from meeting.

In considering localized protrusions, the site and attachment of the protruding body need also to be taken into account. Pedunculated polyps and papillomata, because of their flexible, stalk-like attachments, may be displaced by the transglottal airflow, causing only intermittent obstruction.

d. Asymmetry

Asymmetry of vocal fold structure may cause the two vibrating folds to move out of phase with each other, with complex consequences for the acoustic waveform. This discrepancy will disrupt the fine co-ordination between airflow and vocal fold configuration,

causing perturbation of the laryngeal waveform. Structural asymmetry is a feature of many laryngeal pathologies, including carcinoma, vocal polyps and papillomata.

Tissue layer integrity

In addition to the above comments, there are considerations about the integrity of tissue layers to be taken into account. A degree of independent behaviour of the body and covering tissues is important in determining the fine detail of phonatory vibration (Smith, 1961, Perello, 1962). Any loss of integrity between the tissue layers can therefore be expected to affect vibratory patterns by changing this relative independence.

C. PROPOSED TYPOLOGY OF VOCAL FOLD DISORDERS

It has already been mentioned that our immediate concern is with disorders of the true vocal fold, since these are the most likely to have direct consequences for vibration. Sub-glottic and supra-glottic disorders are not considered here. The scope of this typology is further limited by excluding all disorders which are specific to childhood. The reasons for this are two-fold. The first relates simply to the needs of the present project, which will use speech samples drawn largely, if not exclusively, from the adult population. The second reason is that, as mentioned earlier, the mature layered structure of the larynx is not fully developed until after puberty.

The proposed system of classification is outlined in Figures 8a and 8b. This is not intended to be a definitive solution to the problem of devising a phonatory classification of organic vocal fold pathology. It should be seen rather as a preliminary attempt to highlight some of the mechanical factors which must be considered in order to predict the vibratory characteristics of any disorder. The structural vocal fold pathologies which are most commonly described in the literature are listed in Figure 9, and each is given a classificatory code which corresponds with the codes in Figures 8a/b. Brief descriptions of these disorders are included in the Appendix. This is by no means a complete list of all the disorders which involve structural changes of the larynx, but it will be used to give some idea of the possibilities and limitations of an acoustic screening procedure for detecting vocal fold pathologies.

Allocations of disorders to categories within this framework are often tentative, because it has not yet been possible to gather sufficient information about histological details for many of the disorders mentioned. It should also be stressed that such a framework does not always necessarily relate directly to medical and pathological considerations. For example, the differing structures and mechanical properties of vocal polyps and polypoid degeneration demand that they be given different classifications in this system. They may, however, both be seen as forms of chronic inflammatory reaction to chemical or mechanical irritation, thus sharing a common underlying pathology (Luchsinger and Arnold 1965, Boone, 1977, Aronson 1980).

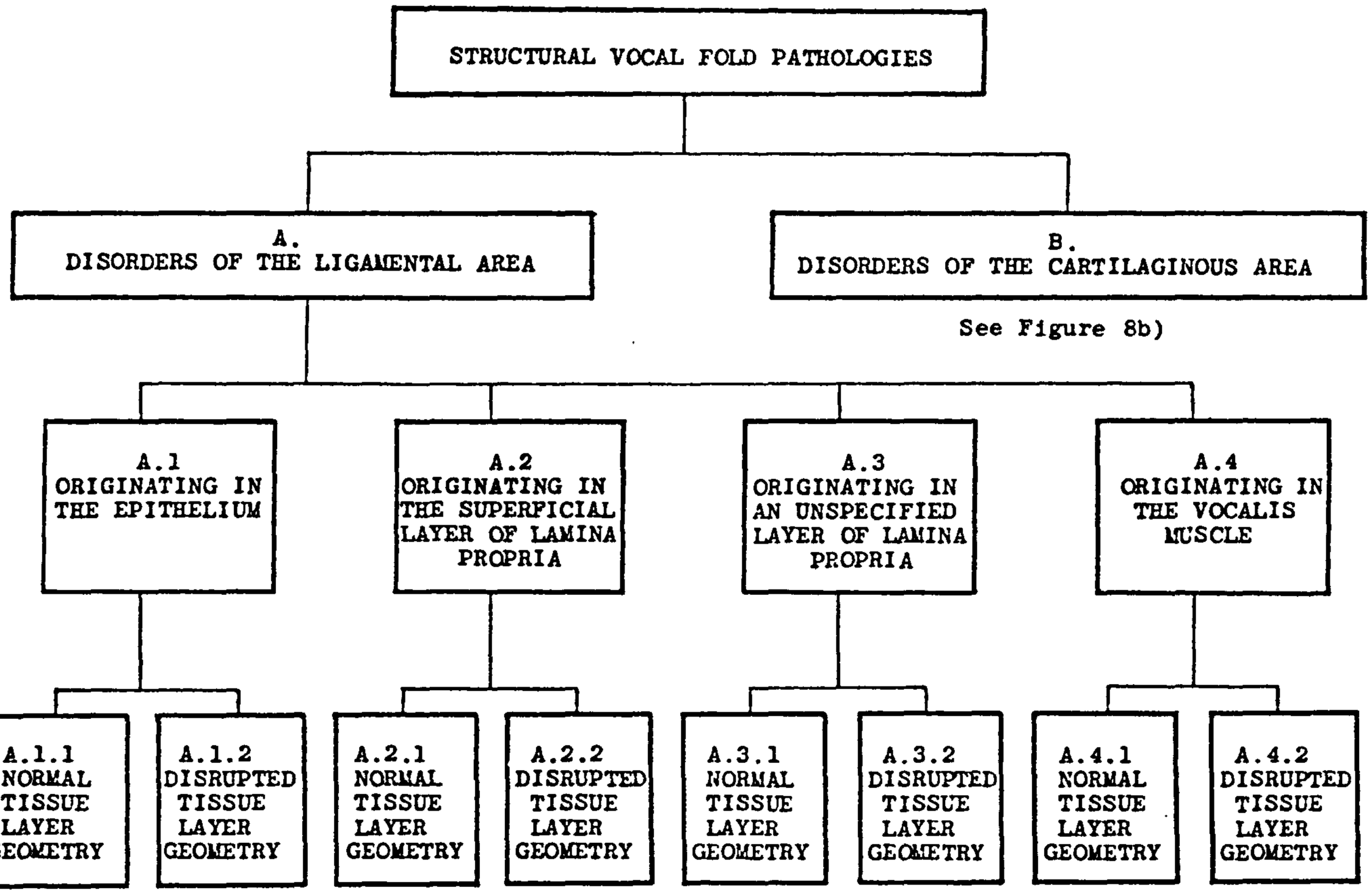


Figure 8A. The proposed system of classification. Disorders of the ligamental portion.

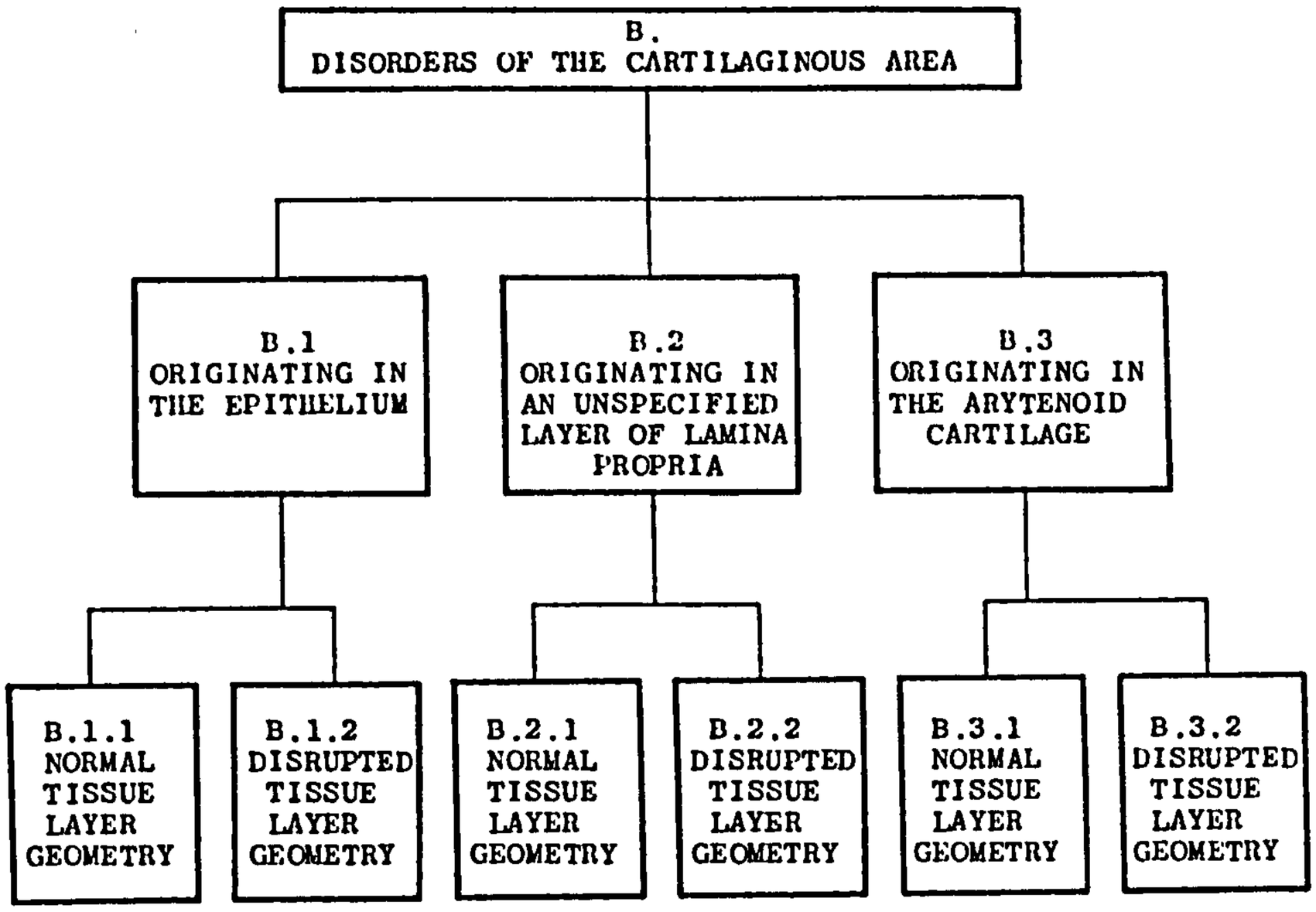


Figure 8b. The proposed system of classification.
Disorders of the cartilaginous portion.

The divisions laid out in Figure 8 are also to some extent over-specific, in that they imply a rather more orderly situation than exists in reality. Many disorders show so much variation in form, in different individuals and at different stages of their development, that they could have been allocated to more than one category. The proposed framework imposes somewhat artificial boundaries in these cases, but the allocation to categories attempts to reflect the most characteristic form of each disorder.

The combination of disorders originating in any of the three separate layers of the lamina propria into one overall category, as in categories A3 and B2, is suggested because medical writers are often not specific about which layers are involved in a structural change. It may be that such distinctions, because of a lack of a biological barrier between the layers, are of no direct medical relevance, even though there may be possible consequences for details of vibratory pattern. Further examination of individual cases may allow a more detailed categorization.

Figure 10 summarises pathologies in terms of the presence or absence of mass and stiffness changes, protrusion into the glottal space, symmetry, and tissue layer geometry. An important point emerges from this, concerning the potential power of acoustic screening to differentiate between disorders. Some clinically separable disorders may be expected to impose rather similar mechanical constraints on vibration, and hence on acoustic output, so that they are

A. Disorders of the ligamental portion

A.1. Disorders originating in the epithelium

A.1.1. Normal tissue layer geometry

Hyperplasia
Keratosis
Carcinoma-in-situ

A.1.2. Disrupted tissue layer geometry

Squamous cell carcinoma
Verrucous carcinoma (a specific form of squamous cell carcinoma)

A.2. Disorders originating in the superficial layer of the lamina propria

A.2.1. Normal tissue layer geometry

Reinke's oedema

A.3. Disorders originating in any unspecified layer of the lamina propria

A.3.1. Normal tissue layer geometry

Vocal nodules
Sessile vocal polyps
Acute laryngitis
Chronic laryngitis
Chronic hyperplastic laryngitis
Fibroma

A.3.2. Disrupted tissue layer geometry

Pedunculated polyp

A.4. Disorders originating in the vocalis muscle

A.4.1. Normal tissue layer relationships

Sarcoma

B. Disorders of the cartilaginous portion

B.2. Disorders originating in any unspecified layer of the lamina propria

B.2.1. Normal tissue layer geometry

Acute oedema

B.2.2. Disrupted tissue layer geometry

Contact ulcer

Figure 9. A list of structural vocal fold pathologies, arranged according to the classification system outlined in Figure 10.

PATHOLOGY	Disrupted tissue layer geometry	Mass change	Stiff-ness change	Protru-sion	Asymmetry
A. LIGAMENTAL PORTION A.1. EPITHELIAL					
A.1.1. Hyperplasia		+			+
A.1.1. Keratosis		(+)	+	(+)	+
A.1.1. Carcinoma-in-situ		+	+	(+)	+
A.1.2. Squamous carcinoma	+	+	+	+	+
A.1.2. Verrucous carcinoma	+	+	+	+	+
A.1.2. Adult papilloma	+	+	+	+	+
A.2. SUPERFICIAL L.P.*					
A.2.1. Reinke's oedema		+		N.L.	
A.3. UNSPECIFIED L.P.*					
A.3.1. Vocal nodules		+		+	(+)
A.3.1. Vocal polyps (sessile)	(+)	+	+	+	(+)
A.3.1. Acute laryngitis		+		N.L.	
A.3.1. Chronic laryngitis		+		N.L.	
A.3.1. Chronic hyper-plastic laryngitis		+	+	N.L.	
A.3.1. Fibrome		+	+	+	+
A.3.2. Vocal polyps (pedunculated)	+	+	+	+	(+)
A.4. VOCALIS MUSCLE					
A.4.1. Sarcoma		+	?		+
B. CARTILAGINOUS PORTION B.1. EPITHELIAL (as under A.1.) B.2. UNSPECIFIED L.P.*					
B.2.1. Acute oedema		+		+	
B.2.2. Contact ulcer	+	+	+	+	(+)

L.P.* = lamina propria
\+ = presence of a factor
(+) = possible or variable presence
N.L. = non-localised protrusion, not expected to prevent vocal fold approximation

Figure 10. A summary of mechanical characteristics of vocal fold pathologies.

unlikely to be separable by a solely acoustic assessment procedure. An example of this is the grouping of papilloma, squamous carcinoma, and verrucous carcinoma, all of which may show an asymmetric increase in mass and stiffness originating in the epithelium, with protrusion into the glottis and altered tissue layer geometry.

D. CONCLUSION

The three broad objectives of the project were described at the beginning of this article as being the development of an automatic screening system, the acoustic differentiation of various laryngeal pathologies, and the acoustic assessment of longitudinal change in a subject's voice. The second of these goals is perhaps the most challenging of the three. If progress is to be made towards the ability to discriminate acoustically between different vocal pathologies, then we need to have a better understanding of the relationships between the diagnosis of pathology, the structural status of the vocal folds, the mechanics of their vibration, and the resulting acoustic output.

The relationships between these four areas are complex. There will seldom be one-to-one links to be traced between them, and this is particularly true of the link between diagnosis of pathology and the structural status of the vocal folds. A given pathology may to some extent show different structural attributes in different individuals, or at different stages of development. For example, carcinoma-in-situ may present either as a single localized area of increased epithelial thickness, or as a multifocal growth. In development in a given individual, it may progress from an area of simple thickening, with no significant alteration of stiffness, to a substantial protrusion of thickened tissue with a marked increase in stiffness due to keratinization. In addition, two growths with quite different medical diagnoses may share a similar pattern of structural attributes. For instance, a fibroma (see Appendix) and vocal polyps may both involve very similar changes in mass, stiffness and geometry.

Our next step will be to collect patient data, in order to explore in detail the nature of the relationships mentioned above. We are fortunate in benefiting from collaboration with hospitals in Oxford and Lothian. We plan to carry out computer-based acoustic analysis, as described in Miller, Laver and Mackenzie (1983), on tape-recordings of patients of known diagnostic status. Information about structural state and vibratory pattern will be provided by our collaborators, using fiberoptic examination of the larynx, stroboscopy and still photography, backed up by histological examination where appropriate. The hospitals involved are the Radcliffe Infirmary, Oxford, where our collaborators are Mr. T. Harris (Department of Otolaryngology) and Mrs. S. Collins, Department of Speech Therapy); the Royal Infirmary, Edinburgh (Mr. A. Maran, Department of Otolaryngology and Mrs. M. Mackintosh, Department of Speech Therapy); and Bangour General Hospital, West Lothian (Mr. W. Singh, Department of Otolaryngology).

We hope that analysis of this data will allow us to approach the objectives described above, by evaluating the acoustic consequences of structural abnormalities of the vocal folds.

APPENDIX

STRUCTURAL VOCAL FOLD PATHOLOGIES

This appendix includes brief notes on the individual vocal fold pathologies mentioned in the text.

Inflammation

Many of the disorders described below involve some degree of inflammation. This may play a major role in the development of a disorder, as in the various forms of chronic laryngitis, or it may occur as a secondary peripheral response, like that seen in the tissues adjacent to an advancing verrucous carcinoma (Ferlito 1974). The descriptions of specific pathologies may therefore be simplified if they are prefaced by a brief account of the basic characteristics of inflammatory reactions. More detailed accounts of inflammation can be found in, for example, Sandritter and Wartman (1969:20-27).

Inflammation is a complex, coordinated response to tissue damage, which acts to limit infection and to repair tissue. It is convenient to view the response as a two-stage process.

a) The acute stage

The acute stage of inflammation can be thought of as an emergency reaction, which marshalls together the elements necessary for defence and repair. This stage exhibits certain common features, regardless of the size, site or type of injury. The three predominant signs are listed below.

i) Hyperaemia.
This simply describes an increase in blood flow to the area, which is usually achieved by capillary dilation.

ii) Leucocyte infiltration.
The capillaries become more permeable and allow white blood cells (leucocytes) to pass into the affected tissue. Some of these cells are active in limiting infection, by engulfing foreign bodies, or by antibody production.

iii) Swelling due to fluid exudation (oedema).
Fluid also passes out of the dilated capillaries and collects in the intercellular spaces of the tissue.

b) The chronic stage

The chronic stage of inflammation follows a more variable course, depending on the extent, duration and type of damage. Necrotic (dead) tissue and blood clots are resorbed by specialised cells, and the damaged area may be localised and walled off by the deposition of collagen fibres (fibrosis). Active repair of damaged tissue is brought about by the proliferation of new connective tissue and blood vessels. This proliferative repair tissue is often known as granulation tissue, but its exact morphology may vary considerably. In some cases fibrosis may predominate, with a progressive increase in collagen density, and eventually hyaline may also be deposited in the fibrosed tissue. Hyaline is the firm, glassy substance which forms

the matrix of some cartilages, so that this type of granulation tissue will form areas of greatly increased stiffness. Other cases may show no sign of fibrosis, but have a marked growth of capillaries. Wherever possible in the following notes the precise nature of the inflammatory response will be specified, but most often the literature simply mentions "inflammation", with no comment on the relative contributions of fibrosis, capillary proliferation, etc..

A. DISORDERS OF THE LIGAMENTAL AREA OF THE VOCAL FOLD

A.1. Disorders originating in the epithelium

Terminology:

A survey of epithelial disorders is complicated by the lack of a standardised terminology to describe some common types of structural disorder within the epithelium. The terms "hyperplasia", "keratosis", "hyperkeratosis" and "leucoplakia" seem all to be applied to a rather similar group of epithelial conditions which are thought to be aggravated by prolonged mechanical or chemical irritation. The common link between these conditions is the presence, in varying balance, of two types of structural change. The first, which we shall call hyperplasia, is a simple increase in cell number resulting from excessive cell division. The second, keratosis, is the formation of keratin. These two processes are described as separate disorders below, but they do commonly occur in combination. It is assumed that, alone or in varying combination, hyperplasia and keratosis cover all the labels listed at the beginning of this section.

There is considerable controversy over the question of whether or not these conditions should be considered as precursors of malignant change. As long as individual cells appear to have normal structure there is no evidence of malignancy, but there does seem to be a continuum from simple hyperplasia and keratosis, where cell structure is normal, to carcinoma-in-situ, where a large proportion of the epithelial cells are abnormal in structure and malignancy must be suspected. Differential diagnosis is therefore often highly problematic. (Saunders 1964, Hall and Colman 1975, Birrell 1977, Friedmann and Osborn 1978).

A.1.1. Hyperplasia

Tissue of origin: Epithelium.

Mechanical factors: An asymmetric increase in mass, with normal tissue layer geometry.

Site of occurrence: Anywhere within the laryngeal epithelium. Common at the centre of the ligamental area of the vocal fold.

Hyperplasia is an increase in cell number resulting from rapid division of the basal cell layer. The increase in basal cell number may cause buckling and distortion of the basement membrane, but the stratified arrangement of cells is maintained, and the cells appear normal.

A.1.1. Keratosis

Tissue of origin: Epithelium.

Mechanical factors: An asymmetric increase in stiffness, with normal tissue layer geometry. Eventually there may be a significant increase in mass and protrusion into the glottal space.

Site of occurrence: As for hyperplasia.

Keratosis is a condition in which the squamous cells of the epithelium begin to produce keratin, which is laid down as a horny layer at the surface of the epithelium. It may form a large, whitish mass, which protrudes into the glottal space and may interfere with vocal fold approximation. Smoking seems to be a major aetiological factor in the development of keratosis. (Auerbach, Hammond and Garfinkel 1970).

A.1.1. Carcinoma-in-situ (Intra- epithelial carcinoma)

Tissue of origin: Epithelium.

Mechanical factors: An asymmetrical increase in mass, with normal tissue layer geometry. Variable increase in stiffness and protrusion into the glottal space.

Site of occurrence: Anywhere within the laryngeal epithelium.

Carcinoma-in-situ is usually regarded as the earliest recognisable stage of cancer of the larynx, although it is not an inevitable precursor of invasive cancer, and not all cases of carcinoma-in-situ necessarily progress to become fully invasive. The difficulty of making a differential diagnosis between simple hyperplasia, keratosis, and carcinoma-in-situ has already been mentioned. This is because carcinoma-in-situ always involves hyperplasia and it may also co-occur with some degree of keratosis. The feature which sets carcinoma-in-situ apart, and which indicates the onset of malignancy, is the presence of a high proportion of abnormal cells and the loss of the normal orderly arrangement of cells within the epithelium. This disorder thus displays a histological pattern of haphazardly dividing cells which may have quite bizarre structure. The abnormality spreads laterally within the epithelium, but the basement membrane seems to act as a barrier, preventing spread into the lamina propria. The lamina propria may, however, be inflamed. (Auerbach, Hammond and Garfinkel 1970, Bauer and McGavran 1972, Ferlito 1974, Friedmann and Osborn 1978).

A.1.2. Squamous cell carcinoma

Tissue of origin: Epithelium.

Mechanical factors: An asymmetrical change in mass and stiffness, with disrupted tissue layer geometry and protrusion into the glottal space.

Site of occurrence: Anywhere within the larynx. Most common in the ligamental portion of the vocal fold.

The commonest type of laryngeal tumour is carcinoma arising in the squamous epithelium. Carcinomatous change is characterized by a

loss of the normal control of epithelial cell division. The epithelial cells divide at an abnormal rate, and form a disorderly mass. The cells are recognized as being malignant by their abnormal structure, and by their tendency to infiltrate not just the surrounding epithelial tissue, but also the underlying tissues. Squamous cell carcinomas vary greatly in their structure, and in their pattern of invasion, so that it is difficult to generalise about their expected mechanical correlates. An increase in mass is almost always found, except in those cases with ulceration. Ulceration may occasionally expose and destroy even the laryngeal cartilages, so that a considerable amount of tissue is lost. Stiffness depends on cell density and on the degree of keratinization, both of which are very variable. The size of the lesion may also fall within a wide range. Some specific forms of squamous carcinoma are recognized, one of which is described below (verrucous carcinoma). (Ferlito 1974, Michaels 1976, Friedmann and Osborn 1978, Shaw 1979).

A.1.2. Verrucous carcinoma (a specific type of squamous carcinoma)

Tissue of origin: Epithelium.

Mechanical factors: An asymmetrical increase in stiffness and mass, with localised protrusion into the glottal space and disrupted tissue geometry.

Site of occurrence: Anywhere within the larynx. Commonest in the ligamental portion of the vocal fold.

