INDIVIDUAL FEATURES IN VOICE QUALITY

John D. M. H. Laver

A thesis submitted for the degree of
Doctor of Philosophy
University of Edinburgh,
1975
ACKNOWLEDGMENTS

The research reported in this thesis has been under way for an uncomfortably long time. My pleasure in acknowledging the help I have received from many friends and colleagues is all the more warmly felt for the long duration over which I have had that support. The influence of David Abercrombie, my supervisor, and who first introduced me to phonetics, must be transparently obvious on many pages here, and I am happy to record my consciousness of my debt to him. I would also like to acknowledge my gratitude to a number of colleagues for particular help: to L.A. Iles, and my second supervisor until his retirement, Walter Lawrence, for their advice on acoustic matters; to E.T. Uldall, for generously allowing me to use part of her synthetic program for 'The North Wind and the Sun' in my attempts to synthesise different voice qualities; to J.A. Kemp, for help with translations of Latin and Greek materials; to David Cruickshank, who has persuaded PAT, over a number of years of the voice quality synthesis project, to do all sorts of things that I'm sure her designer, Walter Lawrence, scarcely anticipated might be asked of her, - and I owe thanks to David Cruickshank also for his help with the schematic anatomical illustrations.

I would particularly like to thank two colleagues in other universities for their helpful critical comments, Martin Kloster Jensen and Jack Windsor Lewis.

I have a special debt I must acknowledge, and I do so gratefully, to my colleagues John Lyons, Alan Davies, Gill and Keith Brown, for the continuing and finally successful pressure they exercised on me to finish this thesis.
Lillias Wylie typed the manuscript with exemplary speed and patience, and I thank her warmly.

I have left the recognition of my greatest debt to the end: my wife and colleague, Sandy Hutcheson, has given me the most generous support, professionally and domestically, and this thesis would certainly never have reached submission at this time or in this shape, without her unselfish tolerance and constructive help over a very long period.

John Laver
Department of Linguistics
January, 1975
Chapter 1: We know little more about voice quality now than was known in phonetic theory a century ago. But it can be argued that there is an intimate and crucial relationship between voice quality and phonetic quality, the prime datum of spoken language: phonetic theory, as part of general linguistic theory, should be able to offer a scientific analysis of voice quality, in order to clarify this relationship, as well as to be able to understand some of the factors that distinguish the voices of different speakers. An approach to such an analysis is initiated in this chapter by constructing a semiotic typology of conventional labels for voices, which leads on to a theoretical discussion of the principles, both impressionistic and scientific, on which different types and aspects of voice quality can be classified.

Chapter 2: The history of the analysis of voice quality is discussed, from the time of Cicero and Quintilian to the present day, with special emphasis on the period from the beginning of the seventeenth century onwards.

Chapter 3: A phonetic system for the description of voice quality is put forward, in terms of both physiological and acoustic parameters. A descriptive phonetic terminology is proposed. An experiment to establish the feasibility of synthesising different voice qualities by means of a resonance analogue speech synthesiser (MAT) is reported, and an illustrative tape recording of seventy two different synthetic voice qualities is provided.

Chapter 4: The place of voice quality in a semiotic analysis of spoken communication is discussed. Two particular areas are covered: the indication by voice quality factors of physical, psychological and social attributes of a speaker; and the particular semiotic relation between voice quality and phonetic quality, with some discussion of the
consequences of this relation not only for general linguistic theory, but also for the facilitation of improved techniques of analysis in areas of current interest in sociological linguistics concerning social differentiation by language, in urban dialectology.
CONTENTS

TITLE PAGE 1
ACKNOWLEDGMENTS ii
ABSTRACT OF THESIS iv
INTRODUCTION 1

PART I
CHAPTER 1. PRINCIPLES OF LABELLING VOICES 7

Introduction 7

A. DESCRIPTIVE LABELS FOR VOICES 13

1. Some phonetic correlates of impressionistic descriptive labels 17
   i) segmental correlates of impressionistic labels 16
   ii) voice dynamic correlates of impressionistic labels 18
   iii) voice quality correlates of impressionistic labels 19