This tumour is a specific type of squamous cell carcinoma, which presents as a slowly growing warty mass, and may be multicentric. The epithelium becomes hyperplastic and highly keratinized, with folds and finger-like protrusions extending deep into the lamina propria. Epithelial pearls (dense deposits of keratin) may develop, forming localised areas of extreme stiffness. Verrucous carcinoma is of relatively low malignancy, and advances by displacement of cells rather than by infiltration. Adjacent tissue usually shows a marked inflammatory response. The tumour may grow large enough to cause dysphagia (swallowing difficulty) and respiratory obstruction. (Ferlito 1974, Biller and Bergman 1975, Michaels 1976, Friedmann and Osborn 1978, Maw et al. 1982).

A.1.2. Adult papilloma

Tissue of origin: Epithelium.

Mechanical factors: An asymmetrical increase in mass and stiffness, with disrupted tissue layer geometry and localised protrusion into the glottal space.

Site of occurrence: Commonest at the edge of the ligamental portion of the vocal fold or at the anterior commissure.

Papilloma is a benign warty tumour, which, in adults, forms multiple branchlike projections of highly keratinized epithelium. There may be extrusion of thin columns of lamina propria into the tumour, so that tissue geometry is substantially disrupted. Papillomata are usually unilateral and solitary, and most are pedunculated. These growths are not common in adults, and their medical significance derives from reports that a small proportion of papillomata undergo

malignant transformation. (Hall and Colman 1975, Birrell 1977, Friedmann and Osborn 1978, Shaw 1979).

A.2. Disorders originating in the superficial layer of the lamina propria

A.2.1. Reinke's oedema (Polypoid degeneration, Chronic oedematous laryngitis)

Tissue of origin: Superficial layer of the lamina propria.

Mechanical factors: A symmetrical mass increase with non-localised protrusion into the glottal space. Tissue layer geometry is normal but with weakened adherence between layers.

Site of occurrence: Both vocal folds are affected along their full length.

Reinke's oedema is a specific form of chronic laryngitis which is characterised by a loosening of the attachment between tissue layers in the ligamental portion of the vocal fold. This allows oedematous collection of fluid along the full length of the vocal fold. The overlying epithelium is normal, or only slightly hyperplastic, and if fluid is allowed to drain away the lamina propria appears to be relatively normal. Only in long-standing cases does mild hyperaemia occur. Reinke's oedema is a disorder of middle age, and seems to be exacerbated by alcohol and smoking. It is interesting that clinical descriptions of Reinke's oedema suggest similarities with the age related changes described by Hirano et al 1982, (see section on vocal ligament). One of the main vocal symptoms is a decrease in fundamental frequency. (Saunders 1964, Fuchsinger and Arnold 1965, Kleinsasser 1968, Saunders 1964, Birrell 1977, Friedmann and Osborn 1978, Salmon 1979, Aronson 1980, Fritzell, Sundberg and Strange-Ebbeson 1982).

A.3. Disorders originating in any unspecified layer of the lamina propria

A.3.1. Vocal nodules (early stage)

Tissue of origin: Lamina propria (probably the superficial layer).

Mechanical factors: A symmetrical or asymmetrical increase in mass, with localised protrusion into the glottal space and normal tissue layer geometry. Stiffness is increased only slightly.

Site of occurrence: Usually on the edge of the vocal fold in the centre of the ligamental portion.

Vocal nodule formation is thought usually to be precipitated by local mechanical trauma. The first stage is probably a haemorrhage of the small blood vessels within the lamina propria, which is followed by a localised inflammatory response. The nodules appear as small soft, red swellings, and they may be bilateral, at the centre of the ligamental section of each fold. Nodules may recover spontaneously if further mechanical abuse of the larynx is avoided. If they

become established fibrosis, epithelial hyperplasia or capillary proliferation may occur, creating a much firmer growth. There is some disagreement about the pathological relationship between vocal nodules and vocal polyps. Some writers consider polyps to be chronically established nodules which have undergone late stage inflammatory change, so the following section on polyps can be taken to represent a later stage in nodule development. (Arnold 1962, Luchsinger and Arnold 1965, Michaels 1976, Perkins 1977, Boone 1978, Friedmann and Osborn 1978, Salmon 1979, Aronson 1980).

A.3.1. Sessile vocal polyps

Vocal polyps may be sessile or pedunculated. Pedunculated polyps have disrupted tissue layer geometry, and must therefore be placed in the category A.3.2. (see Figure 9). Histological characteristics of both forms are, however, similar, so they will be discussed together below.

Tissue of origin: Lamina propria (probably the superficial layer).

Mechanical factors: An asymmetrical (or rarely symmetrical) increase in mass and stiffness, with localised protrusion into the glottal space. Tissue layer geometry is significantly disrupted only if the growth is pedunculated.

Site of occurrence: Usually at the edge of the ligamental portion of the vocal fold.

Long term mechanical abuse of the vocal folds may result in the establishment of localised chronic inflammatory changes. These appear as small, stiff swellings on the edge of the vocal fold, which may be unilateral or bilateral. In bilateral cases the polyps are seldom the same size, so that true symmetry will be rare. The extent and constancy of protrusion into the glottal space will vary, because polyps may be sessile or pedunculated. Stiffness depends on the histological make-up of each polyp. Some are predominantly fibrotic, with a dense, disorganized network of collagen fibres, and this type may eventually develop patches of hyalinization. Others are built up largely from vascular tissue, and may be much less stiff than the fibrotic type. The epithelium overlying a polyp may also become hyperplastic. (Arnold 1962, Luchsinger and Arnold 1965, Kleinsasser 1968, Greene 1972, Hall and Colman 1975, Michaels 1976, Birrell 1977, Boone 1977, Perkins 1977, Friedmann and Osborn 1978, Salmon 1979, Aronson 1980).

A.3.1. Acute laryngitis

Tissue of origin: Lamina propria.

Mechanical factors: A symmetrical increase in mass, with normal tissue layer geometry. Approximation may be limited by associated acute oedema affecting the cartilaginous area of the fold.

Site of occurrence: The whole of the larynx may be involved.

Acute laryngitis, which may have many causes, including infection, sudden irritation or mechanical abuse, shows all the features of a generalised acute inflammation. There is hyperaemia throughout the larynx, and infiltration of leucocytes, so that the vocal folds

appear to be rounded and thickened in cross section. The swelling due to oedema is usually most marked in the mucous covering the arytenoids (see section on acute oedema, B.2.1.), so that approximation of the ligamental area of the vocal folds may be prevented. In severe cases the epithelium may become necrotic, and ulceration results as the dead tissue is sloughed off. The underlying muscle may also become inflamed. (Luchsinger and Arnold 1965, Hall and Colman 1975, Boone 1977, Birrell 1977, Friedmann and Osborn 1978, Salmon 1979, Aronson 1980)

A.3.1. Chronic laryngitis

Tissue of origin: Lamina propria.

Mechanical factors: A symmetrical increase in mass, with non-localised protrusion into the glottal space, and normal tissue layer geometry.

Site of occurrence: The whole larynx may be involved.

Chronic inflammation of the larynx may be rather variable in form. The simplest presentation includes hyperaemia and swelling, with an increase in mucous secretions covering the folds, and in severe cases the inflammatory response may involve the vocalis muscle Chronic laryngitis may be a response to long-standing exposure to irritants such as dust or smoke, or to habitual Reinke's oedema and chronic hyperplastic laryngitis). (Saunders 1964, Hall and Colman 1975, Turner 1977, Friedmann and Osborn 1978, Aronson 1980)

A.3.1. Chronic hyperplastic laryngitis (Chronic hypertrophic laryngitis)

Tissue of origin: Lamina propria.

Mechanical factors: A symmetrical increase in mass and stiffness, with non-localised protrusion into the glottal space, and normal tissue layer geometry.

Site of occurrence: The whole larynx may be involved.

Some authors differentiate a type of chronic laryngitis which is characterized by a generalised hyperplasia of the epithelium, and in terms of mechanical factors it makes sense for us to follow this example. The vocal folds are swollen and hyperaemic, as in other forms of laryngitis, but this is associated with changes in the overlying epithelium. The ciliated epithelium above and below the vocal fold becomes hyperplastic, and takes on a squamous pattern, whilst the squamous epithelium at the edge of the vocal folds becomes keratinized. The vocal folds become progressively more irregular and swollen, and may appear very dry. (Kleinsasser 1968, Birrell 1977, Salmon 1979)

A.3.1. Fibroma

Tissue of origin: Lamina propria.

Mechanical factors: An asymmetrical increase in mass and stiffness, with localized protrusion into the glottis, but no significant disruption of tissue layer geometry.

Site of occurrence: Anywhere within the larynx. Commonest on the edge of the ligamental portion of the vocal fold.

This rare, benign tumour usually presents as a smooth, sessile body on the edge of the vocal fold. It contains a network of collagen fibres, and may be difficult to distinguish from a fibrous polyp. (Birrell 1977, Perkins 1977, Shaw 1979)

A.3.2. Pedunculated vocal polyp

See earlier section on vocal polyps (A.3.1).

A.4. Disorders originating in the body of the vocal fold

A.4.1. Sarcoma

Tissue of origin: Vocalis muscle or lamina propria.

Mechanical factors: An asymmetrical increase in mass.

Site of occurrence: Not specified.

Sarcoma is a very rare type of malignant tumour, which may affect connective tissue and muscle. Sarcoma arising from the vocalis muscle is one of the few disorders (excluding atrophy due to muscle paralysis) which originates in the body of the vocal fold. The rather brief comments in the references below allow only tentative suggestions about mechanical correlates. (Friedmann and Osborn 1978, Shaw 1979)

B. DISORDERS OF THE CARTILAGINOUS AREA OF THE VOCAL FOLD

B.1. Disorders originating in the epithelium

All of the epithelial disorders already described in the preceding section on disorders of the ligamental area of the vocal fold may also affect the epithelium overlying the arytenoid cartilages. Most of these are, however, more common in the ligamental area.

B.2. Disorders originating in the lamina propria

B.2.1. Acute oedema of the larynx

Tissue of origin: Lamina propria.

Mechanical factors: Symmetrical mass increase, with non-localised protrusion into the glottal space, and normal tissue layer geometry.

Site of occurrence: The mucosal covering of the arytenoid cartilage.

Oedema is a symptom with many possible underlying causes. These include chemical or thermal irritation, infection, allergy, and cardiac or renal failure. It merits some special comment, however, because of its characteristic distribution. Fluid tends to collect first in the mucosa overlying the arytenoid cartilage, and whilst it may spread upwards to the ventricular folds and the epiglottis,

the firm adherence of the tissue laters in the ligamental area limits its anterior spread. The ligamental area, therefore, tends not to be affected except when chronic inflammation leads to Reinke's oedema. The swelling will usually be symmetrical, and is likely to prevent full approximation of the unaffected ligamental portion of the vocal folds. (Birrell 1977, Friedmann and Osborn 1978, Salmon 1979)

B.2.2. Contact ulcer (Contact pachydermia, Contact granuloma)

Tissue of origin: Superficial layer of the lamina propria.

Mechanical factors: An increase in stiffness with a redistribution of mass, localised protrusion into the glottal space, and disrupted tissue layer geometry. The degree of symmetry is variable.

Site of occurrence: The mucosa overlying the vocal processes of the arytenoid cartilages.

Contact ulcer is generally thought to develop from a localised area of inflammation over the vocal process of the arytenoid cartilage, which is the point of maximum impact during adduction of the cartilages for phonation. A pile of granulation tissue develops, and the centre of this becomes worn away to expose the cartilage. The result is a central crater, surrounded by an outgrowth of connective tissue and epithelium. The epithelium may be markedly hyperplastic and keratinized. Contact ulcers are usually bilateral, but there is often some discrepancy in size of the ulcers on the two folds. Vocal abuse and psychogenic factors have both been implicated in the aetiology. (Luchsinger and Arnold 1965, Boone 1977, Birrell 1977, Perkins 1977, Salmon 1979, Aronson 1980)

REFERENCES

Arnold, G. E. (1962) 'Vocal nodules and polyps: laryngeal tissue reaction to habitual hyperkinetic dysfunction'. J. Speech and Hearing Res., 27, 205-216.

Aronson, A. E. (1980) Clinical Voice Disorders. An Interdisciplinary Approach. New York: Thieme-Stratton Inc.

Auerbach, O., Hammond, E. C., and Garfinkel, L. (1970) 'Histological changes in the larynx in relation to smoking habits'. Cancer, 25, 92-104.

Baer, T. (1973) 'Measurement of vibration patterns of excised larynxes'. J. Acoust. Soc. Am., 54, 318 (A).

Bauer, W. C., and McGavran, M. H. (1972) 'Carcinoma in situ and evaluation of epithelial changes in laryngopharyngeal biopsies'. J. of the Amer. Med. Assoc., 221, 72-75.

Berg, van den, J. (1962) 'Modern research in experimental phoniatrics'. Folia Phoniatrica, 14, 81-149.

Berg, van den, J., Vennard, W., Berger, D. and Shervanian, C. C. (1960) Voice production. The vibrating larynx (film). Utrecht: SFW-UNFI.

Birrell, J. F. (1977) Logan Turner's Diseases of the Nose, Throat and Ear (8th edn.). Bristol: John Wright and Sons Ltd.

Boone, D. R. (1977) The Voice and Voice Therapy. New Jersey: Prentice-Hall.

Broad, D. J. (1977) Short course in speech science. Santa Barbara: Speech Communications Research Laboratory.

Davies, D. V. and Davies, F. (1962) Gray's Anatomy (33rd edn.). London: Longmans, Green and Co. Ltd.

Farnsworth, D. W. (1940) 'High speed motion pictures of the human vocal cords' (and film). Bell Laboratories Record, 18, 203-208.

Ferlito, A. (1974) 'Histological classification of larynx and hypopharynx cancer'. Acta Otolar. Suppl., 342, 17.

Fields, S. and Dunn, F. (1973) 'Correlation of echographic visuability of tissue with biological composition and physiological state. J. Acoust. Soc. Am., 54, 809-812.

Freeman, W. H. and Bracegirdle, B. (1966) An Atlas of Histology. London: Heinemann Educational Books Ltd.

Friedmann, I., and Osborn, D. A. (1978) 'The larynx' in W. St. C. Symmers (Ed.), Systemic Pathology, Vol.1, 248-267.

Fritzell, B., Sundberg, J., and Strange-Ebbesen, A. (1982) 'Pitch change after stripping oedematous vocal folds'. Folia Phoniatrica, 34, 29-32.

Greene, M. C. L. (1972) The voice and its disorders. (3rd. edn.) Philadelphia: Lippincott.

Goerttler, K. (1950) 'Die Anordnung, Histologie und Histogenese der quergestreiften Muskulatur in menschlichen Stimmband'. Zeitschrift fur Anatomie und Entwickelungsgeschichte, 115, 352-401.

Hall, S. I., and Colman, B. H. (1975) Diseases of the Nose, Throat and Ear: a handbook for students and practitioners. Edinburgh: Churchill Livingstone.

Hardcastle, W. J., The physiology of speech production. New York: Academic Press.

Hiller, S. M., Laver, J., and Mackenzie, J. (1983) 'Acoustic analysis of waveform perturbations in connected speech'. Edinburgh University Department of Linguistics Work in Progress, 16, 40-68.

Hirano, M. (1974) 'Morphological structure of the vocal cord as a vibrator and its variations'. Folia Phoniatrica, 26, 89-94.

Hirano, M. (1981) Clinical Examination of Voice. New York: Springer-Verlag.

Hirano, M., Gould, W. J., Lambiase, A., and Kakita, Y. (1981) 'Vibratory behaviour of the vocal folds in a case with a unilateral polyp.' Folia Phoniatrica 33, 275-284.

Hirano, M. Kurita, S., and Nakashima, T. (1981) 'The structure of the vocal folds'. In K. N. Stevens and M. Hirano (Eds.), Vocal Fold Physiology. Tokyo: University of Tokyo Press.

Hirano, M., Kakita, Y., Ohmaru, K., and Kurita, S. (1982) 'Structure and mechanical properties of the vocal fold'. In N. Lass (Ed.), Speech and Language: Advances in Basic Research and Practice. New York: Academic Press, 211-297.

Hiroto, I. (1966) 'Patho-physiology of the larynx from the standpoint of vocal mechanism'. Practica Otologica Kyoto, 59, 229-292.

Ishizaka, K., and Flanagan, J. L. (1972) 'Synthesis of voiced sounds from a two-mass model of the vocal cords'. Bell System Tech. J., 51, 1233-1268.

Kaplan, H. M.,(1960) Anatomy and physiology of speech. New York: McGraw-Hill.

Kleinsasser, O. (1968) Microlaryngoscopy and endolaryngeal microscopy. London: Saunders.

Laver, J. (1980) The phonetic description of voice quality. Cambridge: Cambridge University Press.

Luchsinger, R. and Arnold, G. E. (1965) Voice-Speech-Language. Clinical communicology: its physiology and pathology. London: Constable.

Maw, A. R., Cullen, R. J., and Bradfield, J. W. B. (1982) 'Verrucous carcinoma of the larynx'. Clinical Otolar., 7, 305-311.

Matsushita, H. (1969) 'Vocal cord vibration of excised larynges - study with ultra-high-speed cinematography'. Otologia Fukuoka, 15, 127-142 (in Japanese).

Michaels, L. (1976) 'Histopathology of nose and throat'. In R. Hinchcliffe and D. Hamson (Eds.), Scientific Foundations of Otolaryngology, 667-700. London: William Heinemann Medical Books Ltd.

New, G. B. and Erich, J. B. (1938) 'Benign tumours of the larynx: a study of 722 cases'. Arch. Otolaryngol. 28, 841.

Perello, J. (1962) 'The muco-undulatory theory of phonation'. Ann. Otolar., 79, 722-725.

Perkins, H. (1977) Speech Pathology, An Applied Behavioral Science. St. Louis: The C.V. Mosby Co.

Romanes, G. J. (Ed.)(1978) Cunningham's Manual of Practical Anatomy, Vol. 3, Head and Neck and Brain (14th edn.). Oxford: Oxford University Press.

Salmon, L. F. W. (1979) 'Acute laryngitis'. In J. Ballantyne and J. Groves (Eds.), Scott-Brown's Diseases of the Ear, Nose and Throat (4th edn.), Vol. 4, 345-380.

Salmon, L. F. W. (1979) 'Chronic laryngitis'. In J. Ballantyne and J. Groves (Eds.), Scott-Brown's Diseases of the Ear, Nose and Throat (4th edn.), Vol. 4, 381-420.

Sandritter, W. and Wartman, W. B. (1969) Colour atlas and textbook of Tissue and Cellular Pathology (4th edn.). Chicago: Year Book Medical Publishers. Inc.

Saunders, W. H. (1964) The Larynx. New Jersey: CIBA Corp.

Shaw, H. (1979) 'Tumours of the larynx'. In J. Ballantyne and J. Groves (Eds.), Scott-Brown's Diseases of the Ear, Nose and Throat (4th edn.), Vol. 4, 421-508.

Smith, S. (1961) 'On artificial voice production'. Proceedings of the 4th International Congress of Phonetic Sciences, Helsinki, 96-110.

Titze, I. R. (1973) 'The human vocal cords: a mathematical model, Part I'. Phonetica, 28, 129-170.

Titze, I. R. (1974) 'The human vocal cords: a mathematical model, Part II'. Phonetica, 29, 1-21.

Titze, I. R. and Strong, W. J. (1975) 'Normal modes in vocal cord tissues'. J. Acoust. Soc. Am., 57, 736-744.

APPENDIX FOUR

AUTOMATIC ANALYSIS OF WAVEFORM PERTURBATIONS IN CONNECTED SPEECH

Steven M. Hiller, John Laver and Janet Mackenzie

ABSTRACT

Details of an algorithm for the automatic acoustic measurement of waveform perturbations in connected speech are presented. A number of measures of perturbations are defined. Results are reported for the application of the algorithm and the perturbation measures to normal voices and a pathological voice, and discussion is offered of the role of the system in screening voices for potential laryngeal pathology.

The automatic analysis of waveform perturbations in connected speech is an extension of a longstanding research interest in the Phonetics Laboratory in the topic of voice quality (Laver 1967, 1968, 1974, 1975, 1979, 1980; Laver & Hanson 1981; Laver, Wirz, Mackenzie & Hiller 1981, 1982; Laver, Hiller & Hanson 1982). Laver (1980) was an early attempt at providing a comprehensive account of perceptual and physiological aspects of normal voice quality, with some preliminary discussion of acoustic aspects. In a recent three-year project ('Vocal Profiles of Speech Disorders' Medical Research Council Grant No. 9781192N, 1979-82), a research team in the Laboratory developed, from this initial base, a perceptual coding system for describing both normal and pathological voice quality. The system was called 'Vocal Profile Analysis', and has now been taught to some 200 speech therapists in a number of different countries. A preliminary account of the system was given in Laver, Wirz, Mackenzie & Hiller (1981), and a full version, supported by illustrative cassette tapes of pathological voices, will be available soon in Laver, Wirz, Mackenzie & Hiller (1984). Now, in a second three-year project ('Acoustic Analysis of Voice Features' MRC Grant No. 8207136N, 1982-85), we are beginning to explore in more detail an acoustic method for characterizing the pathological voice, developing speech signal-processing programs for use on the Laboratory's computer facilities.

This article is a progress report on acoustic and computing aspects of this second MRC project. A companion article (Mackenzie, Laver and Hiller 1983) in this volume reports on anatomical and mechanical aspects of structural pathologies of the vocal folds, and their consequences for perturbatory details of the laryngeal waveform. The project is directed by John Laver; Steve Hiller is responsible for computing aspects, and has written all the computer programs discussed below. Janet Mackenzie is responsible for the speech pathology work. Another member of the project is Robert Hanson, who is a Visiting Senior Scientist from Bell Laboratories, Indian Hills, Chicago: his role is to visit the project each year and advise on signal processing and acoustics.

OBJECTIVES

The broad objective of the project is to explore the feasibility of an automatic acoustic screening system for the early detection of laryngeal pathology. Our first goal is to find acoustic

parameters, such as dysperiodicity of the fundamental frequency of the laryngeal waveform, which can be used to differentiate the healthy population from those with laryngeal pathologies that perturb the laryngeal waveform. Our later objective is to try to differentiate between the various pathologies of the larynx, initially at a descriptive level, and then possibly from a more diagnostic point of view, on the basis of different degrees and types of waveform perturbations (and other anomalies, such as inter-harmonic spectral noise from incomplete glottal closure due to growths on the vocal folds, paralysis of the vocal folds, etc.). Our third objective is to differentiate between stages of progression, either of a given disease, or of rehabilitative improvement. Even the first of these goals poses considerable difficulties. This is true for various reasons - not the least of which is the fact that almost all current speech signal processing programs available today have inbuilt assumptions that are biased towards the normal model of speech. The more one moves towards abnormal pathology, the more these assumptions are violated, and the less effective the signal processing programs very often become. One of the benefits of working in this area, though, is precisely that these discontinuities (and some continuities) between the normal model and the model we need to develop for the abnormal are highlighted. There is also an important sense in which the study of abnormal malfunction throws light on normal function.

If it is socially important to develop a method of screening the general population for such states as early laryngeal cancer, it is perhaps worthwhile asking the question 'Why choose an automatic acoustic method?' - rather than, say, a perceptual, auditory method, or a physiological method such as electrolaryngography (Fourcin 1974). A number of comments can be offered in reply to this question. Firstly, the provision of an acoustic facility allows an objectivity that a solely auditory approach cannot reliably match. Secondly, as an instrumental technique, an acoustic facility (like physiological facilities) provides a permanent written record which can be repeatedly consulted at leisure, copied for communication purposes, and which allows a detailed quantification of the material analysed. Thirdly, an acoustic facility involves a recording technique that is easily portable, easily used in clinical and other environments, and one which is completely non-invasive. It is a technique that is relatively familiar and unfrightening to patients, and the technology for recording is cheap both in capital and recurrent terms. Because of the portability of acoustic recordings, the analysis facility can be remote from the recording facility in both time and space. This allows a single analysis facility, in some central location, to service a large number of varyingly distant clinics. There are, however, a number of disadvantages to an acoustic facility of this sort. Acoustic signals are inherently contaminable by environmental noise in a way that is less true of physiological signals from such techniques as electrolaryngography. In addition, the remoteness of a central analysis facility brings into consideration factors of communication-links and turn-round time that are less relevant to the technology of local physiological analysis. If an automatic acoustic analysis facility were to be proved feasible for clinical application, then favourable financial criteria come into play. Tape recording facilities are already widespread in hospitals, and the possibility of a single, remote analysis facility minimizes the overall financial outlay, compared with the cost of equipping a wide range of clinics with stand-alone physiological instrumentation. However, a sensible eventual policy

PAGE
NUMBERS
CUT OFF
IN
ORIGINAL

might be to combine the advantages of the two complementary approaches, with a central acoustic facility and local physiological facilities.

An alternative approach would be to adopt local physiological instrumentation and combine it with methods of local acoustic analysis which could be developed for use with microcomputers within each clinic. The one problem with this alternative solution is that, given the currently limited capacity and speed of microcomputers, initial data-acquisition would have to be achieved by special-purpose hardware. Once such combinations of microcomputer plus special-purpose hardware became available, or the speed and capacity of microcomputers increased sufficiently, then the equipment could also be used interactively with the patient as a clinical instrument of assessment and rehabilitation. It is taken for granted that all these approaches combine instrumental techniques with auditory observation by the therapist concerned.

INTONATION VERSUS PERTURBATION

From now on, it will be convenient to concentrate on the role and measurement of just one aspect of speech, that of fundamental frequency (FO).

On close inspection, the succession of pitch periods in voiced speech does not show a perfectly smoothly-changing sequence of durational values, in connected speech. In even the healthiest of voices, the duration of each successive pitch period tends to vary, randomly, from the general trend-line discernible through a sequence of such periods. The trend-line represents the intonational contour, and the local deviations of individual periods from the smooth trend-line, as a perturbation of this trend, are perceived in terms of an auditorily 'rough' phonatory quality. The more dysphonic a voice, the greater is the degree of such perturbation, and the greater is the degree of perceived 'roughness'. One of the problems in choosing a suitable method for the automatic detection of the duration of pitch periods in the acoustic waveform is that there is often then a tension between two quite different needs: the need to establish the smoothed trend which represents the intonational contour, versus the need to register as accurately as possible the momentary deviations (or 'excursions') of individual periods from this smoothed trend, representing phonatory quality. Most pitch period extraction algorithms involve a good deal of smoothing in their inherent design, and as such are well-suited to gathering intonational data. There are very few algorithms available that are capable of tracking the exact durations, cycle by cycle, of the perturbed train of periods that is characteristic of not only dysphonic, pathological voices, but also of many types of normal voices.

The present project is interested in both sorts of data, intonational and perturbational. The algorithm we chose was a parallel-processing method working in the time domain, devised originally by Gold and Rabiner (1969). It was chosen in the light of criteria emerging from comparative studies of a number of pitch period detection algorithms (Rabiner, Cheng, Rosenberg, and McGonegal 1976; Laver, Hiller and Hanson 1982). The Gold and Rabiner method was felt suitable for the project's needs in that it can work on connected speech from both male and female speakers, is resistant

to poor signal-to-noise ratios from recordings in hospital environments, as well as being resistant to interharmonic spectral noise, and retains accuracy of period duration estimation in conditions of fairly acute waveform perturbation in both fundamental frequency ('jitter') and intensity ('shimmer'). Steve Hiller has written a version of the Gold and Rabiner algorithm, and we have developed a number of automatic measures of waveform perturbation. These will be described in turn.

1. AUTOMATIC PITCH PERIOD ESTIMATION SYSTEM

1.0. INTRODUCTION

The basic scheme of the parallel processor, as a very fast program able to be implemented on a general purpose computer, has been described by Rabiner and Schafer (1978:136) as follows:

1. Initial processing of speech signal creates a number of impulse trains which retain the periodicity of the original signal and discard features which are irrelevant to the pitch detection process.

2. This processing permits very simple pitch detectors to be used to estimate the periodicity of each impulse train.

3. The estimates of these simple pitch period detectors are logically combined to infer the period of each laryngeal cycle in the speech waveform.

The idea of parallelism in period detection is that the outputs of a number of simple parallel measures of periodicity for a given speech segment are the inputs to a sophisticated majority logic measure which determines the segment's official pitch period. Gold and Rabiner (1969) suggested that parallelism, as implemented in an automatic pitch period estimator, may simulate the visual observations of a human examining a speech waveform for periodicity.

1.1. THE ALGORITHM

A block diagram of the parallel processor is shown in Figure 1 (adapted from Gold and Rabiner, 1969). The input speech is low-pass filtered to reduce formant information and then processed to produce several functions representing different aspects of periodicity in the waveform. A simple pitch period detector is then applied to each function to determine the periodicity displayed by that function. The various measures of periodicity derived from the functions are then combined in a sophisticated manner to determine the most likely pitch period for the input speech. In addition, processes are required for determining the presence of speech (i.e., discrimination between speech and silence) as well as the likelihood of the resultant pitch period representing a voiced or voiceless segment. The general structure of the program follows the more elaborate version of Gold and Rabiner's (1969) parallel processor in order to accommodate the widest variety of voice types. In the present implementation, the program completes the parallel processing of a given window of speech data and then the window is shifted forward in time to try to capture the next pitch period.

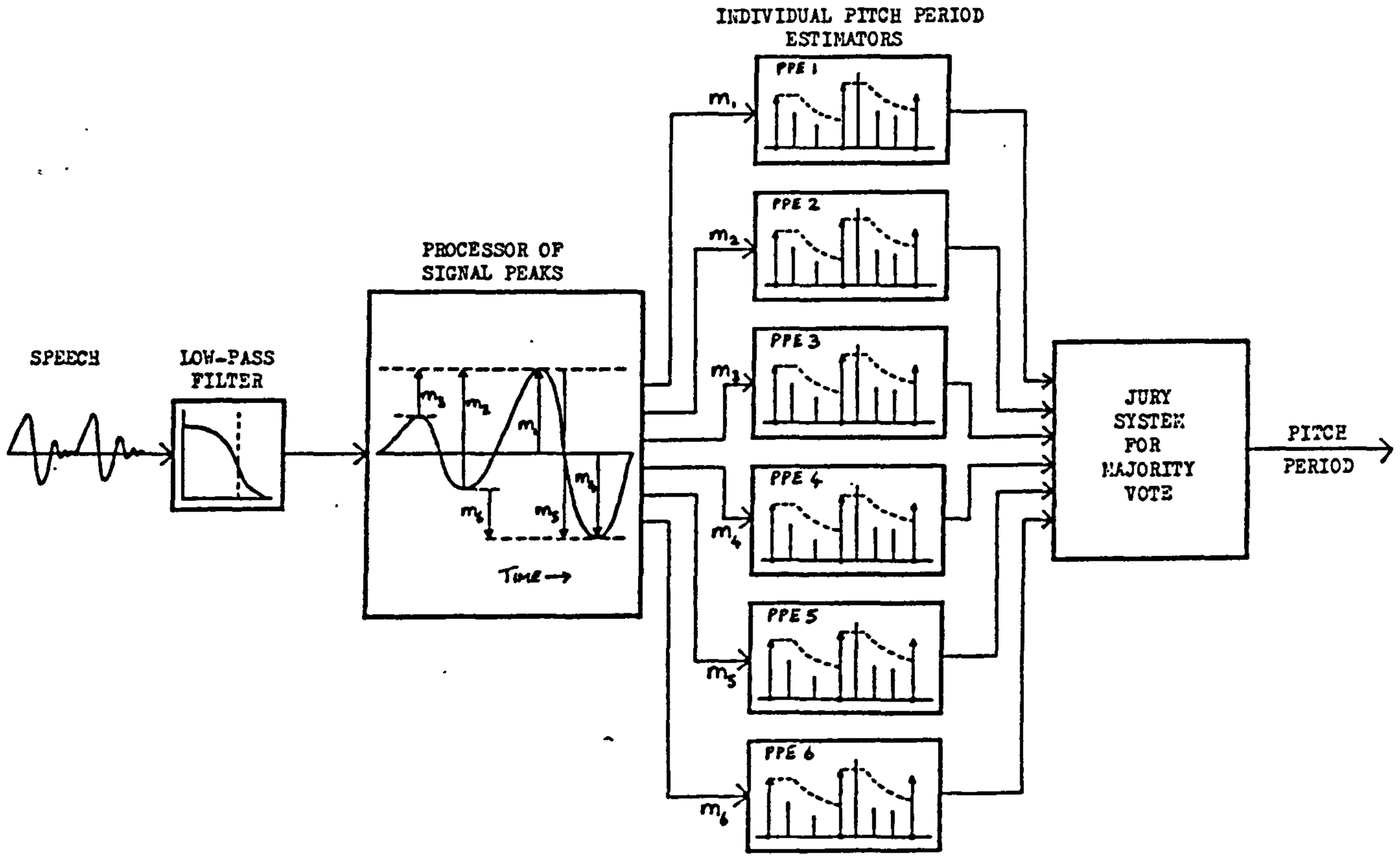


FIGURE 1. Block diagram of the parallel processor for pitch period estimation (adapted from Gold and Rabiner 1969).

1.1.1. Low-Pass Filtering

The input speech signal is low-pass filtered to produce a signal which has been spectrally shaped to contain mostly fundamental frequency information, thus simplifying the period extraction task. In the present system, the low-pass filtering is completed prior to the digitization process by an analog filter. The filter is a Butterworth type which produces a -24 dB/octave rolloff beyond a specified stop band frequency. The cutoff frequency is set to 400 Hz for male voices and 600 Hz for females. This filter also acts as an anti-aliasing filter to prevent spectral distortions during sampling.

1.1.2. Sampling rate

At present, the low-pass filtered signals are digitized at a sampling rate of 10 KHz, as suggested by Gold and Rabiner (1969), thus providing the resolution of pitch periods to within .1 msec. This appears to be a reasonable resolution for typical male fundamental frequencies but increased sampling rates may be required for the higher fundamental frequencies of females and children (Horii, 1979). The digitized signal is then filed for further signal processing.

1.1.3. Silence detection

The pitch period estimation begins by determining the presence of speech within a given window of input data. The silence detection technique is a simple one described by Gold (1964), in which the segment of data is searched for two samples which exceed a pre-determined 'silence' threshold. If the threshold is exceeded then the remainder of the estimation is completed, otherwise the pitch period result is set to zero and the next frame of data is processed. The silence detection threshold is determined interactively for each voice sample by calculating the peak intensity level of the background noise presented in each tape recording. Gold and Rabiner (1969) noted that the parallel processor worked well in low signal-to-noise ratio conditions. This point has been supported for a number of voice samples recorded in rather noisy clinical environments in which good pitch period estimation was possible.

1.1.4. Processing of signal peaks

If speech is present, then the smoothed speech is examined for the presence of "peaks and valleys" (i.e., maxima and minima) which represent periodic behavior in the waveform. Several measures of amplitude are calculated as each valley and peak is located. The amplitude measurement scheme is displayed in Figure 1. This scheme uses six amplitude measurements, which were defined by Rabiner and Schafer (1978, 137) as follows:

1. m1(n): An impulse equal to the peak amplitude occurs at the location of each peak.

2. m2(n): An impulse equal to the difference between the peak amplitude and the preceding valley amplitude occurs at each peak.

3. m3(n): An impulse equal to the difference between the peak amplitude and the preceding peak amplitude occurs at each peak. (If this difference is negative the impulse is set to zero.)

4. m4(n): An impulse equal to the negative of the amplitude at a valley occurs at each valley.

5. m5(n): An impulse equal to the negative of the amplitude at a valley plus the amplitude at the preceding peak occurs at each valley.

6. m6(n): An impulse equal to the negative of the amplitude at a valley plus the amplitude at the preceding local minimum occurs at each valley. (If this difference is negative the impulse is set equal to zero.)

The use of six different measures of waveform characteristics is designed to cover a range of different types of waveform, varying from a simple sinusoid to a signal composed of a weak fundamental component with a strong second harmonic. Each type of peak and valley measurement produces an impulse train made up of positive impulses representing the amplitudes and locations of the measurements.

1.1.5. Pitch period estimation of the peaks

Each impulse train is evaluated for periodicity by a peak detecting circuit based on an exponential decay function (Gold, 1962). Figure 1 demonstrates the basic operation of this exponential circuit. Following the detection of a possible pitch period marker, the circuit is reset and held for a blanking interval during which no detection occurs. After the blanking interval, the circuit begins to decay. The decay continues until an impulse of sufficient amplitude exceeds the decay threshold, and then is once again reset. In this manner, possible pitch period information is stored and extraneous data discarded. The decay behavior of the exponential circuit (i.e., blanking time and decay rate) is dependent upon local pitch period trends in order that reasonable limits are set for the detection of the next period.

1.1.6. Final computation of the pitch period

For each analysis interval, the peak detecting circuit produces six estimates of the pitch period, one for each of the six impulse trains. These estimates of periodicity are combined with the two most recent sets of estimates from the six parallel pitch period detectors. The final determination of the pitch period is based on a comparison of all the estimates. The estimate with the greatest level of agreement among the six immediate candidates is declared the official pitch period for the speech segment. It should be noted that this method of calculating pitch period causes the loss of some period information at the onset of phonation.

1.1.7. Voiced/voiceless decision

Gold (1964) described the technique used for determining whether the chosen pitch period represents a voiced segment of speech. Voiced/voiceless decisions are determined from the level of agreement between the chosen pitch period estimate and the other period

measures. For voiced speech, the agreement level will be high since each simple detector represents redundant information concerning the periodic behavior of the waveform. There is a lack of redundancy associated with noisy voiceless speech and therefore a low level of agreement for any pitch period estimate. A voiced/voiceless decision threshold can be determined from the distributions of the agreements calculated for voiced and voiceless speech (Gold, 1964).

1.2. ANALYSIS CONDITIONS FOR OBTAINING MICROPERTURBATORY DATA

Since the main objective of the present research is the capture of valid cycle-to-cycle perturbation information, a number of analysis conditions linked to the pitch period estimation process need to be considered. The general approach behind our implementation of the parallel processor is to apply the system to an interval of speech data, accept the last pitch period within an analysis interval detected by the exponential decay system as the representative period, and then shift the window forwards to include the next pitch period. The analysis conditions of most importance to the system are thus the nature of the analysis interval (the analysis 'window'), the shifting of the window, and the waveform feature to be used as a pitch period marker.

1.2.1. Analysis interval conditions

Each pitch period estimation is completed on a segment of filtered speech data selected by a rectangular analysis window. The interval within the window is set to accommodate the largest probable pitch period to be produced by a given speaker. At present, the analysis interval is set to a duration of 25 msec (40 Hz) for male speakers and 20 msec (50 Hz) for female voices. Given the rather long durations of the analysis interval, it is normal for more than one pitch period to be present in the window at any one occasion of period detection. The program has been designed to produce an estimate of period for the last complete cycle in the window.

1.2.2. Shifting of the analysis window

Cycle-to-cycle data is estimated by shifting the rectangular window along the data in such a way as to try to bring just one new pitch period into the window. A shift of 10 msec (100 samples at 10 KHz sampling rate) would thus be ideal for a steady fundamental frequency of 100 Hz. However, this ideal situation is seldom reached, because, in continuous speech, fundamental frequency is naturally moving up and down, both for intonational reasons and for microperturbatory reasons. The algorithm is therefore accurate, in the estimation of any two adjacent periods, only within a certain band of fundamental frequencies. The limits of this band are set by the size of the shift factor, basically. If one considers the situation where a new cycle is being brought into the window by one application of the shift factor, then the longest new period that can be accurately detected is one which is no longer than the shift factor itself. If it is longer, then the previous cycle, already estimated once, remains the last complete cycle in the window, and is re-reported. Under-shifts thus result in over-reporting. Conversely, the shortest new period that can be accurately detected is one which is, at a minimum, greater than half the shift factor itself. If it is half the duration or shorter, then (assuming that the next cycle has the same period or less) the algorithm effectively

jumps a cycle and reports the next one as the last in the window. Over-shifting therefore results in under-reporting. Thus, an octave band of accurate FO estimation is provided by a given shift factor - this band demonstrating tolerance to increased FOs and intolerance to decreased FOs, relative to the shift factor. This is perhaps less important if one's interest lies in intonation, but it becomes very relevant if the object of attention is perturbatory behavior, where exact cycle-to-cycle measurement is the goal.

It can be seen that the algorithm retains accuracy of perturbatory tracking only to the extent that the combination of intonational and perturbational movement of FO remains within a frequency-zone whose limits are determined by the shift factor. It is clearly helpful if a shift factor can be chosen, in the examination of a given voice, that relates in duration to some statistical property of the period durations to be found in that voice, to optimize accurate pitch period estimation. The simplest pitch-adaptive strategy would be to set the shift factor to one value for males, another for females, and another for children, on the basis of general values found in these populations. The next step in tuning the shift factor to allow accuracy of pitch period extraction would be to adjust it to some statistic of the individual speaker's typical performance, for example, the mean, median, or mode FO of the habitual speech. Finally, one could try to make the shift factor fully pitch-adaptive, using strategies to change the value of the shift factor dynamically, on the basis of predictions about future short-term period behavior reached from examinations of local past short-term history of FO. These three types of pitch-adaptive strategies will be referred to as sex-specific tuning, speaker-specific fixed tuning, and speaker-specific variable tuning.

All three types of approach were used experimentally in comparing the benefits of fixed and variable settings of the shift factor. For each speaker, we made a preliminary pass through the data, using a sex-specific shift setting of 10 msec (this setting was for male speakers). From this, the median FO was calculated and used to give a fixed shift which was speaker-specific. Alternatively, the sex-specific setting was used as a starting point for processing the speaker's data by means of a variable shift factor. This variable shift was calculated as follows:

1) An assumption was made that there is an underlying orderliness in the train of pitch periods in speech. In the extreme case this would be represented by an FO contour which would be a straight line - level, rising, or falling. Within voices that can be considered to be normal and healthy, microperturbatory excursions can be anticipated to be infrequent, to be small in extent, and to have a normal distribution for size of excursion.

2) What was needed was some means of predicting the slope of the FO trend, from knowledge of recent FO trend behavior. One possibility was to use a moving-average approach to establish the history of recent FO trend. But means are very vulnerable to the influence of single eccentric values. So it was decided to base the prediction of slope of the FO on recent medians. We chose a moving 5-point median.

3) The prediction of slope was calculated as follows: let Sn equal the variable shift factor to be evaluated as an optimized attempt to bring in the next pitch period Pn economically and accurately, and Mn equal the median value of the five estimated periods prior to that next period. Sn can be estimated on the basis of the difference between the two most recent median values (Mn - Mn-1), this difference being a measure of the slope of the FO trend as estimated at the appropriate delay for the median (i.e., Pn-3). If the difference is equal to zero (i.e., the projected slope is horizontal), then let the next variable shift Sn equal the previous shift factor Sn-1. Otherwise, the next shift is determined from a straight-line approximation from the last median value which includes a factor for the delay, that is, Sn = MN + 3(Mn - Mn-1).

With this variable shift, inaccuracies will arise only under certain conditions of FO movement (leaving aside the consideration of perturbations for the moment). These inaccuracies occur at any intonational corner - i.e., at any point of departure from a straight-line trend. It can be seen that there are limiting values for accurate measurement in these changing contours, beyond which error is inherent.

Figure 2 displays two hypothetical pitch period contours, rising and falling, to which the variable shifting logic has been applied. Each contour (the solid line) is plotted as pitch period duration (ordinate) versus the order of the pitch period estimated sequentially in time (abscissa). The first six points of each contour are the six most recently measured periods. Point Pa on the abscissa is the next period (of as yet unknown duration) to be estimated relative to the shift factor produced by the variable shifting algorithm for medians Mn-1 and Mn. It can be seen for each contour that the zone within which accurate estimation of the incoming period can be achieved (the octave band represented by the dotted line at point Pa) has values determined by the local short-term FO behavior. In the case of the rising pitch period contour (i.e. falling intonation contour), there is tolerance to change-over points (i.e. falling to rising intonation) and no tolerance for rising accelerations of period duration (i.e. increasingly negative intonational slope). For the falling pitch period contour (i.e. rising intonation), there is tolerance for falling accelerations (i.e. increasingly positive intonational slope) and no tolerance for change-over points (i.e. falling to rising periods, rising to falling intonational contour).

Similar constraints operate for perturbed waveforms, and the underlying assumption of orderliness in the data in the form of a straight line tendency becomes progressively invalid with increased severity of cycle-to-cycle perturbatory differences. There are two major problems in severely perturbed waveforms for a variable shifting mechanism of this sort. Firstly, the projection of the predicted slope of FO can swing wildly, giving values for Sn which take extreme forms and which thus minimize the likelihood of effectively capturing the next true period. Secondly, with contributory adjacent median values differing widely, it is logically possible for negative shifts to occur. In these circumstances, using a variable shift can actually be counterproductive, and can

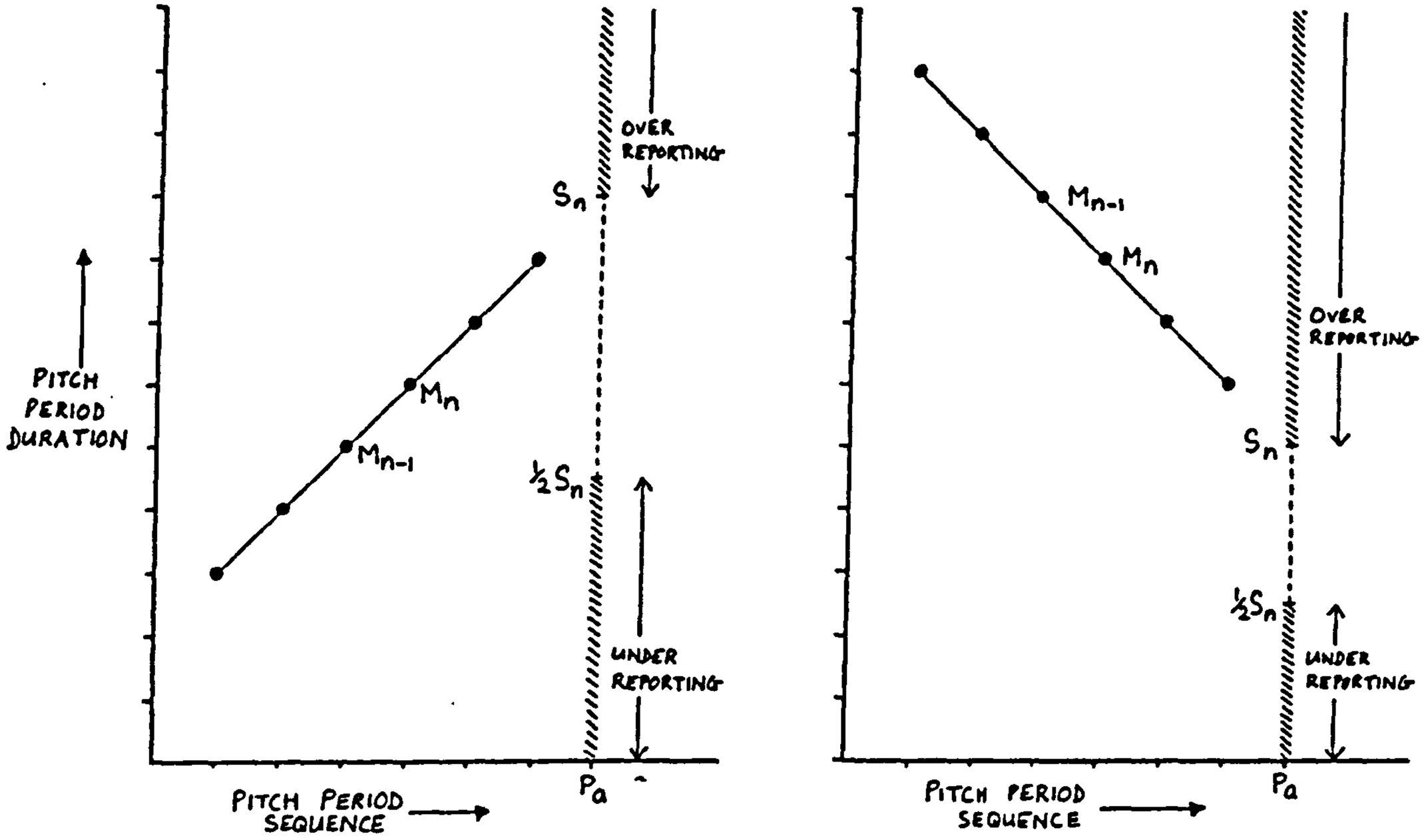


FIGURE 2. Application of the variable shifting algorithm to two hypothetical pitch period sequences (a - rising, b - falling). The two medians Mn and Mn-1, derived from the six most recently measured periods, are used to predict the next period at point Pa. Note how the octave band (at Pa) for accurate period estimation is located relative to recent F0 behavior.

itself contribute artifactually to high perturbation values. A partial solution, up to moderate perturbation levels, is to set range-limits. We set a range-limit of 40 to 240 Hz for male speakers. When the extrapolated shift fell outside this limit, the calculation was cancelled, and the first-pass speaker-specific value was substituted. At the same time, a flag was set for each occurrence of this out-of-range incident, to keep a measure of how often the range-limits were invoked, and the first-pass speaker-specific shift value substituted.