2. Some semiotic aspects of impressionistic labels 22

3. The impressionistic basis of some phonetic labels 23

B. INDEXICAL LABELS FOR VOICES 25

1. The distinction between intrinsic and extrinsic features of the voice 26

2. Intrinsic indexical labels 27

3. Extrinsic indexical labels 28
   i) indexical labels of regional origin 29
   ii) indexical labels of social status 30
   iii) indexical labels of profession 30
   iv) indexical labels of interactional status 30
   v) indexical labels of type of interaction 31
   vi) indexical labels of perlocutionary effect 31
   vii) indexical labels of psychological characteristics 31

C. LABELS FOR VOICES IN SPEECH PATHOLOGY 33

1. Symptomatic labels 34

2. Etiological labels 36

Conclusion 37
CHAPTER 2. THE HISTORY OF VOICE QUALITY ANALYSIS

A. CLASSICAL WRITINGS ON THE VOICE

B. VOICE QUALITY AND PARALINGUISTIC DESCRIPTION

C. EARLY BRITISH WRITINGS ON THE VOICE

D. WRITINGS ON VOICE QUALITY IN THE 19th AND 20th CENTURIES

PART II

CHAPTER 3. THE PHONETIC DESCRIPTION OF VOICE QUALITY

Introduction

A. THE HISTORY OF THE CONCEPT OF AN ARTICULATORY SETTING

1. The concept of an articulatory setting in early writings on the voice

2. The concept of an articulatory setting in the phonetic literature of the 19th and 20th centuries

B. A PHONETIC DESCRIPTION OF VOICE QUALITY

1. Configurational settings of the supralaryngeal vocal tract

   (a) Longitudinal settings
       i) raised larynx voice
       ii) lowered larynx voice
       iii) labial protrusion

   (b) Latitudinal settings
       i) labial settings
       ii) lingual settings
       iii) faucal settings
       iv) pharyngeal settings
       v) mandibular settings

   (c) Velopharyngeal settings

2. Settings of the phonatory mechanism of the larynx

   (a) Phonation types
       i) nodal voice
       ii) falsetto
iii) whisper 228
iv) creak 229
v) harshness 233
vi) breathiness 242

(b) Compound phonation types 246

3. Settings of overall muscular tension throughout the vocal system 251

4. Structural and scalar conventions in the construction of descriptive phonetic labels for composite voice qualities 268
   (a) Structural conventions for descriptive phonetic labels 268
   (b) Scalar conventions for voice quality labels 274

C. THE ACOUSTIC SYNTHESIS OF VOICE QUALITY 276

PART III
CHAPTER 4. A SEMIOTIC VIEW OF SPOKEN COMMUNICATION 289

Introduction 289

A. SEMIOTIC TERMINOLOGY 290

B. VOICE QUALITY AND INDEXICAL INFORMATION 297

1. The indication of physical attributes by voice quality 298
2. The indication of psychological attributes by voice quality 310
3. The indication of social attributes by voice quality 313
4. Stereotyped judgements in voice quality 315

C. VOICE QUALITY AND PHONETIC QUALITY 316

1. Semiotic analysis of the phonic medium 323
2. Voice quality and accent 330

CONCLUSION 332

REFERENCES 335

APPENDICES 357
INTRODUCTION

One of the areas in the production of speech that has not yet been adequately described is that of voice quality, defined by Abercrombie (1967: 91) as being made up of 'those characteristics which are present more or less all the time that a person is talking; it is a quasi-permanent quality running through all the sounds that issue from his mouth'. Henry Sweet (1877: X) suggested that 'It is in (intonation), in the study of voice-timbre and of synthesis generally, that the work of future phoneticians must be concentrated'. We now know a good deal more about intonation, and about the synthesis of sounds, in Sweet's sense. But we know only a little more about voice-timbre, (voice quality in the terms of this thesis) than Sweet did nearly a century ago. Abercrombie was recently still able to write, with justification, that 'voice quality is the least investigated' of the various strands in the production of speech (1967: 91).

This may well be because the primary focus of phonetics is the capacity of speech to act as a vehicle for language. Any definition of the domain of phonetics which gives central importance to linguistic elements, as phonetics in Britain traditionally does (Jones & Laver 1973), will necessarily tend to give peripheral status to elements whose linguistic relevance is not immediately obvious. Indeed, it has sometimes been maintained that the analysis of voice quality is quite external to the study of linguistics (Sapir, 1921: 47). However, in a later chapter of this thesis, it is argued that there is an intimate relationship between voice quality and phonetic quality, on which their definitions in part mutually depend. Since phonetic quality is the prime datum for the study of the spoken mode of language,
the crucial part played by voice quality in the definition of phonetic quality is very relevant to any general phonetic theory, and hence to any well-founded general linguistic theory. Voice quality as such, if considered solely as a feature which characterizes the speaker, clearly lies outside language: but if it can be established that phonetic quality has a figure-to-ground relationship to voice quality, then voice-quality should not be dismissed as irrelevant to linguistics.

It thus seems a tenable position to maintain that this linguistic relevance is sufficient justification for the incorporation of a phonetic description of voice quality into general phonetic theory.

There is an additional justification, which arises from the contemporary widening of linguistic horizons. Many linguists today would accept that they have a double obligation: not only to describe and explain the form of language, but also to give some account of its functions in social interaction. A phonetician, as a social scientist like other linguists, shares this double obligation, and in the case of voice quality, it will be necessary to find descriptive and analytic terminology both for the phenomena of voice quality and for the communicative role of voice quality in social interactions.

Fortunately, incorporating a descriptive model of voice quality into general phonetic theory need not be a matter of adding a set of completely new basic principles of analysis. It is mostly a matter of extending well-established principles of phonetic analysis to a relatively new field of data. This will be discussed at greater length in the chapter on 'Principles of labelling', but briefly the argument is this: the principle of phonetic analysis of segmental aspects of speech production is that the composite physiological
events which together produce one 'segment' are analytically broken down into their component parts, and each independent physiological component is given a separate label. Like segmental articulations, voice quality is susceptible of description in terms of contributory physiological components which are analytically separable. We shall see that many of the concepts originally developed primarily for segmental description are also applicable, with normally only minor modification where necessary, to the description of voice quality.

The problem remains, however, of finding a suitable analytic language for discussing communicative aspects of voice quality in social interaction. We can find such an analytic language by taking a step back from phonetics, as it were, and looking at the relationship of phonetics and linguistics to other disciplines. Both phonetics and linguistics deal essentially with communicative signs, and the discipline which is superordinate to both phonetics and linguistics in that respect is semiotics, the general theory of communicative signs (Morris, 1938: 80). So phonetic theory, to the extent that it is concerned with the means by which communication is achieved through the use of spoken signs, can be said to be part of semiotics. The technical vocabulary of phonetics can thus be seen as a special part of the metalanguage of semiotics. One might therefore reasonably expect that other terms and concepts from semiotic theory could be illuminating when applied to the subject-matter of phonetics. This is indeed the case, and one concept in particular will be useful for this thesis, that of an 'index', or 'indexical sign'. This notion will be discussed at greater length in a later chapter, but at the moment it is sufficient to say that indexical information in speech is information about the characteristics of the speaker. There are
many such indexical features. For example, a speaker's accent can reveal his regional origin; similarly, a speaker's choice of particular words can indicate such factors as his profession, his momentary mood, his social status, and so forth. Most importantly, for this thesis, a speaker's voice quality acts as an index to a surprisingly wide range of information about him, to do with physical features such as sex, age, build, state of health, fatigue, psychological features such as personality and mood, and social features such as place of origin, social status, and sometimes occupation. Abercrombie (1967), who borrowed the term 'index' from C.S. Peirce, the American pragmaticist philosopher of the late nineteenth century, offers an extensive discussion of the notion of indices in speech. Sapir (1927) and Laver (1968) discuss the same area.

In the literature of phonetics, many writers have made occasional comments about voice quality, but almost always only rather cursorily. For a long time, other disciplines such as psychology, psychiatry and speech pathology have been more ready to acknowledge the relevance of studies of voice quality to their professions. Naturally, the study of such topics in speech legitimately attracts the research attention of a number of different disciplines, each with a different professional focus. For the majority of these disciplines, speech is a partial interest, and the main preoccupation of the discipline lies outside speech as such, as in the case of psychology and psychiatry. A discipline like speech pathology, on the other hand, takes speech as its principal data, but brings a specialized interest to bear on a restricted area within the wider field of speech as a whole. Phonetics, as the subject which claims to take as its professional domain the study of all aspects of the medium of spoken language (Abercrombie,
1967: 2), should have something to offer to all these other
disciplines, in the area of voice quality no less than in more
centrally linguistic areas. To psychology and psychiatry, for
example, it should be able to offer a set of descriptive labels
for voice quality, of established reference and universal currency.
To speech pathology, phonetics should be able to offer a valid
characterization, in general phonetic terms, of the range of possible
variations in 'normal' voice quality, against which the deviations
of 'abnormal' and pathological voices can be evaluated.