1.2.3. Pitch period markers

This condition is concerned with the choice of waveform features which yield pitch period markers. Normally, preference is given to the pitch period marker which relates to the positive peak impulse function (see m1 in section 1.1.4 above). The positive peak detector was chosen for two reasons. First, the final computation of the period by majority logic is biased towards the positive peak when comparisons between the various period measures produce equivalent levels of agreement (e.g. in smooth unperturbed segments of voiced speech). Second, the positive peak parameter is the one most directly related to the impulse behavior of the vibrating vocal folds. Further bias towards the positive peak has been added to the system to accommodate small variations in period measurements. It was observed early on that the period durations varied slightly between some of the pitch period detectors for a given pitch period. The slight variations appear to be the result of actual differences for the various features of the low-passed waveform, and perhaps of the effects of the digitizing process. In these cases, it was observed that some other pitch period marker (e.g., m4 - the negative peak measure) had the highest level of agreement even though the positive peak marker was clearly visible and similar in duration. This is the logical consequence of a program which uses past information and redundant features to arrive at a final decision. It was decided, after inspection of typical waveforms, to force the final measure of the period to be the positive peak marker if the difference in duration between some other chosen feature and the positive peak measure was minimal. For the time being, the system is set to choose the positive peak marker if there is a difference of less than or equal to 3 msec between the two period measures. This minimal difference appears to work well for a majority of cases, as will be discussed below. Differences greater than 3 msec are accepted as an indication of perturbed behavior in the waveform and the alternative peak marker duration is stored.

1.3. PERFORMANCE OF THE AUTOMATIC PARALLEL PROCESSOR COMPARED TO VISUAL OBSERVATIONS OF NORMAL SPEECH

It was important to evaluate the performance of the pitch period extraction system when applied to data of known characteristics. In particular, we were concerned with the behavior of the system under the two methods of window shifting (fixed and variable) which we felt would have the greatest effect on accurate perturbation measurement. The following discussion is based on a small pilot study to determine the types of error produced by the automatic system in comparison to visual examinations of speech stimuli.

1.3.1. The pilot study

The automatic pitch period extractor and visual examinations were applied to the stimulus utterance 'A rainbow is a division

of white light into many beautiful colors'. Tape recordings of the utterance were produced by three normal-speaking male adults (RK, JL, SH). The parallel processor was applied to the data in two manners: 1) shifting of the analysis window by a fixed speaker-specific shift factor based on the median period duration derived on a first-pass analysis of the stimulus and 2) variable shifting using a shift factor based on the median shifting logic presented in section 1.2.2. The output of the automated system was compared with visual examinations of the low-pass filtered versions of the stimuli using a cursor program on the minicomputer's visual display unit. The results of the comparisons are summarized in Tables I and II for the fixed and variable shift conditions for each speaker.

1.3.1.1. Under-reporting/over-shifting; over-reporting/under-shifting

There is a marginal advantage in these normal voices for the variable shift. In other words, the distribution of FO values for each speaker falls typically within the accuracy span of the shift-setting of the fixed shift, and making the shift-setting pitch-adaptive brings only a small improvement. It is noteworthy that there is an overall low incidence of pitch period over-reporting for each utterance, given the intolerance of the octave band to FOs deviating towards lower frequencies relative to the local FO trend. This result suggests that the intonational behavior evidenced in the utterances was mostly free of decelerating changes from the local FO trends, and that falling intonational contours typically followed more straight-line tendencies. Further research into a more refined mechanism for variable shifting is currently being undertaken, however.

1.3.1.2. Over-reporting due to shimmer factors in sudden low-amplitude values for waveform peaks

Recalling that an exponential decay function is an integral part of the period detection algorithm, when shimmer factors drop the amplitude of waveform peaks below the exponential threshold, the next true peak is usually beyond the shifted window, and the previously reported cycle is treated as the last complete cycle in the window and re-reported. Values for this type of error were low in both the fixed and the variable shift operations, and the differences were negligible. However, this ability of shimmer factors to contribute to jitter data should persuade us, as Askenfelt and Hammarberg (1980, 1981) suggest, to talk of waveform perturbation, rather than of jitter alone.

1.3.1.3. Non-positive pitch period marker

Despite the bias towards the positive peak parameter, occasionally some other aspect of the waveform receives the majority vote. The figures are very low in both cases due to the additional forcing logic for small variations between simple pitch period detector durations.

1.3.1.4. Voiced-to-unvoiced errors

Occasional low levels of agreement between simple period estimates due to perturbations in the waveform result in an improper unvoiced decision relative to the visual estimation of

Subject	RK N=1yy, CTX=y6	JL N=216, CTX=y3	SH N=211, CTX=y0
Under-reporting/ Over-Shifting	5.0% (10)	4.2% (y)	7.1% (15)
Over-reporting/ Under-Shifting	2.5% (5)	5.1% (11)	1.4% (3)
Over-reporting/ Low Amplitude	1.5% (3)	1.4% (3)	3.3% (7)
Non-positive Peak Detector	3.5% (7)	3.2% (7)	2.4% (5)
Voiced-to- Unvoiced Error	0.5% (1)	0.5% (1)	1.4% (3)

TABLE I
Errors in automatic pitch period estimation, using a FIXED shift factor, relative to visual estimation, in three normal male voices.

Subject	RK N=1y8, CTX=y6	JL N=216, CTX=y3	SH N=225, CTX=y0
Under-reporting/ Over-Shifting	4.5% (y)	3.7% (8)	3.6% (8)
Over-reporting/ Under-Shifting	2.5% (5)	3.2% (7)	2.6% (6)
Over-reporting/ Low Amplitude	1.5% (3)	2.3% (5)	3.1% (7)
Non-positive Peak Detector	3.0% (6)	3.2% (7)	3.1% (7)
Voiced-to- Unvoiced Error	0.0% (0)	0.5% (1)	1.8% (4)

TABLE II
Errors in automatic pitch period estimation, using a VARIABLE shift factor, relative to visual estimation, in three normal male voices.

the waveform. The number of voiced-to-unvoiced errors is very low for the data, and supports the findings of Rabiner et al, (1976).

2. PERTURBATION ALGORITHM

2.0. INTRODUCTION

The design of the algorithm for calculating microperturbatory behavior was primarily based on the nature of fundamental frequency contours extracted from continuous speech. The fundamental frequency curves of continuous speech represent modulations of FO associated with intonational aspects of the utterance as well as short-term microperturbations of FO correlated with efficient use of laryngeal vibration. Continuous speech also introduces the influence of segmental performance into the FO contour, such as pauses, and voicing onsets/offsets, and the effects of stop closures, nasality, etc. In addition, the process of pitch period estimation sometimes produces artifacts in the contour through incorrect period estimations. The primary choice of perturbation algorithm was based on the need for a system which provided intonational information in the form of the underlying smooth curve of the raw pitch periods; this curve being a useful baseline from which to measure the variation of the raw pitch periods from local smoothed behavior. Secondarily, the system had to be able to cope with the segmental and artifactual features evidenced in FO contours of continuous speech.

2.1. THE ALGORITHM

The raw FO curve extracted by the parallel processor is passed through a non-linear smoother to produce a contour equivalent to the smoothed underlying trend of the data. The non-linear smoother was implemented as a running digital filter which enables the determination of excursion behavior of the raw FOs from the local smoothed output of the filter. The smoothed FOs and their associated excursions are statistically evaluated for intonational and microperturbatory measures.

2.1.1. The trend line

The trend line underlying the raw FO curve is constructed by a non-linear smoother presented by Rabiner, Sambur, and Schmidt (1975). A non-linear smoother has advantages over more conventional linear smoothers (e.g., running average) which tend to smear sharp discontinuities present in speech signals as well as being affected by gross errors in the contour. A non-linear smoother was chosen since we wanted to preserve realistic discontinuities present in FO contours - these discontinuities representing transitions from voiced to voiceless states and vice versa - while smoothing microperturbatory roughness and gross pitch period estimation errors. The non-linear smoother implemented is a combination of running median filter plus a Hanning window.

The median filter serves to preserve sharp discontinuities in the FO contour, where the desirable discontinuities must be of a minimum critical duration. Rabiner et al. (1975) noted two significant characteristics of the median filter. First, the size of the median filter is based on the minimum duration which defines an acceptable discontinuity. In the present system, we are con-

cerned with discontinuities which represent transitions from voiced to voiceless states, and vice versa, evidenced in FO contours of continuous speech. Voiced (i.e., greater than 0 Hz) and voiceless (i.e., equal to 0 Hz) segments of a FO contour were operationally defined as those segments consisting of three or more sequential FOs of either state. Therefore, a median filter with a duration of five samples is required to preserve discontinuities of three samples or more. Second, the median filter inherently smooths out sharp discontinuities in the signal which are shorter than the minimum acceptable duration. In our system, very short discontinuities are considered to be gross errors in pitch period extraction and, as a result of the operational definition for segments, one and two point discontinuities are smoothed out by the 5-point median filter. The advantage of this second characteristic of the median filter is that large errors do not affect the surrounding calculations of the trend line.

A Hanning window is used as a linear smoother to filter out the less sharp noise components evidenced in speech signals. In the present research, the noise components represent microperturbatory movements in the raw FO contour. A 3-point Hanning window is used in the non-linear smoother as recommended by Rabiner et al. (1975).

The combination of a 5-point median filter and a 3-point Hanning window results in a filtering delay of 3 points for the non-linear smoother. Rabiner et al. (1975) noted the need for additional logic for determining the beginning and ending points of the output data which are lost due to the filter delays. The primary concern of the present research is quantifying the perturbation behavior in the FO contour and therefore the onset and offset data are not included as part of the perturbation data.

2.2.2. Excursions

Excursions represent the deviations of the raw FOs from the equivalent smoothed values produced by the non-linear smoother. The use of excursion measures from a smoothed trend has been presented in Koike (1973), Kitajima, Tanabe and Isshiki (1975), Davis (1976), Kitajima and Gould (1976), Koike, Takahashi and Calcaterra (1977), and Laver, Hiller and Hanson (1982). Excursions are measured relative to a smoothed trend line in order that slow-moving modulations (e.g., vibrato) and intonational movements of FO are excluded from contributing to perturbation parameters.

An excursion is derived for each output of the non-linear smoother and defined as the difference between the raw FO and its equivalent smoothed FO. Each excursion is stored in four formats: 1) signed excursion in Hz - the difference between raw and smoothed FO in units of Hz with the algebraic sign retained, 2) signed excursion in percent - the ratio of the signed excursion in Hz to its associated smoothed FO multiplied by 100, 3) magnitude excursion in Hz - the absolute value of the signed excursion in Hz, and 4) magnitude excursion in percent - the absolute value of the signed excursion in percent. The signed and magnitude excursions in percent are used to normalize excursion measures extracted from varying FO levels evidenced within a sample of continuous speech of a single speaker as well as between different speakers. Excursion measures are not calculated for voiceless segments of FO contours and regions of very short discontinuities. In the latter case, a flag is set denoting each instance of a short discontinuity.

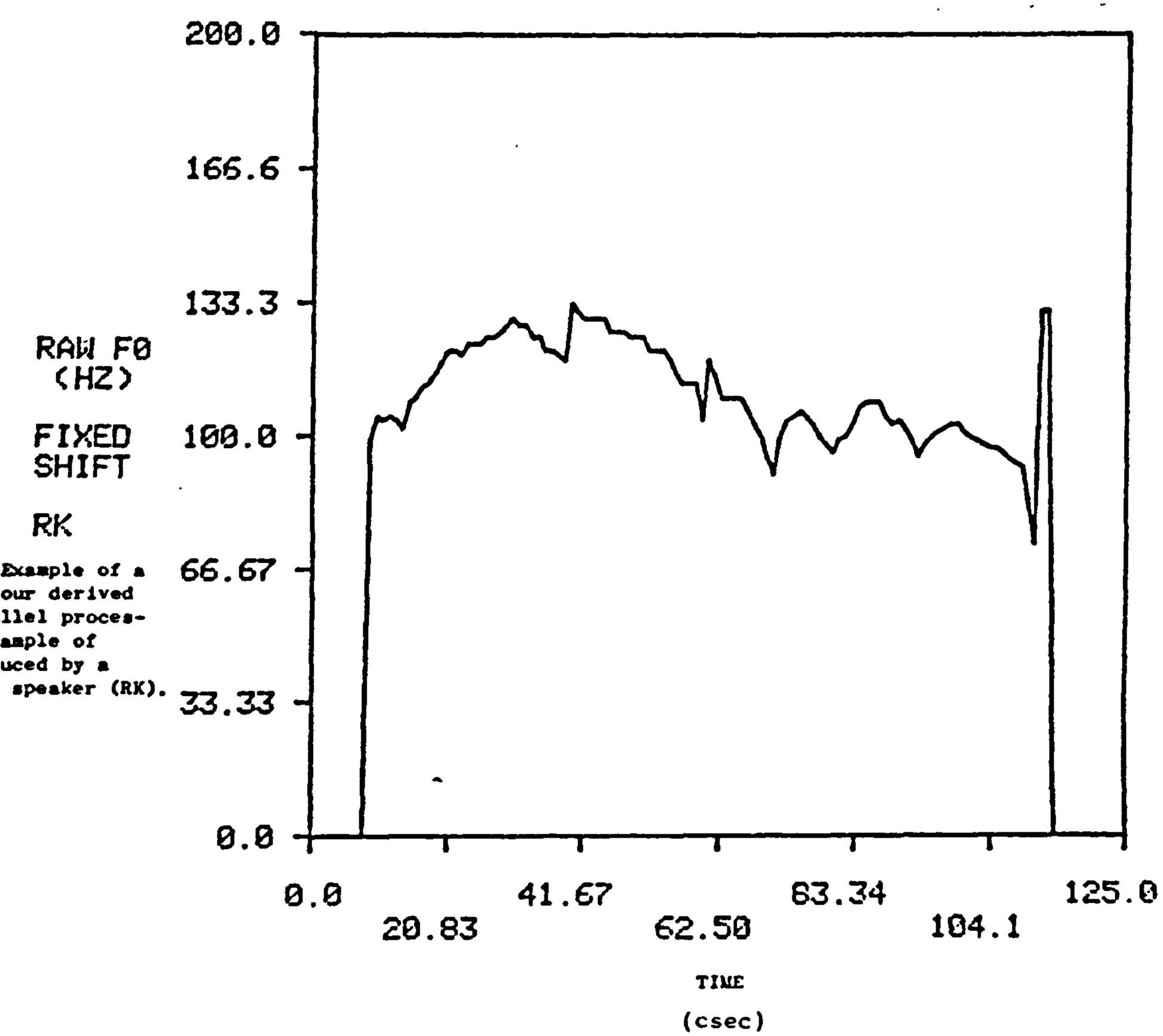


FIGURE 3. Example of a raw F0 contour derived by the parallel processor for a sample of speech produced by a normal male speaker (RK).

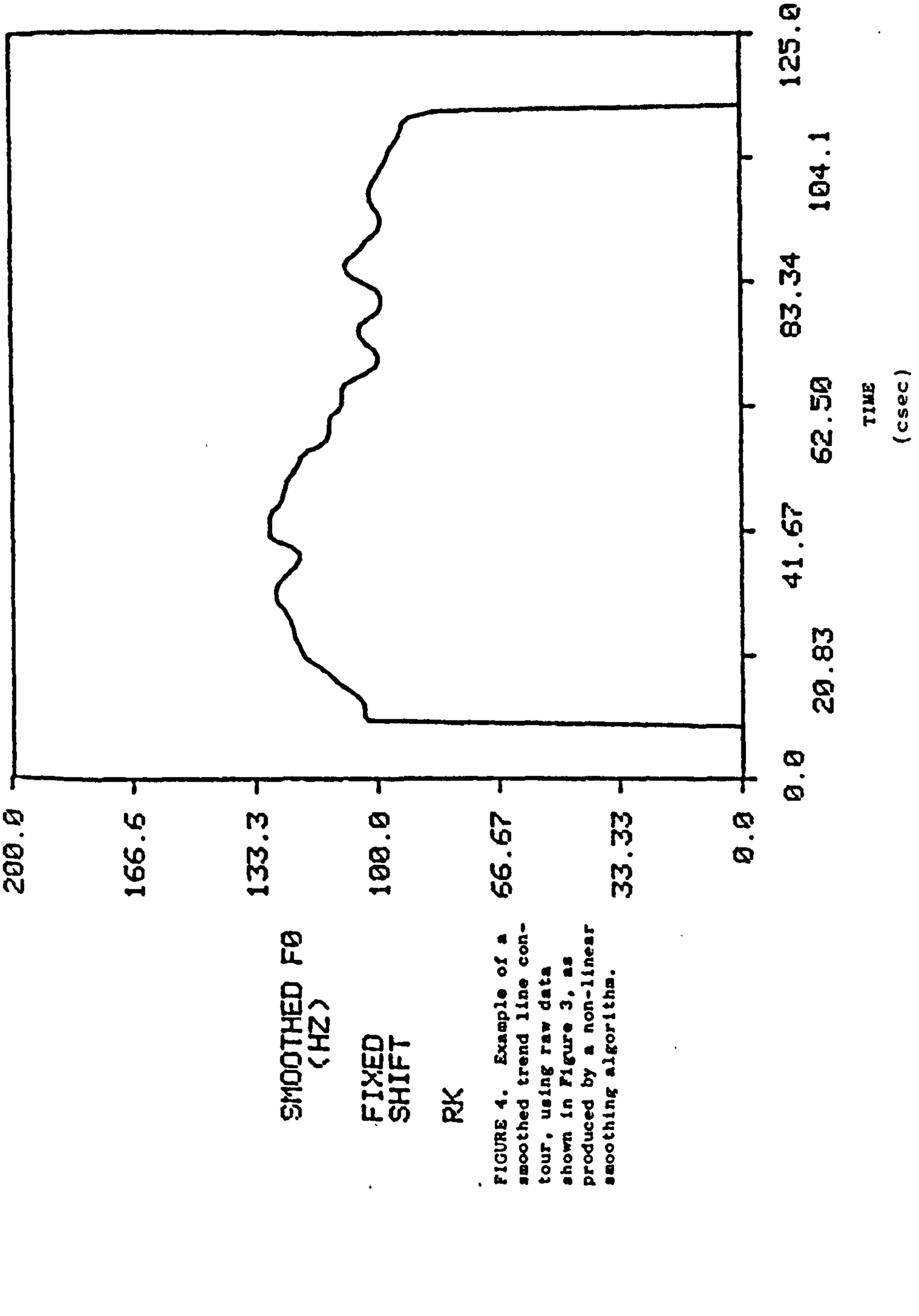


FIGURE 4. Example of a smoothed trend line contour, using raw data shown in Figure 3, as produced by a non-linear smoothing algorithm.

Figures 3 and 4 demonstrate an example of the smoothing of a raw FO contour by the non-linear smoother. Figure 3 is the raw FO contour derived by the parallel pitch period estimator for a small section of the stimulus utterance produced by a normal speaker (RK). This contour is characterised by a normal range of FO values for a male speaker as well as small irregularities in the contour typical of normal phonation. Figure 4 is the equivalent FO contour produced by the smoothing process. This smoothed trend line retains the over-all intonational features of the original raw FO contour but the small irregularities have been removed, thus making the trend line a useful base from which to measure the irregularities.

2.2.3. Perturbation measures

Several statistical measures are determined for each FO contour which describe the magnitude, distribution, and frequency of micro-perturbatory behavior in each sample.

AVEX - The average magnitude of the excursions present in each FO contour, described in units of Hz or percent.

SDEVEX - The standard deviation of the distribution of the excursions present in each FO contour, described in units of Hz or percent.

RATEX - The Rate of Excursions is the percentage of points in the sample where a magnitude excursion in percent is equal to or greater than a pre-set threshold. RATEX is adapted from the "Pitch Perturbation Quotient" (PPQ) of Koike, Takahashi, and Calcaterra (1977) - RATEX differs from PPQ in that a non-linear smoother is used to produce a smoothed FO trend rather than the moving average approach used to calculate PPQ. The non-linear smoother preserves major features of the FO contour while smoothing out noisy and anomalous components in the contour. RATEX is based on magnitude excursions in percent in order to normalize excursion measures calculated for varying FOs evidenced within and between speakers' phonations. The pre-set threshold is used to quantify the number of significant perturbations in any given speech sample (similar to Lieberman's (1963) minimum threshold for his "Perturbation Factor"). The pre-set threshold is set to 3%, because even in the healthiest voice, uttering a monotone vowel, the successive pitch periods typically show approximately 2% frequency jitter, in a normal distribution (Hanson, 1978). A 3% threshold allows us to discount this factor. Thus RATEX reflects the incidence of significant excursions in the sample.

DPF - The Directional Perturbation Factor which has been adapted from Hecker and Kreul (1971). DPF is the percentage of changes of algebraic sign calculated for differences between adjacent raw FO measures (thus not based on a smoothed trend line). A 3% threshold for the magnitude of the difference between adjacent FOs is also included in this measure to exclude the normal distribution of FO differences.

ANOMALIES - This category includes both short discontinuities, as defined earlier, and anomalous FOs outside the pre-set range for acceptable frequencies (i.e., 40-240 Hz for males, 75-450 Hz for

females). All such anomalies are rejected from the perturbation calculations, but their occurrence flagged.

OUT OF RANGE - The total number of occasions when, in the variable shift calculation, the projected value of the incoming period fell outside the pre-set limits for acceptable FO values.

2.2.4. Intonation measures

Several measures of overall intonational behavior are calculated for each utterance based on the smoothed trend line. These measures include the mean, median, mode, and standard deviation of the FO distribution, limited to the pre-set limits for acceptable frequencies.

2.2.5. Application of the perturbation algorithm

The perturbation algorithm was applied to the FO contours extracted from three normal voices (RK, JL, SH) and one very pathological voice (MA2/RIE12), each speaker having produced the test sentence 'A rainbow is a division of white light into many beautiful colours'. Tables III and IV present the resultant perturbation and intonational measures for each of the voices - Table III contains data derived via a speaker-specific fixed shift factor and Table IV displays data derived via a speaker-specific variable shift factor. Inspecting the perturbation measures AVEX, SDEVEX, RATEX, and DPQEX, it can be seen that a clear separation of the pathological speaker from the normal speakers exists. Similar results were noted for the ANOMALIES and OUT OF RANGE measures between the pathological and normal speakers.

Figures 5, 6, 7 and 8 display examples of FO intonational and perturbational distributions produced by the perturbation programs for one normal speaker (RK) and one pathological speaker (MA2/RIE12). All data presented in these figures were derived via the speaker-specific fixed shift method of pitch period estimation. Figures 5 and 6, for the normal and pathological speakers respectively, are histograms of the long term FO intonational behavior based on the smoothed trend line output of the non-linear smoother. The FO histogram of the normal speaker (mean = 109.2 Hz) in Figure 5 shows a distribution which is more normally distributed, narrower and more peaked compared to the FO distribution of the pathological speaker (mean = 126.2 Hz). The FO histogram of the pathological speaker shows a bimodal distribution. Figures 7 and 8 are histograms of the short term perturbational data based on the signed magnitude of the excursions in Hz for the two speakers. These two figures also demonstrate substantial differences for the phonatory behavior of the two speakers with a much narrower and more peaked distribution for the normal speaker compared to the pathological speaker. The differences between the two perturbational distributions are reflected in the greater AVEX, SDEVEX, and RATEX measures of the pathological speaker compared to the normal speaker.

CONCLUSION

Having developed a successful pitch detection algorithm, and plausible measures of perturbation, the next stage of the project is to apply these to an extensive set of voices. These fall into two categories. The first involves the recorded voices of patients

Subject	RK N=228, CTX=y6	JL N=300, CTX=y3	SH N=283, CTX=y0	MA2/RIE12 N=848, CTX=70
AVEX	3.57 Hz 3.24 %	3.38 Hz 3.65 %	4.85 Hz 3.48 %	16.20 Hz 12.26 %
SDEVEX	11.y6 Hz 10.48 %	7.y7 Hz y.12 %	15.08 Hz y.86 %	24.42 Hz 18.02 %
RATEX	14.04 %	21.33 %	20.4y %	52.12 %
DPQEX	8.41 %	20.13 %	15.0y %	34.01 %
ANOMALIES	5	2	6	42
F0 MEAN	107.20 Hz	105.60 Hz	115.20 Hz	126.20 Hz
F0 MEDIAN	yy.00 Hz	104.y0 Hz	113.00 Hz	128.30 Hz
F0 SD	13.11 Hz	13.6y Hz	18.71 Hz	37.57 Hz

TABLE III

Automatic F0 perturbation analysis for three normal male voices (RK, JL, SH) and one dysphonic male voice (MA2/RIE 12), using a FIXED shift factor.