No such comprehensive description is available in the literature
on general phonetic theory; and the scale and detail of a complete
description of non-pathological voice quality are such as to make it
quite unrealistic to hope to be able to present it in its entirety
in this thesis. It should be possible, however, to draw the main
structural outline of the descriptive system, in terms of the
individual components in voice quality, with some physiological and
acoustic detail, and to offer a commentary on semiotic aspects of
voice quality as a factor in social interaction.

Part I of the thesis prepares the way for the exposition of
the descriptive phonetic system, through a discussion in the first
chapter of the semiotic principles underlying various types of labels
for voices; and an account in the second chapter of the history of
the analysis of voice quality from Cicero and Quintilian onwards.

Part II is the outline, in the third chapter, of the phonetic
description of voice quality. This account of the descriptive
system constitutes the major section of the thesis.
Part III, in the last chapter, considers a semiotic view of spoken communication. There are two principle areas covered: the indication by signs in voice quality of physical, psychological and social attributes of the speaker; and the relation between voice quality and phonetic quality, including a discussion of the possible contribution of voice quality features to accent.

Three appendices are included: two previously published articles on voice quality; and an illustrative tape recording of the acoustic synthesis of seventy two voices with different voice quality settings.
Part I

Chapter 1. Principles of labelling voices

Introduction

In many areas of perception, sets of conventional labels are available in everyday language for the categorization of particular varieties of quality. This applies most validly to the perception of colour: not only are there terms for the different hues, but also for different degrees of saturation and intensity. It is reasonably meaningful, for example, to the large majority of readers, to write of 'a dull, deep purple', or 'a light, bright turquoise'.

There are also conventional labels for different qualities of odour, though of slightly less specificity than colour terms. Readers understand fairly uniformly what is meant by labels such as 'an acrid smell', or 'a spicy smell'.

The situation is different, however, though it may not at first sight appear so, when we come to descriptive labels for different sorts of voices. As laymen, we have a very wide and subtle vocabulary for talking about the perception of voices. Our widespread interest in this area is reflected in the fact that it is quite common to come across comments on aspects of voices such as 'accent' and 'voice quality' in reading newspapers and magazines. Similarly, many novelists and writers show an intuitive grasp of the wealth of indexical information conveyed by the voice, and take advantage of our interest in this topic by using descriptive comments on the voice as an economical literary device for investing
their characters with individuality. However, the apparent precision of very many written descriptions of voices can be very misleading, if one looks at them from a technical phonetic point of view: this is partly because of the difficulty of deciding to which aspect of the voice reference is really being made - to accent versus voice quality, for example; and partly because of the difficulty, even when the aspect has been determined, of knowing what vocal phenomena are involved. The following quotations may show something of the variety and ambiguity of reference of such comments:

1. '..... in her fluent and flashy New York English, with brash, metallic nasalities and soft lapses into furry Russianisms ....' (Nabokov, V., 1957, Pnin, Doubleday, New York)


3. 'When she asked a question I noticed that she spoke with her teeth together but with her lips moving very freely. This gave her voice a harsh, resonant quality which I thought suited her looks .... her clenched-teeth voice ....'. (Anis, K., 1955, That Uncertain Feeling, Victor Gollancz, London)

4. '..... the Arabs shouted to each other in their harsh, far-carrying voices ....' (Thesiger, W., 1959, Arabian Sands, Longmans Green, London)

5. (The voice of a pimp) 'was carefully wrapped, but it had the suppressed clangour of a sewer-cover being lifted from its hole and dropped to the asphalt'. (Mailer, N., 1965, An American Dream, Deutsch London)
6. '.... the stretch of the Earls Court Road which taxi-drivers call Kangaroo Alley, near the station. There the beer is iced, man, and the voices flat and nasal ....' (Observer Weekend Review, May 9, 1965, on 'Kensington Colonials')

7. '.... a light voice is a grave handicap to a politician. First, a general example: the male American voice, deeper and more resonant and altogether heavier than the Englishman's voice, is far more effective in getting and holding attention and in convincing. Let an idea, however, foolish, be put forward by such a voice, and it at once acquires weight and seems worthy of consideration; but let exactly the same idea, even in the same words, be put forward in, say, the light voice of Sir Alec (sc. Sir Alec Douglas Home, J.L.), and the silliness of what the speaker is saying is at once apparent'. (Hyams, E., New Statesman, March 29, 1966, on 'Party Voices')