Subject	RK N=225	JL N=2y3	SH N=306	MA2/RIE12 N=726
AVEX	2.66 Hz 2.2y %	3.y4 Hz 3.y3 %	4.5y Hz 3.y6 %	14.61 Hz 10.62 %
SDEVEX	7.82 Hz 6.55 %	10.18 Hz 10.20 %	12.24 Hz 10.7y %	22.81 Hz 16.15 %
RATEX	15.56 %	20.48 %	23.20 %	48.35 %
DPQEX	8.70 %	17.85 %	13.11 %	35.47 %
ANOMALIES	1	1	10	35
OUT OF RANGE	2	0	5	y3
F0 MEAN	10y.20 Hz	107.50 Hz	11y.50 Hz	126.y0 Hz
F0 MEDIAN	yy.20 Hz	105.y0 Hz	117.30 Hz	128.30 Hz
F0 SD	15.47 Hz	13.45 Hz	1y.60 Hz	37.35 Hz

TABLE IV

Automatic F0 perturbation analysis for three normal male voices (RK, JL, SH) and one dysphonic male voice (MA2/RIE 12), using a VARIABLE shift factor.

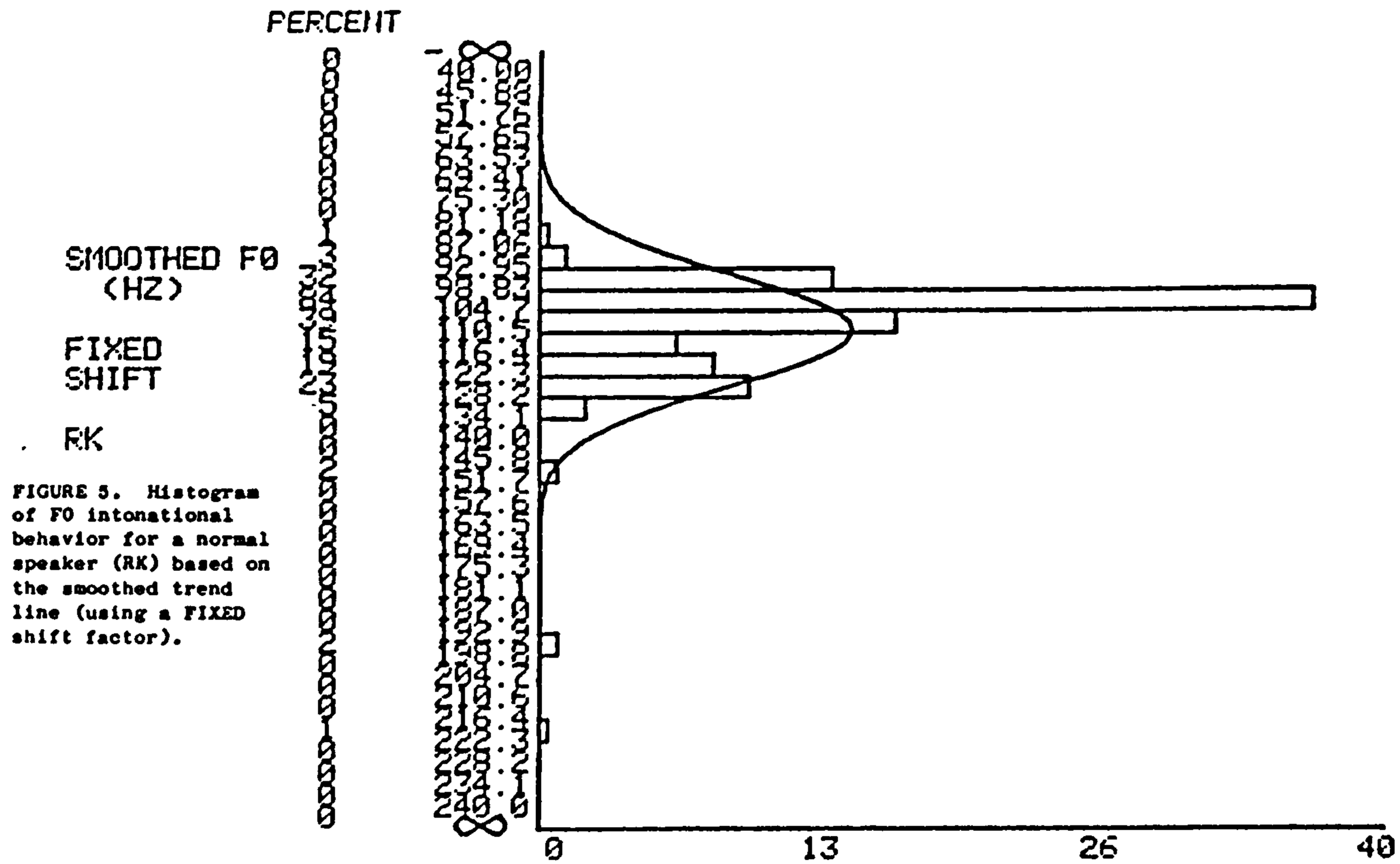


FIGURE 5. Histogram of F0 intonational behavior for a normal speaker (RK) based on the smoothed trend line (using a FIXED shift factor).

STATISTICS OF	40.00 TO	240.0	
MINIMUM	83.55		
MAXIMUM	222.2		
AVERAGE	109.2		
STANDARD DEV	15.47	MEDIAN =	99.02
NUMBER OF POINTS	226	MODE =	101.9

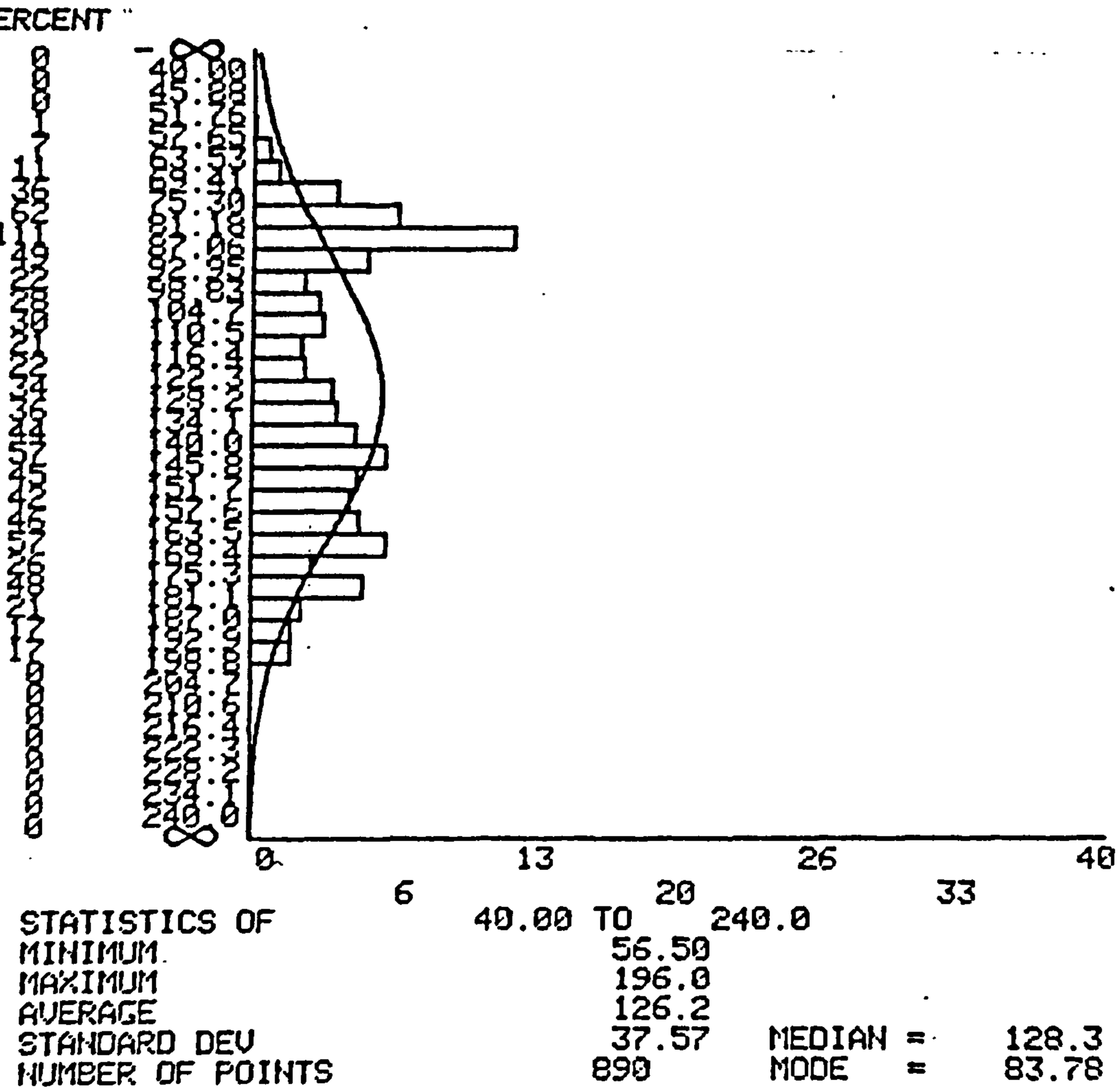


FIGURE 6. Histogram of F0 intonational behavior for a pathological speaker (MA2/RIE12), based on the smoothed trend line (using a FIXED shift factor).

STATISTICS OF	40.00 TO 240.0		
MINIMUM	56.50		
MAXIMUM	196.0		
AVERAGE	126.2		
STANDARD DEV	37.57	MEDIAN =	128.3
NUMBER OF POINTS	890	MODE =	83.78

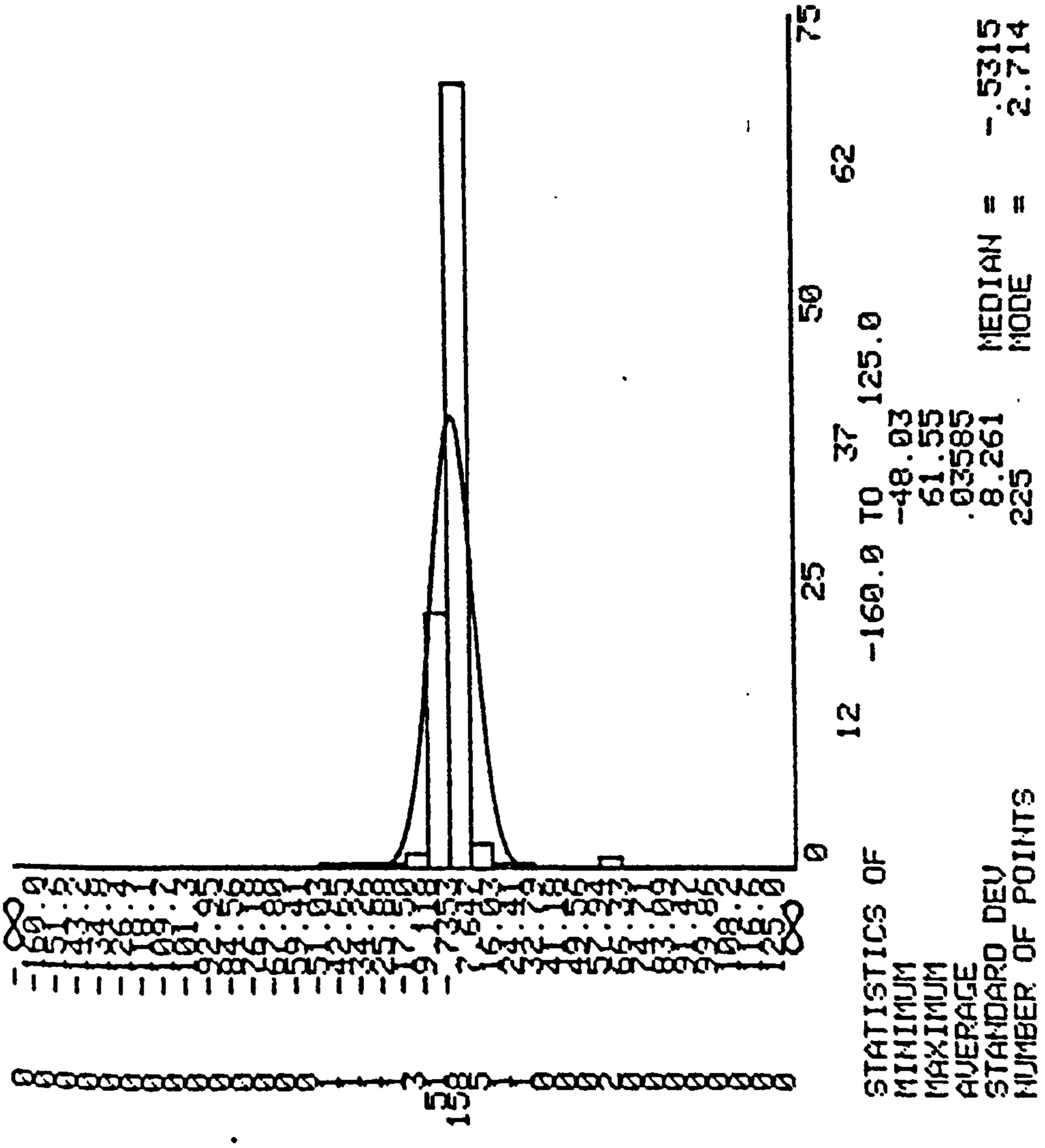


FIGURE 7. Histogram of signed excursions in Hz for a normal speaker (RK) using a FIXED shift factor.

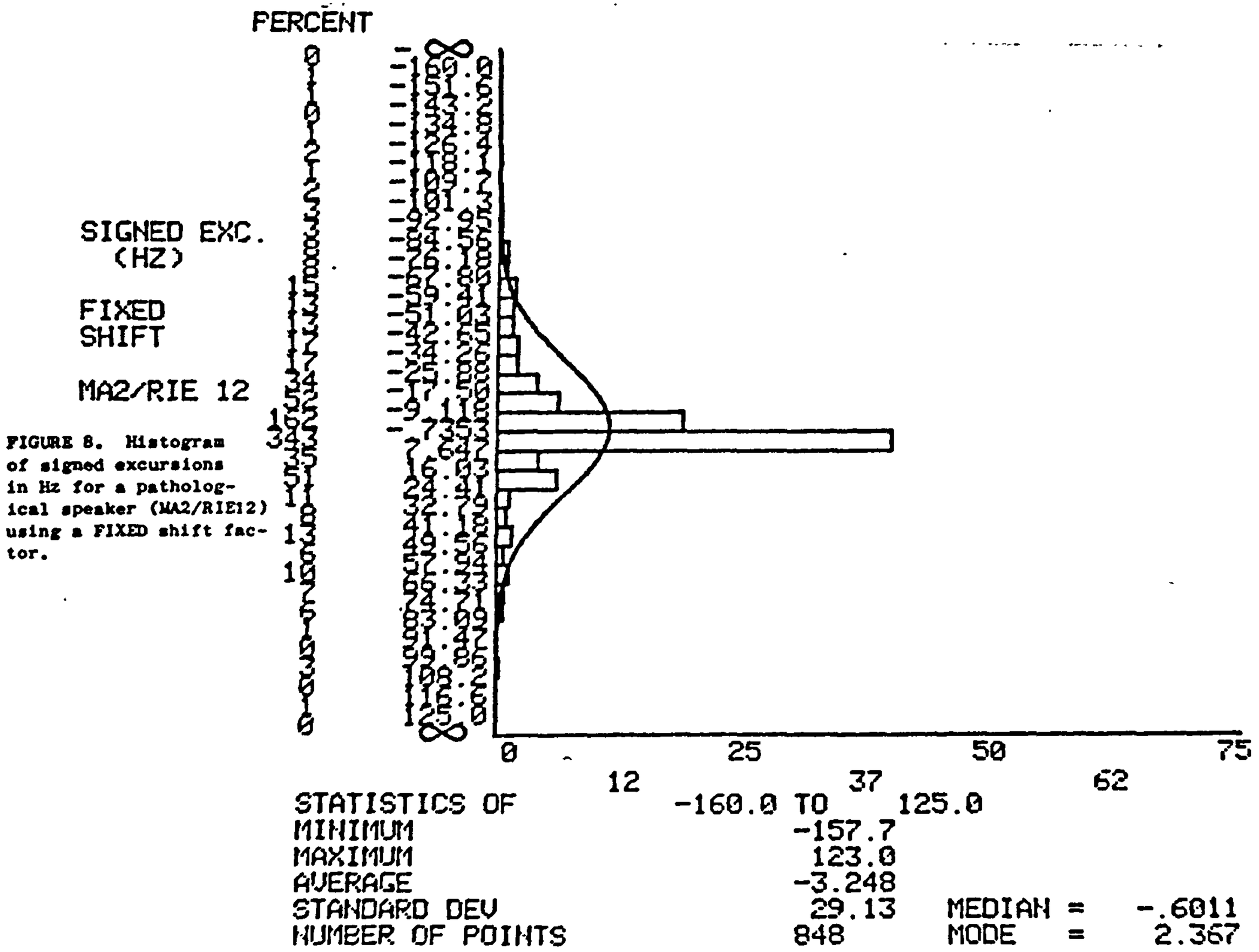


FIGURE 8. Histogram of signed excursions in Hz for a pathological speaker (MA2/RIE12) using a FIXED shift factor.

from voice clinics, for which a wide range of diagnostic information about their vocal pathology will be made available. The two main collaborating institutions for this part of the project are the Otolaryngology Departments of the Radcliffe Infirmary, Oxford, and the Royal Infirmary, Edinburgh. The project will seek to correlate acoustic perturbational data with the type and degree of pathology present, as discussed in more detail in Mackenzie, Laver and Miller (1983: see this volume). The second category consists of a control group of some hundred voices of each sex. A fairly clear picture is available of most of the acoustic characteristics of the normal voice (Laver and Hanson, 1981), but this does not yet include a full knowledge of typical ranges of perturbation in the healthy voice. This is needed to establish the phonatory norm from which pathological voices can be held to deviate.

The hypothesis underlying the work of the project is that increasing perturbation, beyond a threshold yet to be established, reflects increasingly severe pathology. This hypothesis will obviously have to be refined, and the range of perturbation which characterises stages of different pathologies will have to be made more specific, but as a preliminary conceptual step it seems profitable to distinguish between two general levels of perturbation. The first of these is the range of perturbation that characterises the normal, healthy larynx: we can refer to perturbation in this range as being "microperturbation". The second is the range of perturbation that characterises the unquestionably pathological larynx: we can call this more extreme type "macroperturbation". As an initial estimate, the threshold for passing from microperturbation to macroperturbation possibly lies somewhere in the range between 30 to 40% RATEX, with an associated AVEX of 10% or more and SDEVEX of 15% or more - i.e., where roughly between a third or more of all individual periods in phonation deviate substantially and variably from the local smoothed trend line.

Given that our interest is in screening the general population for potential laryngeal pathology, rather than only in quantifying the phonatory consequence of unquestionable pathology, it is the border zone towards the end of the microperturbatory range, up to the threshold of definitely pathological macroperturbation, that attracts our attention. This is the zone of perturbation where, within the frame of reference of a screening system, an individual subject can be held to be 'at risk', as indicated in Figure 9. This 'risk zone' is where early signs of pathology will surface, we speculate. It may well be that the phonation of a given speaker found to be in the risk zone will be one where the relatively high degree of microperturbation shown is due to the dysperiodic symptoms of a particular habitual but healthy phonation type, such as creaky voice (vocal fry), rather than of pathology. But false alarms of that sort are the price one pays for the benefit of a screening system designed to catch symptoms of vocal pathology as early as possible. A major part of our empirical research will consist of tuning the boundaries of the risk zone as far as possible to reduce false alarms and maximize the early detection of laryngeal pathology. This tuning process will include the investigation of the differential power of the perturbation measures to distinguish between the populations of normal and pathological speakers.

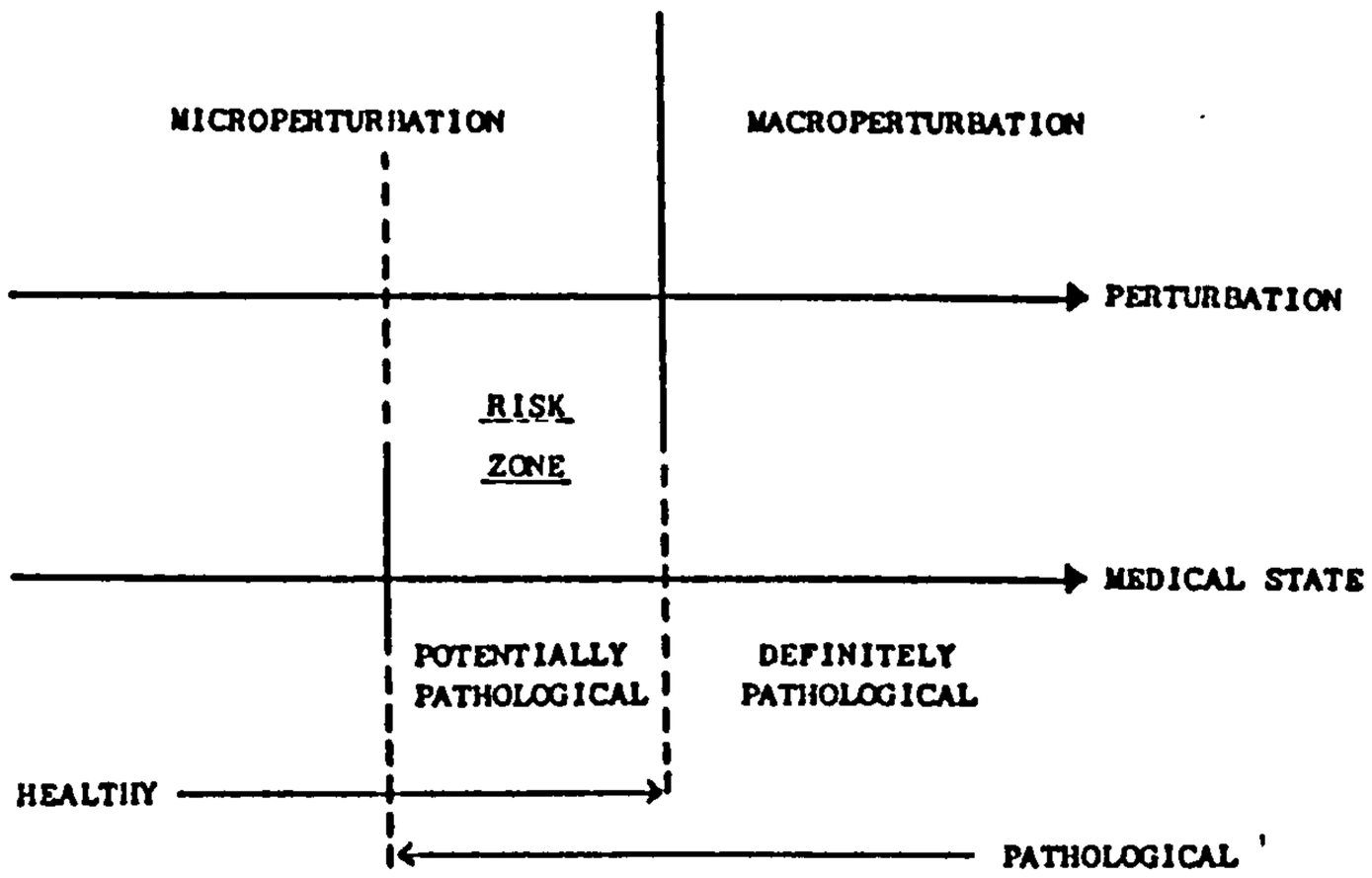


FIGURE 9. A schematic diagram of the relationship between waveform perturbation and vocal fold pathology.

REFERENCES

Askenfelt, A. and Hammarberg, B. (1980) 'Speech waveform perturbation analysis'. Speech Transmission Laboratory Quarterly Progress and Status Report, 4, 40-49.

---------- (1981) 'Speech waveform perturbation analysis revisited'. Speech Transmission Laboratory Quarterly Progress and Status Report, 4, 49-68.