8. 'Beautiful bronzy voice .... The voice is steeped in the black pitch of soul .... There is striking beauty in this harled, bronzed sound set as it is against sparse electric piano, basic guitar chording and an occasional female harmony'. (Clark, A., The Scotsman, December 4, 1972)

9. '.... the Edinburgh Baroque Choir under Richard Walker, sensibly chose a fairly short programme which enabled their bright, young voices to shine without tiring'. (Wilson, C., The Scotsman, December 4, 1972)

10. 'Did you know your voice has a colour? It's true. It can be brown, pink, chocolate, coffee - any one of 101 different shades. I learned this from Harold Landey. He's the man who supplies voices for TV commercials.

He has 5000 voices on file.
The colour coding makes it easy to pick out the different types.

A dark brown voice is good for selling beer. But it would make bread sound stodgy.

A light brown American twang is grand for detergents.

An airy fawn voice is fine for furniture polish .... (Sunday Post, May 3, 1970)

11. 'It does not do, it seems, to cultivate a British accent in the American State Department. An official claims that the department "has a lousy image, and we earned it sending out stuffed shirts with phoney British accents to represent the Foreign Service".

Presumably the accent employed by these quasi-British representatives of the land of the free, not to say, easy speech, was Oxford English. It would be difficult to equate nasalised Brum and the richly-inlaid verbal sonorities of up-country Buchan with stuff-shirtedness. In any case, what once seemed to be regarded in some circles as the accent most representative of Britain, BBC English, is now being replaced by a sort of mid-Atlantic accent of indeterminate enunciation, now much in favour with television announcers and demonstrators of vacuum cleaners, egg whisks and jumping furry toy spiders in stalls connected with exhibitions of vital household necessities.

The time may come, alas, when we are all speaking in classless, de-localized accents, when it will be impossible to point out the musical precision of Inverness speech, the glottic hilarities and artistic elisions of the Glasgow accent, the illusions of refinement connected with the Edinburgh version of the Queen's English, not to mention the Border tongue, so suggestive
in some cases of nuggets of gold being mined from some deep pit after a painful and complicated excavation process'.
(Morris, A., The Scotsman, December 5, 1969, on 'Some blunt words on an acute accent')

12. 'Juliette Greco performs a vocal striptease. She is coldly erotic; she titillates the body and seduces the mind. She has a square face, a flat nose, lanky hair and a lopsided mouth - and yet is thought beautiful. At the Festival Hall last Monday, describing the plots of her songs, her husky, brown voice imparted an aura of significance which they rarely possessed'. (Palmer, T., Observer Weekend Review, March 15, 1970, on 'Poems in her throat')

Some of the labels in these quotations, such as 'nasal', 'husky' and 'harsh' may result in reasonably accurate communication, because their referential link with the qualities of sound concerned are long-established and familiar to the great majority of readers. This is much less likely to be true of labels like 'bronzy', 'furry', 'heavy', 'brown' and 'suppressed clangour'. They are all the more misleading for being apparently so evocative, - each reader may conjure up a detailed image of the voice referred to, but it may well be that no two readers would agree closely about the nature of the imagined voice.

To capitalize on the richness of the observations which underlie lay comments on speech, we have to be able, as phoneticians, to elucidate their characteristic vagueness and ambiguities.

We have had more success in doing this with some types of comment than others. For example, we can be phonetically reasonably precise about what speech phenomena are being referred to by lay usages such as 'a London accent', or 'a lisp'. We have had slightly less success so far in being able to say exactly what might be meant by such phrases as 'he has a very broad accent', or 'he has very flat
One particular type of phrase in everyday lay usage which has so far largely evaded efforts towards phonetic specificity is the formula 'a .... voice', where the adjective preceding 'voice' might be one of literally hundreds of labels on the pattern of 'dulcet, venomous, hollow, educated, foreign, British, metallic, angry, colourless, dark brown, sharp, brisk, soporific, pleasant, melodious, ingratiating, sepulchral' and so forth. Abercrombie (1967: 94) and Crystal (1969) have recently drawn attention to the relative neglect of this area of phonetics.