Davis, S. B. (1976) 'Computer evaluation of laryngeal pathology based on inverse filtering of speech'. Speech Communication Research Laboratory Monograph, 13.

Fourcin, A. J. (1974) 'Laryngographic examination of vocal fold vibration'. In B. Wyke (ed.), Ventilatory and Phonatory Control Mechanisms. London: Oxford University Press, 3-5-333.

Gold, B. (1962) 'Computer program for pitch extraction'. J. Acoust. Soc. Am., 34, 442-448.

---------- (1964) 'Note on buzz-hiss detection'. J. Acoust. Soc. Am., 36, 1659-1661.

Gold, B. and Rabiner, L. R. (1969) 'Parallel processing techniques for estimating pitch periods of speech in the time domain'. J. Acoust. Soc. Am., 46, 442-448.

Hanson, R. J. (1978) 'A two-state model of FO control'. J. Acoust. Soc. Am., 64, 543-544.

Hecker, M. and Kreul, E. (1971) 'Descriptions of the speech of patients with cancer of the vocal folds. Part 1: Measures of fundamental frequency'. J. Acoust. Soc. Am., 49, 1275-1282.

Horii, Y. (1979) 'Fundamental frequency perturbation observed in sustained phonation'. J. Speech and Hearing Res., 22, 5-19.

Kitajima, K. and Gould, W. J. (1976) 'Vocal shimmer in sustained phonations of normal and pathological voices'. Annals. Otol., 85, 377-381.

Kitajima, K., Tanabe, M., and Isshiki, N. (1975) 'Pitch perturbations in normal and pathological voice'. Studia Phon., 9, 25-32.

Koike, Y. (1973) 'Application of some acoustic measures for the evaluation of laryngeal dysfunction'. Studia Phon., 7, 17-23.

Koike, Y., Takahashi, H., and Calcaterra, T. C. (1977) 'Acoustic measures for detecting laryngeal pathology'. Acta Otolar., 84, 105-117.

Laver, J. (1967) 'The synthesis of components in voice quality'. Proceedings of the VI International Congress of Phonetic Sciences, 523-535, Czechoslovak Academy of Sciences, Prague.

Laver, J. (1968) 'Voice quality and indexical information'. Brit. J. Disorders Comm., 3, 43-54.

---------- (1974) 'Labels for voices'. J. Inter'l. Phonetic Assoc., 4, 62-75.

---------- (1975) Individual features in voice quality. Doctoral dissertation, University of Edinburgh.

---------- (1979) Voice Quality : a Classified Bibliography. Amsterdam: John Benjamins B.V.

---------- (1980) The Phonetic Description of Voice Quality. Cambridge: Cambridge University Press.

Laver, J. and Hanson, R. J. (1981) 'Describing the normal voice'. In J. Darby (ed.), Speech Evaluation in Psychiatry. New York: Grune & Stratton, 51-78.

Laver, J., Hiller, S. M., and Hanson, R. J. (1982) 'Comparative performance of pitch detection algorithms on dysphonic voices'. Proceedings of IEEE Conference on Acoust., Speech, and Signal Proc., 192-195.

Laver, J., Wirz, S., Mackenzie, J. and Hiller, S. M. (1981) 'The perceptual protocol for the analysis of vocal profiles'. Work in Progress, Department of Linguistics, Edinburgh University, No. 14:139-155.

Laver, J., Wirz, S., Mackenzie, J. and Hiller, S. M. (1982) Vocal profiles of speech disorders. Final Report on MRC Grant No. 9781192N, University of Edinburgh.

Laver, J., Wirz, S., Mackenzie, J. and Hiller, S. M. (forthcoming 1984) Vocal Profiles. Cambridge University Press.

Lieberman, P. (1961) 'Perturbations in vocal pitch'. J. Acoust. Soc. Am., 33, 597-603.

---------- (1963) 'Some acoustic measures of the fundamental frequency periodicity of normal and pathological larynges'. J. Acoust. Soc. Am., 23, 361-363.

Mackenzie, J., Laver, J., and Hiller, S. M. (1983) 'Structural pathologies of the vocal folds and phonation'. Work in Progress, Department of Linguistics, Edinburgh University, No. 16:80-116.

Rabiner, L. R., Cheng, M. J., Rosenberg, A. E., and McGonegal, C. A. (1976) 'A comparative performance study of several pitch detection algorithms'. IEEE Trans. Acoust., Speech and Signal Proc., ASSP-22, 552-557.

Rabiner, L. R., Sambur, M. R., and Schmidt, C.E. (1975) 'Applications of a non-linear smoothing algorithm to speech processing'. IEEE Trans. Acoust., Speech and Signal Proc., ASSP-22, 552-557.

Rabiner, L. R. and Schafer, R. W., (1978) Digital Processing of Speech Signals. New Jersey: Prentice-Hall, Inc.

APPENDIX FIVE

Journal of Phonetics (1986) 14, 517–524

An acoustic screening system for the detection of laryngeal pathology

John Laver, Steven Hiller, Janet Mackenzie and Edmund Rooney

Centre for Speech Technology Research, Department of Linguistics, University of Edinburgh, U.K.

1. Introduction

This project has two main aims: the development of a computer-based system of acoustic analysis which can screen voices for the presence of laryngeal pathologies; and the differentiation of such pathologies using acoustic measures alone. A system based on measurement of fundamental frequency and waveform perturbations has been developed (Hiller, Laver & Mackenzie, 1983, 1984; Laver, Hiller & Mackenzie, 1984). This paper is a discussion of possible procedures for distinguishing a group of speakers with known pathologies from a large control group, as a prelude to the development of screening techniques.

An automatic system which can detect possible laryngeal pathology has several potential applications.

(1) *Screening* of an unselected population, alongside existing screening programmes in hospitals, "well-man/well-woman" clinics, etc. An acoustic system has the advantage of being completely non-invasive, and the recording procedure is simple, causes minimal distress to subjects and is highly portable (so that screening could be extended to schools, factories, etc.).

(2) *Assessment of priorities* among a preselected population, consisting of patients already complaining of hoarseness, or those visiting their GPs with voice problems. The use of an acoustic system could speed the process of referral for laryngeal examination where the possibility of serious pathology was indicated.

(3) *Diagnostic support* where a particular laryngeal pathology is already suspected. This depends on the discriminability of the various pathologies using acoustic measures.

(4) *Longitudinal monitoring* to assess change in phonatory efficiency in patients undergoing treatment (surgery, speech therapy, radiotherapy or chemotherapy), or to track deterioration in progressive disease.

2. Acoustic system

The analysis system, implemented on a VAX 11/750 computer, produces measurements of fundamental frequency (F_0) and waveform perturbations in approximately 40 s of recorded text read from the "Rainbow Passage" (Fairbanks, 1960). The measurement system uses an elaborated version of the Gold & Rabiner (1969) parallel processing pitch detection algorithm, with phase compensation for low-frequency distortion introduced by tape recording techniques; low-pass filtering to remove higher frequency resonance effects from the waveform (600 Hz for males, 800 Hz for females); non-linear smoothing

0095–4470/86/030517 + 08 $03.00/0

© 1986 Academic Press Inc. (London) Ltd.

to derive an intonational "trendline" from the raw pitch period estimates; and parabolic interpolation at waveform peaks to provide greater resolution of pitch period values (Hiller *et al.*, 1983, 1984).

Intonational data are derived from the smoothed F_0 trendline, giving its mean value (F_0-AV) and its range, represented as the standard deviation of the trendline values (F_0-DEV). Statistical analyses are then made of pitch period perturbation (jitter) and amplitude perturbation at waveform peaks (shimmer). The following measures are taken for both jitter (*J*) and shimmer (*S*).

(1) Average magnitude of excursions of the raw F_0 contour from the local trendline (AVEX).

(2) Standard deviation of (signed) excursions from trendline (DEVEX).

(3) Rate of excursions (RATEX): this is the percentage of points in the sample where the magnitude of excursions is greater than or equal to 3% of the local trendline value. A value of 3% was chosen because even the healthiest of voices, performing monotone, steady-state vowels, typically shows a level of (jitter) perturbation of about 2% (Hanson, 1978).

(4) Directional perturbation factor (DPF). This measure, adapted from Hecker & Kreul (1971), is the percentage of changes in algebraic sign between adjacent pitch or amplitude estimates in the raw contours. A 3% threshold is also applied to this measure.

3. Subjects and data collection

The collection of data on pathological subjects has been made possible by collaboration with the ENT departments of the Radcliffe Infirmary, Oxford and the Royal Informary, Edinburgh. One hundred and nine speakers whose laryngeal state had been established by medical examination in these departments were recorded on high quality analogue recorders (Revox A77 and Uher 4000). The first 40 s of each speech sample were then digitized at 20 kHz, and analysed using the above acoustic system. A control group of 121 speakers was recorded and analysed in the same way. It has not been possible to subject the control speakers to a laryngeal examination, but none reported any history of laryngeal disorder or other relevant complaint.

Table I gives details of each group used, including the percentage of self-reported smokers at the time of recording. Speakers from the control group are in general younger than those of the pathological group, but this bias will be rectified as the control group nears its target of 200 speakers (100 of each sex).

The speakers from the pathological group show a wide variety of laryngeal disorders. Table II presents a summary of the types of disorder present in the pathological group.

TABLE I. Subject data by group (n = 230)

Group	Sex	Number	Age range (mean)	% Smokers
Control	M	63	18–63 (31.7)	17.5
Control	F	58	18–73 (28.7)	17.2
Pathological	M	55	25–82 (53.9)	27.3
Pathological	F	54	24–75 (53.2)	44.4

TABLE II. Classification of laryngeal disorders diagnosed in pathological group and number of cases ($n = 109$)

Type of Pathology	Males	Females
Disorders of the ligamental area		
Epithelial disorders (e.g. carcinoma, papilloma, keratosis)	17	2
Reinke's oedema	0	4
Polyps, nodules	8	22
Cysts	2	2
Miscellaneous mild oedema, redness, laryngitis	11	15
Disorders of the cartilaginous area	8	5
Palsies	8	4
Supra-glottic lesions	1	0
Total	55	54

4. Group separation and screening procedures

The two groups of subjects may be expected to show a certain amount of internal diversity. The pathological speakers evidence a variety of disorders (as shown in Table II), each of which may have different effects on the structural—and hence vibratory —properties of the vocal folds (Mackenzie, Laver & Hillier, 1983). The control group is, it is hoped, more homogeneous, but could contain speakers with undetected laryngeal pathologies or functional disorders. Some overlap between the groups' phonatory behaviour is possible, then, but in general they should be separable if a screening procedure is to be feasible.

The project has considered a number of approaches to demonstrating the separation of the groups, with a view to developing screening tools. These include:

(1) a simple graphic approach, showing the relation between the groups on bi-variate plots, with a plausible screening boundary to separate them,

(2) a multivariate statistical technique (linear discriminant analysis) as a means of using data from all 10 parameters simultaneously.

4.1. *Bivariate plots*

Our bivariate plots have a screening boundary derived from principal components analysis. This approach has the advantage of allowing the relationship between the two groups—and that of individual patients to the control group—to be easily visualised. In order to facilitate the comparison of pathological subjects and controls, all subjects' scores were transformed to Z-scores and expressed as multiples of the *control standard deviation* from the *control group mean* for their sex. Given that two standard deviations on any one parameter should include approximately 90–95% of control subjects (assuming normal distributions), any subject whose score on a given parameter deviates from the control group mean by more than two SD may be considered to be at risk of pathology. Table III presents the numbers of subjects in each group (pathological and control) who deviate from the control group mean by more than two SD on each parameter in turn.

TABLE III. Subjects deviating from control group mean for each parameter by more than 2 SD (figures in parentheses are percentages)

	Males		Females	
Parameters	Pathological	Control	Pathological	Control
F_0-AV	12 (21.8)	3 (4.7)	12 (22.2)	2 (3.4)
F_0-DEV	7 (12.7)	2 (3.2)	8 (14.8)	5 (8.6)
J-DEVEX	14 (25.4)	2 (3.2)	15 (27.7)	2 (3.4)
J-AVEX	21 (38.2)	3 (4.7)	16 (29.6)	3 (5.2)
J-RATEX	3 (5.4)	2 (3.2)	3 (5.5)	3 (5.2)
J-DPF	34 (61.8)	2 (3.2)	31 (57.4)	2 (3.4)
S-DEVEX	25 (45.4)	4 (6.3)	20 (37.0)	2 (3.4)
S-AVEX	28 (50.1)	6 (9.5)	22 (40.7)	3 (5.2)
S-RATEX	16 (29.1)	3 (4.7)	11 (20.4)	1 (1.7)
S-DPF	42 (76.4)	2 (3.2)	35 (64.8)	1 (1.7)

On this basis, no single parameter of the ten distinguishes between the two groups sufficiently for the purposes of screening; but a combination of two parameters—one F_0 parameter and one perturbation parameter, for example—is more successful (Laver *et al.*, 1984).

Figure 1 shows 55 male patients with known structural pathologies of the larynx plotted on a scattergram of mean F_0 versus shimmer DPF. The axes are marked in multiples of standard deviation, and the origin of both axes corresponds to the control group mean for each parameter. S-DPF was the best single discriminator between the groups for both sexes. F_0-AV was included because of the possibility that some pathological subjects may be able to maintain normal levels of perturbation (as represented by S-DPF) by boosting laryngeal tension, at the expense of slightly higher than normal mean F_0 (Mackenzie *et al.*, 1984). Principal components analysis was applied to the control group data to give an ellipse (at the 2 SD level) indicating the covariance between the parameters. The boundary of this ellipse forms the screening threshold for the detection of pathology.

Fifty (90.1%) of the pathological males fall outside the ellipse, and would be recognized as pathological by this approach. Six (9.5%) control males fall outside, and register as false positives. It is worth noting that two of the pathological males who fail to be detected have epithelial disorders. Both, however, are cases of keratosis with oedema.

For the females, 43 (79.6%) pathological subjects fall outside the ellipse, with six false positives (10.3%). Both female patients with epithelial disorders are successfully detected.

4.2. *Linear discriminant analysis*

Linear discriminant analysis (Klecka, 1980) is a statistical technique for discriminating between two (or more) nominal groups on the basis of several parameters simultaneously. A *discriminant function* is derived by weighting and combining the parameters in such a way that the groups will be maximally separated by their members' scores on this function.

The data are first assessed to see whether there is enough difference between the groups on these parameters to justify the analysis proposed. This is done by computing Wilks'

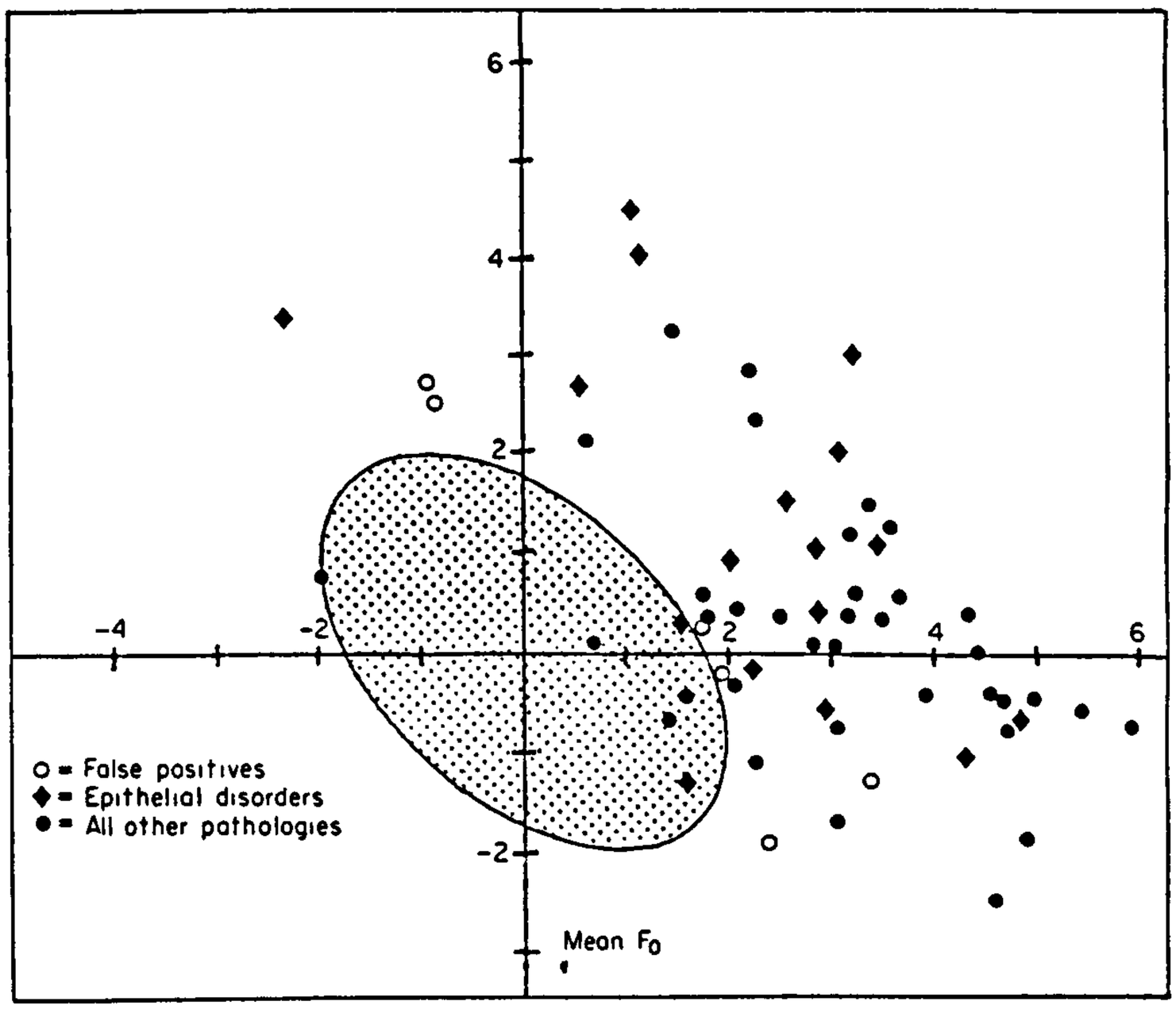


Figure 1. A scattergram of DPF shimmer vs. mean F_0 for male speakers. The shaded area represents a 2 SD ellipse derived from principal components analysis of male controls.

λ (an inverse measure of group separation), with an associated χ^2 test of statistical significance. A significant Wilks' λ at this stage implies that the first discriminant function to be derived will itself be statistically significant. The *substantive* utility of the function can be measured by its canonical correlation: that is, the association between the function and the nominal categories representing the groups present in the data. A high canonical correlation (0.7 or upwards) indicates that the function is discriminating quite successfully between the named groups. The discriminant scores calculated for each subject can then be used to classify the subjects, allowing an additional measure of the usefulness of the function: its rate of success in allocating subjects to their correct groups.

One discriminant function, separating pathological subjects from controls, was derived for each sex separately, from subjects' raw (unstandardized) scores on all 10 parameters, using the DISCRIMINANT subprogram available in the *Statistical Package for the Social Sciences* (1983). Table IV gives the resulting classifications for the males and females separately, along with the canonical correlation coefficient for each function and the Wilks' λ calculated before the derivation of that function, with a χ^2 test of statistical significance. Both functions were highly significant. Two of the incorrectly classified pathological males have epithelial disorders: one was a case of keratosis with oedema (one of the cases referred to above), the other a very early case of squamous carcinoma (undetected at the time of recording).

It is not expected that all 10 parameters will be equally useful for discriminating between pathological and control subjects: some do not separate the groups very well, while others are redundant by virtue of their high correlation with those that do. The

TABLE IV. Classification of male and female subjects into pathological and control groups by discriminant functions derived from all 10 acoustic parameters (figures in parentheses are percentages)

	Correct classifications	Incorrect classifications
A. Males*		
Pathological (55)	47 (85.5)	8 (14.5)
Control (63)	58 (92.1)	5 (7.9)
B. Females†		
Pathological (54)	47 (87.0)	7 (13.0)
Control (58)	55 (94.8)	3 (5.2)

*Wilks' λ before function = 0.299; $\chi^2 = 133.9$; $p < 0.0001$; canonical correlation = 0.837.

†Wilks' λ before function = 0.366; $\chi^2 = 105.67$; $p < 0.0001$; canonical correlation = 0.797.

TABLE V. Standardized discriminant function coefficients for each of the functions described in Table IV

Males		Females	
S-DPF	1.66889	S-DPF	2.05087
F_0-AV	0.96325	J-DEVEX	−1.19266
S-RATEX	−0.68534	F_0-DEV	0.78410
J-AVEX	0.54630	S-RATEX	−0.70417
J-DEVEX	−0.52852	J-DPF	−0.58023
F_0-DEV	−0.38440	J-RATEX	0.52369
J-RATEX	0.11701	F_0-AV	−0.39136
S-AVEX	0.08795	S-DEVEX	0.29108
J-DPF	−0.04024	S-AVEX	−0.19908
S-DEVEX	−0.01797	J-AVEX	0.14846

relative contribution of each individual parameter to the function can be learned from the absolute values of the (standardized) weighting coefficients produced by the program. These are given in Table V, in order of importance, for both functions.

It is clear that S-DPF is by far the most important contributor to both functions. It also seems, from measurements of the correlation between individual parameters and the discriminant functions (given in Table VI), that these functions are essentially functions of *perturbation*. The failure of some of the perturbation measures to achieve high weightings in the functions can perhaps be attributed to high degrees of intercorrelation among them.

4.3. *Reservations*

The results of this discriminant analysis need to be treated with caution. Linear discriminant analysis assumes that the data show a multivariate normal distribution, but given the heterogeneous composition of the pathological group it is likely that this assumption is seriously violated in this case. However, the technique is quite robust in the face of such violations. A more serious problem is the fact that the groups are still rather small, given the number of parameters being used to derive the functions, and it is therefore not

TABLE VI. Pooled within-groups correlation coefficients between parameters and each discriminant function

Males		Females	
S-DPF	0.67050	S-DPF	0.66738
S-RATEX	0.45745	S-RATEX	0.47688
J-DPF	0.37772	J-DPF	0.26889
J-RATEX	0.27290	S-AVEX	0.24010
S-AVEX	0.26127	F_0-AV	−0.21667
J-AVEX	0.21225	J-RATEX	0.16737
F_0-AV	0.16901	J-AVEX	0.13843
F_0-DEV	0.15553	J-DEVEX	0.06467
J-DEVEX	0.12085	F_0-DEV	0.05434
S-DEVEX	0.01593	S-DEVEX	0.04271

possible to put great reliance on the functions obtained, despite their statistical significance. It must be remembered that a function derived for a set of data is an *optimal* one, designed to force the groups as far apart as possible. Success in achieving a high degree of separation is then a descriptive measure of structure in the actual data set. The classification rates obtained, however, cannot safely be asserted to be necessarily predictive of future success in classifying another set of subjects with the same function. The recommended procedure for testing a function's true discriminating power is to split the sample, deriving the function from, say, half of the subjects (randomly selected) and measuring its success in classifying the remainder. However, this will not be possible until the groups are larger.

It was also felt inappropriate at this stage to attempt to derive an optimal set of parameters for discrimination, despite clear indications that certain parameters (especially S-DPF) were more useful than others.

5. Conclusions

The two principal objectives of the project are (1) the development of a screening system, and (2) the differentiation of disorders. The separation of the two groups of subjects —and the feasibility of screening—have been clearly demonstrated using two techniques, both of which form potential screening tools. Work is continuing into assessing the acoustic consequences of different pathologies, but the use of a technique such as discriminant analysis for the task of differentiation, though promising, cannot be attempted without considerably larger numbers of subjects in each pathology group. The potential applications of the system to assessing priorities among patients, and monitoring progress or deterioration, remain to be examined.

This project is funded by the Medical Research Council (Grant No. 8207136N: 1982–1985).

We are very grateful for the collaboration and co-operation of Mr T. Harris (Department of Otolaryngology) and Mrs S. Collins (Department of Speech Therapy), of the Radcliffe Infirmary, Oxford; and Mr A. Maran (Department of Otolaryngology), and Mrs M. Mackintosh and Mrs R. Nieuwenhuis (Department of Speech Therapy), of the Royal Infirmary, Edinburgh.

References

Fairbanks, G. (1960) *Voice and articulation drill book*. New York: Harper Row.

Gold, B. & Rabiner, L. (1969) Parallel processing techniques for estimating pitch periods of speech in the time domain, *Journal of the Acoustical Society of America*, **46**, 442–448.