It would be a mammoth task to try to give an exhaustive specification of what the phonetic correlates of all such labels might be, and perhaps one which is in principle impossible, since quite often a particular label means different things to different users. But it should at least be possible to construct a semiotic typology of these labels, in which suggestions are offered as to what aspect of the semiotic process of speech the labels are making implicit or explicit reference. At first sight, it looks as though an everyday label on the pattern of 'a .... voice' should be a label for a particular voice quality. Sometimes it is, but it is often a comment about other aspects of speech production as well or instead, or sometimes about imputed aspects of the speaker's identity, or even, sometimes about the effect of the voice on the listener. The apparent unity of the everyday formula 'a .... voice' thus conceals a wide variety of different types of referent. One profitable way of approaching the more scientific description of the phenomena and functions of voice quality is to attempt to trace the thread of reference to voice quality through the tangled complex of everyday descriptive language about voices.
The purpose of this chapter, then, is to try (following Abercrombie, 1967: 90) to identify, in a broad semiotic framework, the type of referent involved in lay descriptions of voices, in the hope of facilitating an explicitly phonetic and semiotic analysis of voice quality.

Throughout this chapter, one major reservation should be remembered by the reader. Nearly all the statements here are based on my personal impressions about the meanings of the labels discussed. They should certainly not be regarded as statements of fact. It may be that most of the interpretations of the different labels offered here will find a measure of agreement with the reader's interpretations, and I hope this will be the case. But the comments made about the meanings of the various labels have no greater authority than that of any reader, and their speculative, tentative nature should be kept firmly in mind.

The first and most fundamental semiotic distinction between different types of labels for voices concerns whether the label refers to the sounds produced by the speaker or to the characteristics of the speaker producing the sounds. It will be convenient to call the first sort of label a descriptive label, and the second an indexical label.

A. Descriptive labels for voices

Descriptive labels themselves fall into two broad categories, impressionistic labels and phonetic labels. Impressionistic labels are labels which need an audible demonstration of the type of voice referred to before the listener or the reader can construct an accurate interpretation of the label. Pike (1943: 16) has
called such labels 'imitation labels'. Phonetic labels, on the other hand, should be part of established phonetic vocabulary, and as such should have exact and agreed definitions subscribed to by all trained phoneticians (in the ideal situation, at least). The system of phonetic descriptions of voice quality outlined in Chapter 3 below offers a set of phonetic labels, as a first attempt at a comprehensive taxonomy which could be generally meaningful to phoneticians. It is acknowledged that this descriptive system falls a long way short of 'the ideal situation'; it is offered here as a beginning. One would hope that phonetic theory should eventually be able to suggest translations of all common impressionistic labels for voices into plausible phonetic labels of the sort discussed below.

One characteristic of most impressionistic labels is that they tend to use single adjectives for individual voices (as in 'heavy voice'), and for these different single adjectives to have a very variable domain of reference. It will be helpful in trying to make the phonetic reference of impressionistic labels explicit to follow Abercrombie in dividing the medium of speech into 'three strands, separable though closely woven together, all simultaneously and continuously present and together making up the totality of the medium' (Abercrombie, 1967: 89). These three strands are: segmental features, features of voice dynamics and features of voice quality (Abercrombie ibid.). Some impressionistic adjectives refer implicitly to a single strand in the production of a particular voice, as in a 'sibilant voice', where one might conclude that reference is being made to features of segmental pronunciation. Others refer implicitly to two strands, as in a 'sepulchral voice', where
reference to features of both voice dynamics and voice quality seems likely. Still others seem to refer to all three strands, segmental pronunciation, voice dynamics and voice quality together, in labels such as a 'twittering voice'. Most impressionistic labels seem to be of the more holistic sort, where multiple reference is implicit in a single label, as in 'velvety, rich, golden, flat, thin, bird-like,' and 'thick'.

Phonetic labels used for voices share the same characteristics as the rest of the taxonomy for phonetic descriptions: holistic description is avoided, and each controllable independent component is individually identified, giving a composite label. A standard anatomy and physiology is assumed, with the consequent advantage that the labels can in effect act as instructions to the reader or listener about how to control his own vocal organs in such a way as to produce the same vocal effect as the phenomenon under description. Interpersonal differences of anatomy and physiology are disregarded as irrelevant in this most basic assumption of a universal phonetic theory - that the vocal apparatus, and all parameters of vocal control, are standard to all human beings. So phonetic labels for voices, like phonetic labels for segments, are composite multi-term labels conveying componential instructions to the reader about how he himself can achieve the phonetic effect being described. Examples of phonetic labels for voices would be 'a slow, loud, ventricular whispery voice', or 'a fast, soft, nasal, lax, breathy creaky voice'.

It will be immediately obvious that many terms in the composite phonetic labels would be as much in need of prior audible demonstration for intelligibility as the holistic impressionistic labels, were an
individual phonetician to meet a particular phonetic label for the first time in the form given above. In the last resort, all descriptive labels commonly used in phonetics have to have had a basis of audible demonstration during the training of the individual phonetician. There are two differences at this level, however, between the systems of impressionistic and phonetic labels. One is to do with the number of different labels available in each system, and the other is to do with the number of interpretive conventions about the meaning of the labels in each system. With regard to the size of the labelling inventory, the number of individual phonetic terms which can be used to make up the available list of composite phonetic labels is fairly limited, but capable in combination of producing a very large number of such composite labels. This means that the phonetician has only to master (through training by audible demonstration in part), a small number of articulatory-auditory correspondences in order to be able to recognize and produce this large number of composite phonetic events. The overall list of impressionistic labels is much larger than the list of individual phonetic labels, relying as it does to an important degree on the rich metaphorical resources of the language, but it is weaker than the phonetic system because of the second factor, to do with interpretive conventions. In the phonetic system, as mentioned above, the conventions for interpreting the reference of the comparatively small number of individual productive components are rigorous, explicit, and universal amongst trained phoneticians, ideally at least. Communication in writing about any phonetically composite voice quality, between individual phoneticians trained in the system, is relatively simple, and accurate, without any need for audible demonstrations of the
composite events to which reference is being made. In the impressionistic system, each different holistic label needs its own individual interpretive convention, and the users of the system have to rely therefore on the general conventions of everyday language for effective communication with their readers or listeners.

A number of inefficiencies in the impressionistic system arises from this. Firstly, communication is less than adequate in cases where, in writing about voices, the writer is mistaken in his belief that the referent of a label he uses is accurately accessible to the reader through a culturally-disseminated convention. In cases of this sort, where the writer and the reader have different interpretations of the same label, both writer and reader are likely to be unaware that accurate communication is not being achieved. The difficulty arises because the convention specifying the meaning of the label is tacit, and not explicit as in the case of the phonetic system. The same difficulty underlies a similar breakdown of accurate written communication about voices when different writers attach different labels to the same referent. In all these ambiguous situations, the reader has at best only approximate information about the writer's communicative intentions, and at worst can be seriously misled.

The existence of many impressionistic elements in supposedly scientific, precise phonetic description will be discussed more fully later in this chapter.

1. Some phonetic correlates of impressionistic descriptive labels

We go on now to explore in some detail the types of correspondences
between impressionistic labels for voices and the different phonetic strands in the production of speech, expanding on suggestions by Abercrombie (1967: 90). In this analysis, the reader may well find himself in disagreement about the phonetic correlates claimed to exist for a particular impressionistic label, precisely for the reasons of ambiguities of interpretation inherent in impressionistic labelling outlined above. However, the object here is less to secure exact phonetic agreement about the correlates than to show the variety of implied reference to different aspects of sound production that impressionistic labels may have.

1) **segmental correlates of impressionistic labels**

There seem to be very few impressionistic labels for voices which refer primarily to segmental pronunciation, or which include reference to segmental pronunciation together with other features. Adjectives for voices whose reference can plausibly be thought to have connotations of segmental pronunciation include not only 'sibilant', and 'twittering', as mentioned above, but also 'bleary, blurred, distinct, guttural, lisping, orotund, precise, slurred', and perhaps terms like 'clipped' and 'thick'.

ii) **voice dynamic correlates of impressionistic labels**

Impressionistic labels frequently have implications of reference to various voice dynamic features. These include features of pitch-range, pitch movements, loudness-range, tempo, continuity and phonational register.

Terms with implications for features of pitch-range include 'bird-like, bleating, colourless, creaky, creaky, dark brown, deep,
flat, fluting, grating, gravelly, graveyard, gravy, gruff, heavy, high, light, light blue, monotonous, pale, pale pink, piercing, piping, rich, reedy, rumbling, sepulchral, shrill, silvery, sombre, sonorous, twittering, whining'.

Labels with implications for particular types of pitch movements within a particular pitch-range include 'colourless, creaky, croaky, far-carrying, full, gravelly, gruff, husky, light blue, loud, low, orotund, pale, pale pink, piercing, querulous, raucous, resonant, rumbling, sharp, shrill, small, soft, sombre, sonorous, staccato, strident, strong, weak', and perhaps 'brazen' and 'rasping'.