Hanson, R. (1978) A two-state model of F_0 control, *Journal of the Acoustical Society of America*, **64**, 543–544.

Hecker, M. & Kreul, E. (1971) Descriptions of the speech of patients with cancer of the vocal folds. Part 1: measures of fundamental frequency, *Journal of the Acoustical Society of America*, **49**, 1275–1282.

Hiller, S., Laver, J. & Mackenzie, J. (1983) Automatic analysis of waveform perturbations in connected speech, *Edinburgh University Department of Linguistics, Work in Progress*, **16**, 40–68.

Hiller, S., Laver, J. & Mackenzie, J. (1984) Durational aspects of long-term measurements of fundamental frequency perturbations in connected speech, *Edinburgh University Department of Linguistics, Work in Progress*, **17**, 59–76.

Klecka, W. (1980) *Discriminant analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences 07-001). Beverly Hills, London: Sage.

Laver, J., Hiller, S. & Hanson, R. (1982) Comparative performance of pitch detection algorithms on dysphonic voices. In *Proceedings of the IEEE International Conference ASSP 1982*, pp. 192–195.

Laver, J., Hiller, S. & Mackenzie, J. (1984) Acoustic analysis of vocal fold pathology, *Proceedings of the Institute of Acoustics*, **6**(4), 425–430.

Mackenzie, J., Laver, J. & Hiller, S. (1983) Structural pathologies of the vocal folds and phonation, *Edinburgh University Department of Linguistics, Work in Progress*, **16**, 80–116.

Mackenzie, J., Laver, J. & Hiller, S. (1984) Acoustic screening for vocal pathology: preliminary results, *Edinburgh University Department of Linguistics, Work in Progress*, **17**, 98–110.

SPSSSx User's Guide (1983) New York: McGraw Hill.

APPENDIX SIX

VOICE QUALITY AS AN EXPRESSIVE SYSTEM IN MOTHER-TO-INFANT COMMUNICATION: A CASE STUDY

H. Marwick, J. Mackenzie, J. Laver and C. Trevarthen

ABSTRACT: A study was carried out into a mother's use of her voice as an expressive and communicative instrument in play with her 18-week-old infant. A phonetic description of voice quality was used. It was possible to specify in detail the mother's vocal modulations using this system, and it was found that changes in the mother's voice quality closely reflected changes in her communicative intentions, as measured by other indices.

A collaborative study was carried out into a mother's use of voice as an expressive and communicative instrument in her interaction with her infant. We applied the system of voice quality analysis devised by Laver (1980) and category systems devised by Trevarthen and Marwick (1982)* for the description of mother-infant interpersonal behaviours.

Microanalysis from film or television of interactions which occur spontaneously between a mother and infant has shown that infants possess refined temporal regulation of expressive and explorative actions and an ability to interact in synchrony or contingent alternation with expressive moves of adult partners (Stern et al, 1975; Brazelton et al, 1975; Als, 1979; Trevarthen et al, 1981). Rapid development in the second month in the infant in facial, vocal and gestural signs of an integrated positive affect response causes, in turn, changes in the maternal behaviours (Sylvester-Bradley and Trevarthen, 1978; Trevarthen 1979a, 1983a). Mothers adjust the quality, pace and temporal pattern of their behaviour, presumably to obtain a strong response from the infant who is thus aided in the expression of inherent capacities for interaction (Stern et al, 1977; Kaye, 1977: Brazelton et al, 1974). Throughout the first year, the communication between infant and mother becomes increasingly complex and subtle, a process that culminates in the infant using vocalizations with a variety of communicative intentions (Halliday, 1975). It can be observed that both mother and infant use an intricate network of gaze, facial expression, a variety of vocalizations, laughter, touch, hand movements and body posture to engage each other's emotions and sense of humour. Both mother and infant show an intense interest in each other's utterances. The changes in interpersonal content and affective tone of maternal speech with age of the infant over the first few months after birth suggests that mothers are picking up information from responses and emotional expressions that guide them to produce forms of vocal output that are optimal for the infant's powers of perception.

Intensive observation of mother-infant behaviour enabled Trevarthen and Marwick to devise two category systems that describe mother-infant behaviours at two levels to constitute a comprehensive assessment of face-to-face interactions, revealing their form and the processes by which they are regulated (Trevarthen and Marwick 1982). These two systems are a 'macroanalysis' of psychological

---

*Note: Trevarthen, C. and Marwick, H. (1982) 'Co-operative Understanding in Infancy' Project report to the Spencer Foundation, Chicago, Department of Psychology, Edinburgh University.

functions in interpersonal communication, describing the affective and intentional level of psychological interaction, and a 'micro-analysis' of representative movements to show their precise anatomical distribution and temporal organization, providing detailed evidence on the mechanisms of development. In the macro-analytic system, distinctions of the mother's and infant's motivational state, interpersonal awareness, communicative intention and co-operative understanding of objects are made on a checklist comprising 211 categories. Figure 1 outlines the major functional areas into which the categories are grouped. Each category is defined in detail and is distinct from all others in its section. Any piece of interaction can potentially be described using categories from a number of sections simultaneously. Such breadth of description is an essential feature of the design of the system to reflect the complex motivational structure and the many simultaneous functions of interactions. This description is highly sensitive to the following: moment by moment changes in the emotional and cognitive goals in both partners, whether spontaneously produced or consequent on experimentally imposed variations in the mother's communicative aims; developmental trends and rates of developmental change in the infant's behaviour and the accompanying adjustments in the mother's behaviour; individual differences in the composition of behaviours in different mother-infant pairs. Interobserver agreement on the macroanalytic categories is 96 per cent for the mother and 94 per cent for the infant, over all sections (Trevarthen and Marwick, 1982).

Figure 1: Category System for Macroanalysis of Interactive Behaviours

This category system describes the interpersonal interaction, mutual awareness and the cooperative understanding of a mother-infant dyad. The derivation of these categories has followed theoretical principles which are reflected in groupings of the terms that cover major functional areas in interpersonal behaviours as follows:

A.	Self-regulatory:	Personal State/Mood Self-directed
B.	Reaction to Environment:	Exploratory Performatory
C.	Interpersonal:	Affect Engagement Play and Tease Modelling and Imitation
D.	Communicative Expression:	Messages Gestures and Utterances Conversation Structure
E.	Tasks:	Cooperative Use of Objects

The microanalytic system, accurate to one TV frame or 0.02 second, enables identification of specific channels of communicative signalling involving different perceptual modalities (principally visual and auditory) and different expressive means (upper and lower facial expression, gestures of arms, hands and fingers, posture and

head orientation and direction of gaze). Figure 2 outlines the microanalytic categories. This analysis detects the inherent rhythmical structure of activity, rate of response to stimuli, co-ordination between expressions in each subject, interactions between the two subjects and the development of joint control over behaviour. Inter-observer agreement on the microanalytic categories is 95 per cent over both mother and infant scores.

Figure 2: Category System for Microanalysis

Tapes are played in slow motion to observe movements of one part of the body at a time. Charts are built up showing the presence or absence of the following forms of behaviour which are located to 0.02 seconds (1 frame). Interobserver agreement for these categories is 95 per cent.

Prescriptive Group	Number of Categories
Gaze	6
Eyes	7
Brows	4
Nose	3
Mouth : Smile	4
Mouth : Open/Closed	6
Mouth : Grimaces	11
Tongue Protrusion	3
Jaw	4
Arms	10
Hands/Fingers	10
Palms	4
Head	7
Body Orientation	5

Although the content of mother's speech and its intonation contour and prosodic features have been analysed by Trevarthen and Marwick, the quality of voice of mothers interacting with infants has not been studied. Voice quality in adult interactions is generally thought of as 'the permanent background vocal invariable for an individual's speech' (Crystal, 1969; 103) or the socially important 'habitual voice of a person', but it is also recognised that, in spite of the relatively fixed characteristics of the voice that serve to identify the individual speaker, we make additional adjustments of the vocal organs which superimpose upon a given utterance a particular attitudinal colouring (Laver and Trudgill 1979). This is what is generally referred to as 'tone of voice' - a powerful expressive instrument. More technically, attitudinal features of 'tone of voice' are often referred to as 'paralinguistic' features.

In the past, voice quality has been loosely described in impressionistic terms, such as, 'soft', 'rough', 'quiet', 'firm', but its subtle variations caused difficulties and prevented systematic study. Not only were the impressionistic descriptive terms not sensitive enough to convey important differences of voice but it was also the case that different observers applied the same terms to different voices.

In an attempt to overcome the lack of a standard method for describing voice quality, Laver (1974, 1980) proposed a phonetic system for description of the normal voice. This system was the starting point for the development of the Vocal Profile Analysis

Vocal Profile

Speaker: ..... **Sex:** ..... **Age:** .....

I VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
A. Supralaryngeal Features										
1. Labial				Lip Rounding/Protrusion						
				Lip Spreading						
				Labiodentalization						
				Extensive Range						
				Minimised Range						
2. Mandibular				Close Jaw						
				Open Jaw						
				Protruded Jaw						
				Extensive Range						
				Minimised Range						
3. Lingual Tip/Blade				Advanced						
				Retracted						
4. Lingual Body				Fronted Body						
				Backed Body						
				Raised Body						
				Lowered Body						
				Extensive Range						
				Minimised Range						
5. Velopharyngeal				Nasal						
				Audible Nasal Escape						
				Denasal						
6. Pharyngeal				Pharyngeal Constriction						
7. Supralaryngeal Tension				Tense						
				Lax						
B. Laryngeal Features										
8. Laryngeal Tension				Tense						
				Lax						
9. Larynx Position				Raised						
				Lowered						
10. Phonation Type				Harshness						
				Whisper(y)						
				Breathiness						
				Creak(y)						
				Falsetto						
				Modal Voice						

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

Analysis Protocol

Date of Analysis: ..... **Tape:** ..... **Judge:** .....

II PROSODIC FEATURES

CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral		SETTING	Scalar Degrees					
		Normal	Abnormal		Normal			Abnormal		
					1	2	3	4	5	6
1. Pitch				High Mean						
				Low Mean						
				Wide Range						
				Narrow Range						
				High Variability						
				Low Variability						
2. Consistency				Tremor						
3. Loudness				High Mean						
				Low Mean						
				Wide Range						
				Narrow Range						
				High Variability						
				Low Variability						

III TEMPORAL ORGANIZATION FEATURES

CATEGORY	FIRST PASS		SECOND PASS			
	Adequate	Inadequate	Scalar Degrees			
			Inadequate			
			1	2	3	
1. Continuity						Interrupted
2. Rate						Fast
						Slow

IV COMMENTS

	Adequate	Inadequate	1	2	3
Breath Support					
Rhythmicality					

	Present	Absent
Diplophonia		

Other Comments:

Figure 3.

© 1981.

Scheme (VPAS), which has been described elsewhere (Laver et al. 1981). This is a perceptual analysis scheme, which uses phonetic concepts and techniques to describe and to quantify aspects of voice quality. Voice quality is described in terms of long-term articulatory adjustments of the larynx and of the supralaryngeal vocal tract. These habitual adjustments, or settings, combine with the anatomically-based fixed characteristics to make up the overall impression of a speaker's voice. A setting is any tendency, underlying segmental phonetic performance, towards the maintenance of a particular configuration of the vocal apparatus. One example of such a setting would be a tendency to keep the lips in a rounded posture throughout speech. Another would be the tendency to make one's phonation sound 'whispery'.

Central to this scheme is the concept of a neutral setting. This is a 'baseline' setting, against which any individual's voice quality can be judged. The neutral setting has clearly defined articulatory and acoustic correlates, at the laryngeal and at the supralaryngeal level. Individual voice quality may be judged to deviate from neutral in any of 10 broad categories. For each category, a speaker may be judged to have a neutral setting, or a non-neutral setting. If the setting is non-neutral, a further judgement is made to determine in what way the voice is non-neutral. Within each category, there may be several possible non-neutral settings, each of which has six scalar degrees to indicate the extent of the deviation from neutral. The overall combination of settings which characterizes a speaker is known as a 'vocal profile'. In a normal vocal profile, very few settings will exceed scalar degree 3. The vocal quality settings are listed on the VPAS protocol form shown in Figure 3. A complete vocal profile would also include judgements of prosodic features and temporal organization. Inter-judge agreement to within one scalar degree per setting has been found to be 94 per cent for the judgement of vocal quality features in non-pathological speakers.

Although this system was designed primarily for the description of habitual voice quality, we hoped that its detail would make it sufficiently sensitive to be used to evaluate the shorter term fluctuations in voice quality which are used as paralinguistic features.

We set out to establish whether Laver's phonetic system for describing voice quality could be applied to the speech of a mother to a young infant and, if so, to investigate the nature of the voice quality and voice quality changes and consider their communicative function and regulatory potential.

Our intention was to describe all the voice quality settings which were used by the mother and thereby chart the flow of change of these settings in her ongoing vocalisations. Having considered the nature and distribution of the voice quality settings we wished then to relate the voice quality of the mother to her other communicative and affective behaviour as measured by our macroanalytic category system of interpersonal behaviour, considering in particular any interdependence of change in the two systems, and any systematic relation between one set of voice quality features and one particular communicative or affective behaviour of the mother. We further wished to compare voice quality with the other expressive systems of facial expression, gaze and gestural behaviour in relation to interactive behaviour.

MATERIAL AND PROCEDURE

We chose for our analysis a section of video and audio tape from an interaction between a mother and her 18-week-old daughter taped in Trevarthen's laboratory. The mother had been asked to chat with the baby and to make her smile.

The mother and infant were video-taped in an observational set-up now standard in Trevarthen's laboratory using one camera and a front surface mirror. A special infant seat is used which holds the baby at 18" in front of the mother with maximum freedom for movement of head and limbs. Mother and infant are along together in a small carpeted room with sound absorbing curtains and studio lighting (Figure 4). Interactions are recorded from an adjacent room through a window. Separate additional studio-quality audio recordings are made throughout the session.

Because of the time involved in carrying out the analysis we took only a 40-second section of tape which we had informally observed to contain a number of variations in voice quality settings. Laver and Mackenzie analysed the voice quality settings from the audio tape, taking the syllable as the unit of analysis and using the scalar coding of the Vocal Profile Analysis scheme. Marwick and Trevarthen analysed the piece of interaction from the video recording using the macroanalysis system, simplified slightly for ease of comparability, and selected aspects of the microanalysis system, namely mouth expression (reduced, to avoid coding the frequently changing lip forms resulting from speech, to two categories of smile and two non-smiling categories), gaze direction and certain action categories appropriate to the section of tape under study.

The analyses were referred to the same time base and combined in graphical form (Figure 5). We shall call voice quality features which were present together a 'vocal set'. Any change in a setting or number of settings results therefore in a change of vocal set (except in certain cases outlined below). Similarly, we shall call macroanalytic categories of interactive behaviour which occurred simultaneously a 'state'. Thus change of one or more macroanalytic categories results in a change in interpersonal 'state'.

RESULTS

In our 40 second sample we found that 16 vocal sets were used, none of which was absolutely identical to any other. We found in addition five very brief variations in voice quality which lasted only for one syllable and occurred within the context of an otherwise stable vocal set. Because they occurred within a stable vocal set we felt that including them as independent vocal sets would give an unrepresentative idea of the number of changes of vocal set that the mother used. We therefore call them 'within-set' variations. All the voice quality changes occurred at the boundaries of discrete pause-defined utterances. Changes between vocal sets could occur in one of two ways. Either one or more settings in the set changed in intensity from its value in the preceding set, or the list of settings making up the set changed from the list in the previous set.

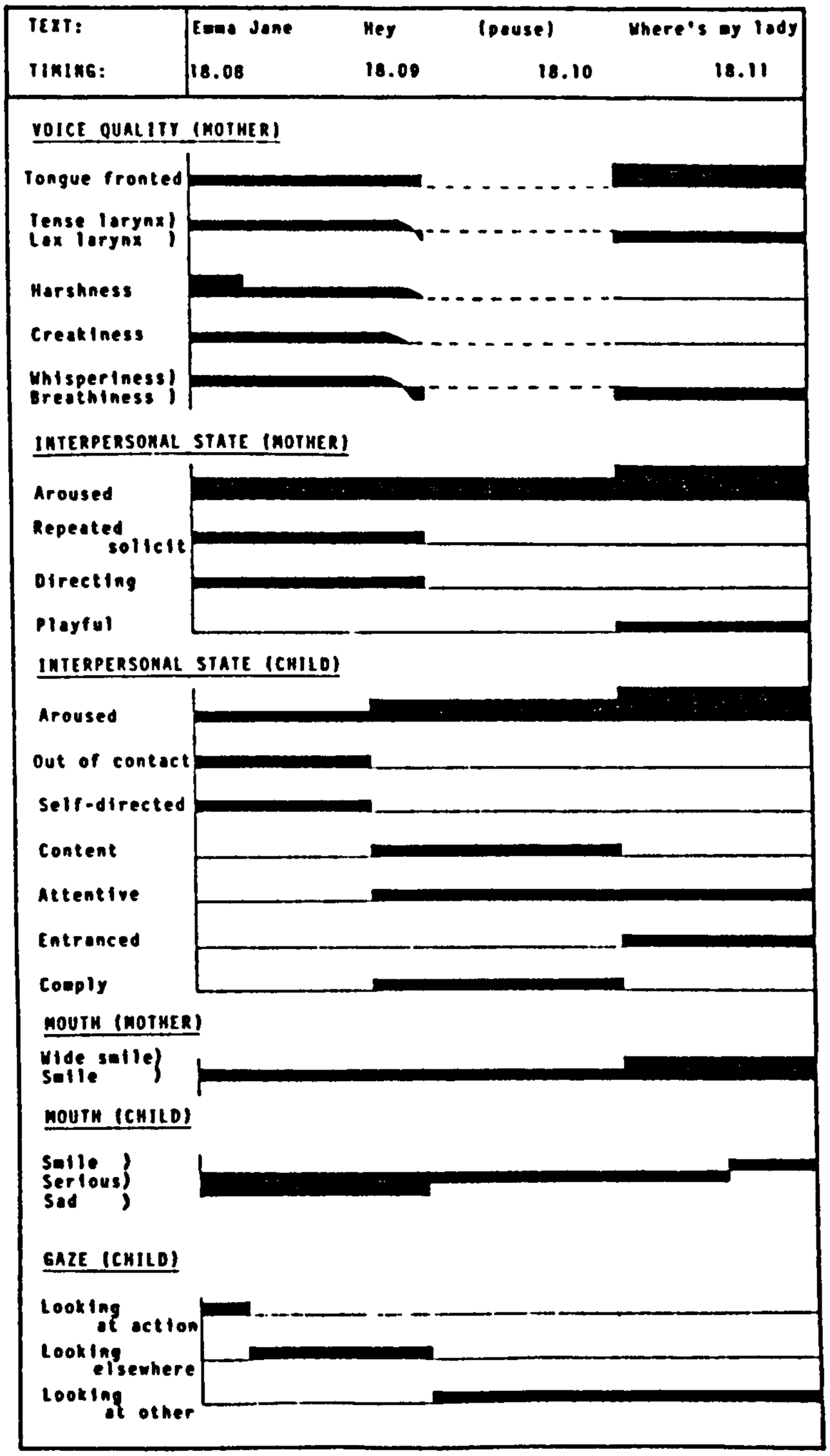
In the macroanalysis of interpersonal arousal, affect and intention, there were 27 changes of category sets noted for both the mother and the infant comprising 28 combinations or 'states' for each of them. Twenty-one of the mother's changes were accompanied

Figure 4: 'A mother talking and playing with her 6-week-old infant in the laboratory set-up'



BEST COPY
AVAILABLE
Variable print
quality

Figure 5: A representational portion of the graphed data. This has been simplified by the exclusion of all features which were absent or which showed no change during the course of this 4-second sample.



by utterances. When we related change in interpersonal state with change in voice quality we found that all vocal set changes accompanied an interpersonal state change (Table 1). Of the remaining interpersonal state changes, three were accompanied by a slight variation in voice quality, and two were not accompanied by any voice quality change. Two slight variations in voice quality were not accompanied by any interpersonal state changes. In both cases where an interpersonal state was not accompanied by any voice quality change the mother was adding further playful actions only to an already established playful engagement. There were no other occasions when this was all that changed. The two slight variations in voice quality that were not accompanied by changes in interpersonal state, were momentary introductions of creak and harshness on syllables with falling intonations (these are susceptible to creak and harshness if those occur even intermittently in the speaker's habitual voice quality profile, which is the case with this mother). We were struck by the direct reflection in change of voice quality of the mother's change of communicative intentions and affect, and concluded that not only was it possible to specify the details of the mother's voice using a descriptive technique based on perceptual phonetic principles, but also that this method of study had highlighted the importance of voice as a communicative and regulatory instrument.

TABLE 1: Relations between changes in interpersonal "state" (which are accompanied by an utterance) and changes in voice quality.

	Voice Quality		
	Vocal Set change	Momentary Within-set change	No Vocal Set change
Change	16	3	2
No change		2	

The next part of the study investigated the nature of the vocal sets and changes accompanying the mother's various intentional, affective and communicative states. Examples of our results are shown in Figure 6.

There is not a simple relation between voice quality and interpersonal intentions, affect and engagements. This is not unexpected, however, as the interpersonal states themselves are not simple and cannot be easily, or indeed usefully, extracted from the advancing interaction which gives them meaning. Having said that, however, we were impressed by the observation that for this mother, at least, the settings accompanying joint affectionate play (example 2) were very different from those accompanying soliciting behaviour (example 1). The settings in the latter case included tense larynx, raised larynx and whisperiness often with harshness and intermittent creak, where, in the first case, the settings were lax larynx, lowered larynx, breathiness and greater nasality and tongue fronting, with no harshness or creak.

Figure 6:

Mother's Interactive "State"	Vocal Set
1. Aroused 2 Happy Affectionate Attentive Repeated Solicit	Larynx tense 1 Larynx raised 1 Harshness 1 Whispery 1 Creak 1 Nasal 1 Tongue fronted 1 Tongue raised 1
2. Aroused 3 Happy Affectionate Attentive Playful (chant, dance) Enjoy	Larynx lax 2 Breathy 1 Larynx lowered 1 Nasal 2 Tongue fronted 2 Tongue raised 2
3. Aroused 3 Attentive Playful (vigorous) Boisterous Repeated Solicit	Larynx tense 1 Intermittent Creak Whisper 1 Intermittent Harshness Larynx raised 1 Nasal 1 Tongue fronting 3 Tongue raising 3

A further observation with regard to the nature of vocal set was that where the mother's interpersonal 'state' contained a mixture of soliciting and playful categories (example 3), as she tried to gain the infant's attention and interest through play, the accompanying vocal set contained an interesting mixture of settings, some of which were associated with soliciting behaviour, others being more intense levels of settings associated with joint play behaviour.

We then compared changes in voice quality with changes in other expressive systems as indicators of interactive state (Tables 2 and 3). Voice quality appears to be a more sensitive indicator of change of state than the other expressive systems we analysed. However, as explained above, the sets of categories for these other systems were somewhat simplified and the comparisons made should, therefore, not be overinterpreted. The comparisons do, nevertheless, demonstrate the applicability of this kind of voice analysis in investigations into the mechanisms of communicative interaction between mothers and babies.

TABLE 2:

Expressive Mode	Percentage of changes of interpersonal "states" accompanied by change in expressive mode
Voice Quality	90 per cent*
Mouth Expression	57 per cent
Gaze Direction	11 per cent
Action and Gesture	54 per cent

* State changes accompanied by utterances

TABLE 3:

Expressive Mode	Percentage of changes of expressive mode accompanied by change in interpersonal "state"
Voice Quality	90 per cent
Mouth Expression	84 per cent
Gaze Direction	30 per cent
Action and Gesture	47 per cent

CONCLUSION

This study has shown that a phonetic system of voice quality description can be used to describe in detail changes in a mother's tone of voice used in speaking to her young infant. It would appear to provide a sensitive and precise means of studying the way in which the mother uses her voice to communicate her changing intentions in interaction with her infant.

REFERENCES

Brazelton, T. B., Koslowski, B. and Main, M. (1974). 'The Origins of Reciprocity : The Early Mother-Infant Interaction', in The Effect of the Infant on its Caregiver, pp.49-76 (M. Lewis and L. A. Rosenblum, Eds.). New York and London: John Wiley and Sons.

Crystal, D. (1969). Prosodic Systems and Intonation in English, Cambridge University Press.

Halliday, M. A. K. (1975). Learning How to Mean: Explorations in the Development of Language, London: Edward Arnold.

Kaye, K. (1977). 'Toward the Origin of Dialogue', in Studies in Mother-Infant Interaction, pp.89-117 (H. R. Schaffer, Ed.). New York and London: Academic Press.

Laver, J. (1974). (1974). 'Labels for voices', Journal of the International Phonetic Association, 4: 62-75.

---------- (1980). The Phonetic Description of Voice Quality, (Cambridge Studies in Linguistics, 31), Cambridge: Cambridge University Press.

Laver, J., Wirz, S., Mackenzie, J. and Hiller S. (1981). 'A perceptual protocol for the analysis of vocal profiles', Edinburgh University Department of Linguistics Work in Progress 14: 139-155.

Laver, J. and Trudgill, P. (1979). 'Phonetic and Linguistic Markers in Speech', pp.1-30 in Scherer, K. R. and Giles, H. (Eds.) Social Markers in Speech, Cambridge:Cambridge University Press.

Stern, D., Jaffe, J., Beebe, B., and Bennett, S. (1975). 'Vocalization in unison and alternation: Two modes of communication within the mother-infant dyad'. Ann. New York Acad. Sci., 263, 89-100.

Stern, D. N., Beebe, B., Jaffe, J., and Bennett, S. L. (1977). 'The Infant's Stimulus World during Social Interaction', in Studies in Mother-Infant Interaction, pp.177-202 (H. R. Schaffer, Ed.). New York and London: Academic Press.

Sylvester-Bradley, B. and Trevarthen, C. (1978). 'Baby talk as an Adaption to the Infant's Communication', in The Development of Communication, pp.75-92 (N. Waterson and C. Snow, Ed.). New York: John Wiley and Sons.

Trevarthen, C. (1977). 'Descriptive Analyses of Infant Communicative Behaviour' in Studies in Mother-Infant Interaction, pp.227-270 (H. R. Schaffer, Ed.). New York and London: Academic Press.

---------- (1979a). 'Communication and Cooperation in Early Infancy. A Prescription of Primary Intersubjectivity', in Before Speech: The Beginning of Human Communication. (M. Bullowa, Ed.). Cambridge Cambridge University Press.

---------- (1983a). 'Interpersonal abilities of infants as generators for transmission of language and culture', in The Behaviour of Human Infants (A. Oliverio and M. Zapella, Eds.). London, New York: Plenum.

Trevarthen, C., Murray, L., and Hubley, P. (1981). 'Psychology of Infants' in Scientific Foundations of Clinical Paediatrics (2nd edn., pp.211-274) (J. Davis and J. Dobbing, Eds.). London: Heinemann.

APPENDIX SEVEN

The College of Speech Therapists Bulletin August 1986 No. 412

The Use of Two Voice Analysis Techniques in Clinic

Officers
President *The Earl of Halsbury, FRS*
Vice Presidents *Mrs Audrey Callaghan*
Sir Sigmund Sternberg, KCSG, JP
Chairman *Mrs D. Cox, MSc LCST*
Deputy Chairman *Miss M. Edwards, MPhil, FCST*
Hon. Treasurer *Mrs A. Jennings, LCST*
Press Officer *Miss Jois Stansfield, MSc LCST:*
118 Sheephousehill, Fauldhouse, West Lothian EH47 9EL
(0501-70484)

General Editorial Team
Catherine Humphreys, Annie Elias
Sally Jones, Jenny Yeatman:
Department of Speech Therapy, The London Hospital, Whitechapel London E1 1BB
(01-377 7177)

Reviews Editor *Mrs G. Taylor, LCST:*
18 Lady Frances Drive, Caistor Road, Market Rasen, Lincs.

Parliamentary Liaison Officer *Mrs Kay Wood, MSc, LCST:*
Ground Floor Flat, 96 Warwick Gardens, London W14.
(01-603 1700)

General Secretary *Mrs H. P. Fishman*

© *College of Speech Therapists*
Harold Poster House
6 Lechmere Rd, London NW2 5BU
(01-459 8521)

In this issue

Voice Analysis 1
Notices 4
Computers 7
Stuttering—Early Intervention 9

Introduction

At the Dysphonia Clinic, Royal Infirmary of Edinburgh (R.I.E.), two types of objective voice quality assessments have been used. An informal attempt was made to evaluate the contributions of these in planning and monitoring of therapy. This subsequent anecdotal report highlights the need for further study.

The first type of assessment, the Voice Profile Analysis Scheme (VPAS), is an auditory perceptual technique, whilst the other involves computer-based acoustic measurement. A brief description of each assessment is given, followed by two case studies. These illustrate the way in which the complementary information offered by the two techniques can be used in the clinic.

Analysis techniques

A. The Vocal Profile Analysis Scheme (Laver et al. 1981) is a phonetic technique for describing the long-term components which contribute to the overall impression of a speaker's habitual voice. Each of these components is called a setting. Analysis of an individual's voice quality involves comparison with a defined baseline setting, known as the "neutral setting". "Neutral" cannot be equated with normality but simply acts as a convenient reference quality. Scalar degrees are used to quantify the deviation of any setting from neutral.

The result of this analysis is a detailed Vocal Profile (see Figure 1a) which comments on:

1. Supralaryngeal and laryngeal features of vocal quality.
2. Prosodic features (pitch and loudness).
3. Temporal organisation (continuity and rate).

Only vocal quality features are recorded in Figure 1a, due to limitations of space.

The existence of clear phonetic information for each setting makes the Scheme quite objective and although the VPAS cannot be used without intensive training, trained judges do show high levels of agreement. A major advantage of the VPAS is that it highlights the interaction between different parts of the vocal tract.

B. The Acoustic Analysis System, in contrast to the VPAS, focuses more narrowly on laryngeal activity. This system, which has recently been developed at the Centre for Speech Technology Research at the University of Edinburgh (Laver et al, 1984; Miller, 1985; Laver et al. 1985) looks in detail at the irregularities in pitch (fundamental frequency, Fø) and loudness (intensity) which result from inefficient vocal fold vibration. The set of computer programs handles 40 second samples of tape-recorded speech, so that the patient need not be exposed to a daunting array of complex machinery.

The acoustic system was designed primarily as a screening tool for discriminating between the output of healthy larynges and those with organic pathology. In addition, it may be possible to discriminate acoustically between different types of organic change (for example vocal nodules, carcinoma, contact ulcers) and between different patterns of laryngeal misuse.

As output the Scheme gives ten acoustic measures:

1. Intonational measures
   (i) Pitch (Fø) mean
   (ii) Pitch (Fø) range (standard deviation)
2. Measures of phonatory irregularity
   (i) Four measures of pitch irregularity
   (ii) Three measures of loudness (intensity) irregularity

From these an acoustic profile of each speaker can be drawn up, which relates each measurement to the results obtained from a control population. An example is shown in Figure 1b. The zero point on each scale, highlighted by the heavy horizontal line, corresponds to the mean value for the control population and each unit on the scale corresponds to one standard deviation.

Approximately 95% of the normal population would be expected to fall within two standard deviations of the control group mean ie: between the light horizontal line. Measurements which fall outside the two standard deviation limits may be a potential indication of abnormality.

Case 1: "Organic with functional overlay"

Patient 1 was a 43 year old part-time Nursery Teacher referred by her GP to the E.N.T. department at the R.I.E. with an eight month history of intermittent

Figure 1a

Vocal Profile Analysis Protocol

Speaker: Patient 1 Sex: ♀ Age: 43

I VOCAL QUALITY FEATURES

o = Initial Assessment
● = Assessment Post-Therapy

CATEGORY	FIRST PASS: Neutral	FIRST PASS: Non-neutral: Normal	FIRST PASS: Non-neutral: Abnormal	SECOND PASS: SETTING	SECOND PASS: Scalar Degrees: Normal 1 2 3	Abnormal 4 5 6
A. Supralaryngeal Features						
1. Labial		o ●		Lip Rounding/Protrusion		
				Lip Spreading	●	
				Labiodentalization		
	●	o		Extensive Range		
				Minimised Range	o	
2. Mandibular	●	o		Close Jaw	o	
				Open Jaw		
				Protruded Jaw		
	●	o		Extensive Range		
				Minimised Range	o	
3. Lingual Tip/Blade		o ●		Advanced		
				Retracted	●	
4. Lingual Body		o ●		Fronted Body	●	
				Backed Body		
		o ●		Raised Body	●	
				Lowered Body		
	●	o		Extensive Range		
				Minimised Range	o	
5. Velopharyngeal		o ●		Nasal	●	
				Audible Nasal Escape		
				Denasal		
6. Pharyngeal	●	o		Pharyngeal Constriction	o	
7. Supralaryngeal Tension	●	o		Tense	o	
				Lax		
B. Laryngeal Features						
8. Laryngeal Tension	●	o		Tense	o	
				Lax		
9. Larynx Position	●	o		Raised	o	
				Lowered		
10. Phonation Type				Harshness		
		●		Whisper(y)	●	
				Breathiness		
				Creak(y)		
		o		Falsetto		
		●		Modal Voice		●

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

Figure 2a

Vocal Profile Analysis Protocol

Speaker: Patient 2 Sex: ♀ Age: 37

I VOCAL QUALITY FEATURES

o = Initial Assessment
● = Assessment Post-Therapy

CATEGORY	FIRST PASS: Neutral	FIRST PASS: Non-neutral: Normal	FIRST PASS: Non-neutral: Abnormal	SECOND PASS: SETTING	SECOND PASS: Scalar Degrees: Normal 1 2 3	Abnormal 4 5 6
A. Supralaryngeal Features						
1. Labial	●	o		Lip Rounding/Protrusion	o	
				Lip Spreading		
				Labiodentalization		
	●	o		Extensive Range	o	
				Minimised Range		
2. Mandibular	●	o		Close Jaw	o	
				Open Jaw		
				Protruded Jaw		
	●	o		Extensive Range	o	
				Minimised Range		
3. Lingual Tip/Blade		o ●		Advanced		
				Retracted	●	
4. Lingual Body		o		Fronted Body	o	
				Backed Body		
		o ●		Raised Body	●	
				Lowered Body		
	●	o		Extensive Range	o	
				Minimised Range		
5. Velopharyngeal		o ●		Nasal	●	
				Audible Nasal Escape		
				Denasal		
6. Pharyngeal		●	o	Pharyngeal Constriction	●	
7. Supralaryngeal Tension		o ●		Tense	●	
				Lax		
B. Laryngeal Features						
8. Laryngeal Tension		●	o	Tense	●	
				Lax		
9. Larynx Position		o ●		Raised	●	
				Lowered		
10. Phonation Type			o	Harshness	o o o o	
		●	o	Whisper(y)	●	
				Breathiness		
			o	Creak(y)	o o o o	
				Falsetto		
				Modal Voice		●

"VOCAL PROFILES OF SPEECH DISORDERS" Research Project. (M.R.C. Grant No. G978/1192)
Phonetics Laboratory, Department of Linguistics, University of Edinburgh.

hoarseness. This was aggravated by periods of intense conversation and voice strain at work. Indirect laryngoscopy demonstrated small vocal nodules in the middle of both folds. These were considered to be too small for surgical intervention. The patient was referred for Speech Therapy in the hope that further growth of the nodules, with the prospect of future surgery, could be prevented.

On initial interview, a case history was taken and a Vocal Profile was completed. Highly relevant factors were uncovered:

1. Dysphonic attacks had started soon after the patient began part-time nursery teaching, which involved much vocal strain.
2. Her home environment was also conducive to vocal strain; she had three young, loud and active children.

Results of a baseline Vocal Profile Analysis are outlined in Figure 1a (only Section 1 (Vocal Quality) is shown).

Though scores lay within the "norm" for all categories, patient 1 presented with a tense voice and a marked degree of whisperiness; lip movement was minimised and the jaw lay in a close position; there was also some degree of pharyngeal construction. All these features contributed to a generally tense long-term vocal behaviour.

Results of baseline acoustic analysis are given in Figure 1b; the patient was using a pitch mean and range very close to the female average; all four measures of pitch irregularity were, however, outside the two standard deviation limits, and two of four measures of loudness irregularity were close to the limits.

These results indicate that the patient's phonation is unusually regular. At first sight, it seems strange that this is undesirable but there are indications that unusually regular phonation may be associated with hypertension; indeed this is a fairly typical profile for speakers with vocal nodules.

A treatment programme was commenced and involved ten half-hour weekly sessions, based on the following features:

1. Counselling on the nature of vocal abuse; the influence of environment on voice quality.
2. Reduction of laryngeal tension and pharyngeal constriction by direct work on lowering the jaw and extending lip movement during speech. The VPAS had already highlighted the inter-relation between lip and jaw settings and laryngeal/pharyngeal settings.
3. Reduction of whisperiness by work on breath control and resonance.

Patient 1 was re-assessed three months later on the VPAS. Results are shown in Figure 1a, where dark shading indicates the change since initial assessment. Though some scores have remained static, there has been a marked improvement towards the neutral position for:

1. Whisperiness (3→1)
2. Laryngeal tension and pharyngeal constriction (2→neutral).

Range of lip movement has increased. These features were compatible with the patient's subjective impressions of improvement in voice quality.

Results of post-therapeutic acoustic analysis are given in Figure 1b: pitch mean was slightly lower and only one irregularity measure still lay outside the two standard deviations limit. These changes in the acoustic profile may all reflect a reduction in tension. The acoustic profile certainly appears to allow instrumental verification of the changes in laryngeal settings shown on the perceptual Voice Profile.

Patient 1 was reviewed again three months later and continued to exhibit this improvement in voice quality. She was subsequently discharged from Speech Therapy; following joint discussion, she agreed to re-establish contact if necessary in the future.

Case 2: "Functional"

Patient 2 was a 37 year old typist referred by her GP to the E.N.T. Department, R.I.E., with episodes of aphonia during the past year. She had experienced complete recovery in between each attack. Indirect laryngoscopy demonstrated no abnormality, though the patient had experienced an aphonic episode three days previously and sounded extremely creaky and whispery at the time of E.N.T. examination, with intervals of intermittent aphonia.

Figure 1b

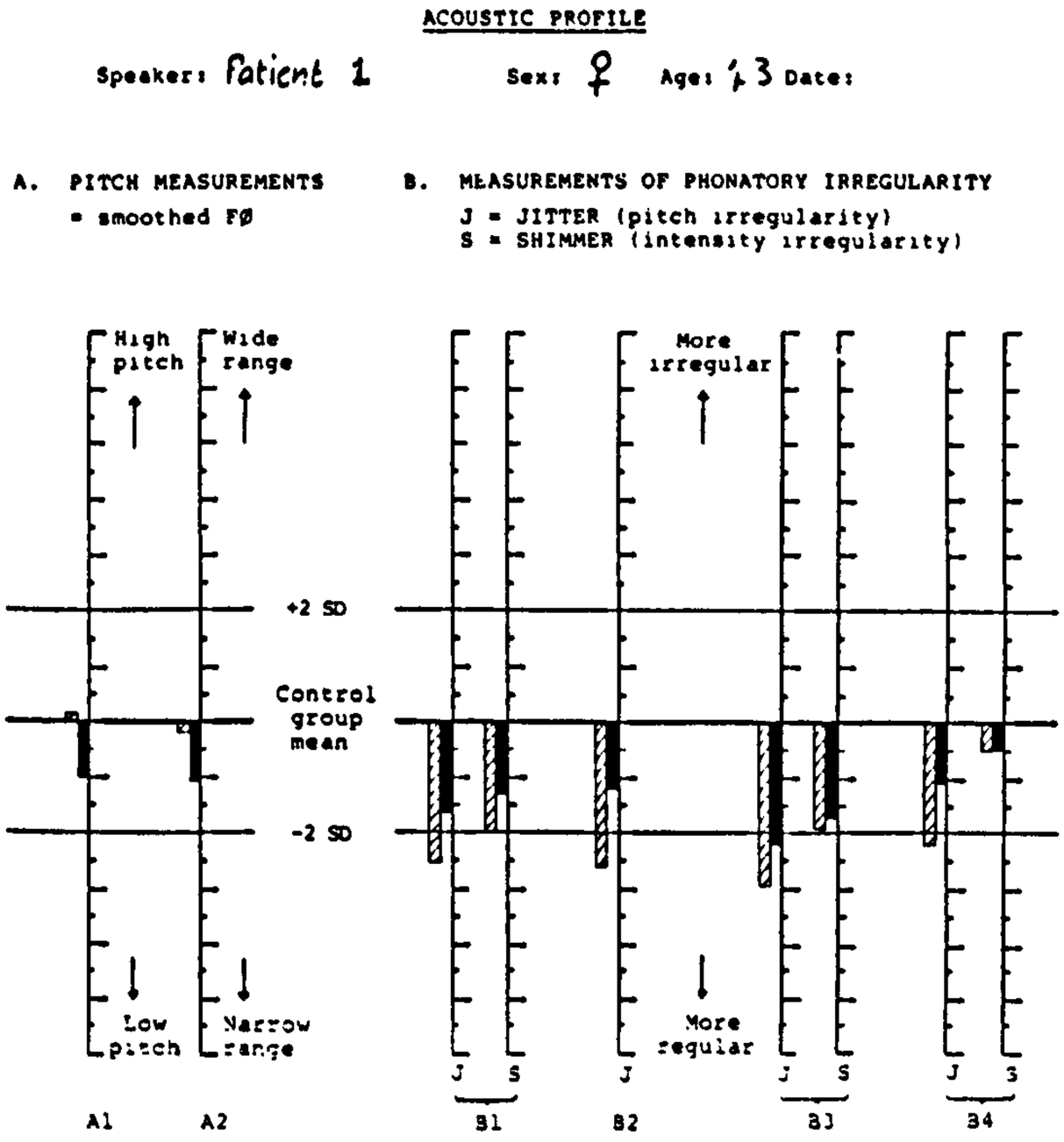
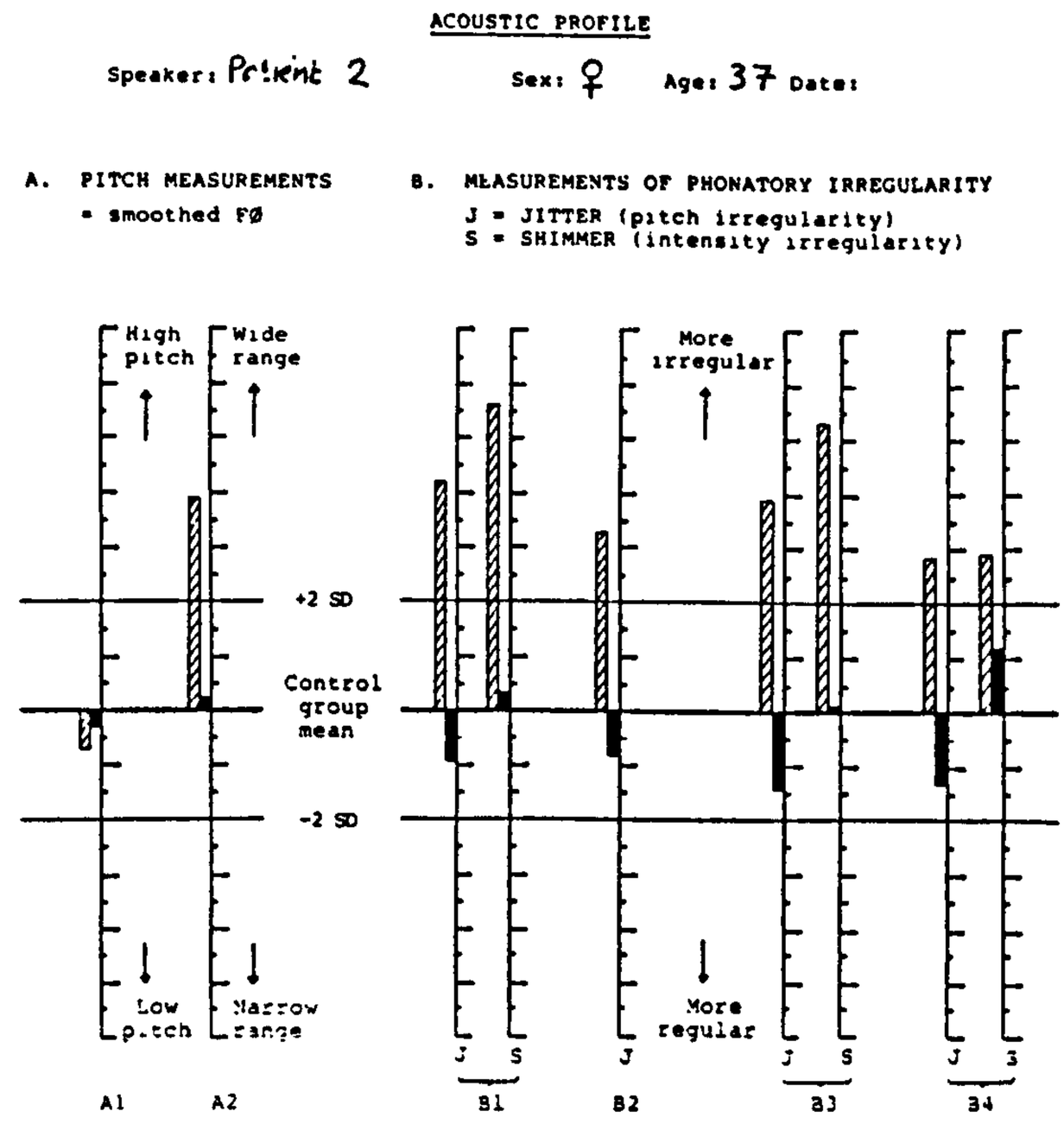


Figure 2b



"ACOUSTIC ANALYSIS OF VOICE FEATURES" Research Project.
(MRC Grant No. G8207136) Centre for Speech Technology Research,
Department of Linguistics, University of Edinburgh.

Following referral for Speech Therapy, she was assessed on the VPAS with findings as indicated in Figure 2a. These demonstrated:

1. An extremely tense superlaryngeal setting and a marked degree of pharyngeal constriction, both extending into the abnormal range.
2. A marked degree of laryngeal tension, an abnormal degree of whisperiness, with an intermittently high degree of harshness and creakiness.
3. Intermittent control of modal voice.

Results of baseline acoustic analysis are shown in Figure 2b: before therapy the acoustic profile looked grossly abnormal; although pitch mean lay within normal limits, pitch range was unusually wide and all measures of pitch and loudness irregularity were abnormally high.

Patient 2's voice disorder was considered a "functional" problem. This was confirmed during the course of subsequent therapy. Counselling highlighted much emotional stress, due to a recent divorce, and domestic worries. The patient was re-assessed after five weekly half-hour sessions, which included counselling and some voice therapy. On the VPAS, (dark shading again indicates the change since initial assessment (see Figure 2a)), she exhibited a marked improvement in voice quality: all parameters now lay within the normal range; whisperiness had reduced and was intermittent; labial, mandibular and lingual categories showed a movement towards the neutral position; moreover, the patient had now achieved modal voice.

Results of acoustic assessment, following therapy, (see Figure 2b), demonstrated that all measures now lay well within the normal range.

Conclusion

These two voice analysis techniques had several advantages with regard to assessment, treatment and management: both were non-invasive and non-threatening for the patient; they allowed an objective, concrete analysis of vocal behaviour and comparison with population norms, thus guiding the clinician towards treatment; they provided a baseline assessment against which to measure the efficacy of therapy. Such objective evaluation reflects a growing demand within Speech Therapy—thus these systematic, structured analyses offered more than a conventional informal subjective voice assessment (though they did not necessarily influence the course of subsequent treatment). The analyses also provided individual positive reinforcement to therapy as an abridged version of each was presented to the patient to give a visual record of changes in performance.

As *complementary* assessments, these techniques were wide-ranging, making use of both perceptual and instrumental data: where the VPAS highlighted the interaction between components of vocal quality (namely, the relation between lips, jaw, tongue, velopharyngeal, pharyngeal and laryngeal settings), the acoustic analysis concentrated on details of vocal fold vibration. In addition, the techniques involved close liaison between the clinician and the research team; this multidisciplinary co-operation tapped a wide range of professional expertise enhancing general patient management.

There are obvious refinements which must be mentioned: Ideally, the clinician involved in treatment should not be involved in assessment—in other words, "blind" re-assessment would have been more objective. In addition, review by indirect laryngoscopy at the E.N.T. Department would have been informative. Nevertheless, despite these facts and despite the training required before the VPAS can be used effectively, the above report provides a basis for future research; this account highlights the need for evaluation of different types of voice therapy; furthermore identification of patient populations (with various pathologies) might be valuable information in the treatment and management of voice disorders.

Acknowledgement
We would like to thank Mrs Marion Mackintosh, formerly Chief Speech Therapist, Royal Infirmary of Edinburgh, for her support.