Implications of tempo are contained in labels such as 'drowling, droning, fast, graveyard, quick, sepulchral, slow, sombre, sonorous, staccato, twittering', and perhaps 'sharp'.

Continuity is referred to in labels like 'clipped, droning, jerky, staccato', and perhaps 'smooth, fluent' and 'fluid'.

Phonational register, as the mode of vibration of the vocal cords which the individual speaker can control on a moment-to-moment basis for linguistic and paralinguistic signals, seems to be implicitly referred to in many impressionistic labels, such as 'breathy, cracked, crackling, creaky, croaky, falsetto, grating, gravelly, gritty, gruff, harsh, hoarse, husky, rasping, rough, rumbling, shrill, sombre, thin, throaty, velvety, whisper, whispery', and perhaps 'thick', and possibly 'blurred'.

iii) voice quality correlates of impressionistic labels

Reference to the voice quality strand in the production of
speech falls into two groups (discussed in detail in Chapter 3, on the phonetic system for describing voice quality): firstly, to the invariant anatomical and physiological foundation of a speaker's vocal equipment, and secondly, to the different types of muscular 'settings' the speaker habitually superimposes on this basic biological foundation (Laver, 1964, 1967, 1968).

There are only a very few impressionistic labels for the first group, the physical foundation. They include some of the conventional terms for different types of singing voices, -treble, tenor, baritone, bass, double-bass, soprano and contralto'. Not all terms for singing voices are of this sort, however, - as for example 'counter-tenor'.

Labels for the second group, the habitual long-term muscular settings, can be divided into a number of different sections, depending on the type of setting to which they refer. Impressionistic labels for 'longitudinal' settings, which modify the basic length of the vocal tract, are more numerous for the modifications achieved by vertical adjustment of the larynx than for those achieved by protrusion of the lips. One possible label for the latter may be 'a pursed-lips voice', but the type of lip activity being referred to in this label may sometimes refer to a constriction of the vocal tract by rounding of the lips rather than a lengthening of the tract by labial protrusion. Terms for the effect of vertical adjustment of the position of the larynx include 'clergyman's (voice), dark brown, graveyard, hollow, light blue, pale pink, pulpit, preacher's, sepulchral'.

'Latitudinal' settings, which are long-term tendencies to modification of the cross-sectional area of the vocal tract, seem to
be referred to by impressionistic labels for settings either of
the pharynx, or involving the lips or the jaw, but not for settings
of the tongue in the oral cavity itself. Labels for latitudinal
settings are 'clenched-teeth (voice), grinning, guttural, 'hot-
potato, orotund, plummy, slack-jawed, strangled'. It is interesting
that there seems to be no impressionistic label corresponding to the
phonetic label 'velarized voice', as a label for an aspect of speech
production, - though there are a number of labels for the
characteristics of the speaker that the listener thinks such a voice
signals.

Velic settings in voice quality are included in the reference
of labels such as 'droning, nasal, sonorous, whining', and perhaps
'plangent'.

Settings of overall muscular tension of the vocal tract
(including the larynx as part of the tract) are referred to by
labels like 'brassy, brazen, dull, hard, metallic, muffled, pig's
whistle, plangent, rasping, raucous, resonant, soft, strident, tinny'.

Laryngeal settings, as the mode of vibration of the vocal cords
which is quasi-permanent in the voice of the individual speaker, are
included in the meaning of quite a large number of impressionistic
labels, most of which can also be used to refer to the shorter-term
voice dynamic feature of phonational register, though not all. The
labels include 'beery, bird-like, brandy, breathy, cracked, crackling,
creaky, croaky, dark brown, dulcet, falsetto, fluting, fruity, ginny,
gravelly, gruff, harsh, hoarse, husky, light blue, mellifluous, mellow,
melodious, metallic, muffled, musical, pale pink, pig's whistle, piping,
rasing, rich, rough, rumbling, shrill, thick, thin, velvety, whisky,
There are some impressionistic labels whose connotations are extremely difficult to unravel. In most cases they seem more likely to be metaphorical comments on indexical aspects of the speaker, rather than descriptive comments on the sound of the voice itself. These are labels like 'oily, treacly, dry, cool, warm, brittle, sugary'.

2. Some semiotic aspects of impressionistic labels

Having tried to show something of the variety of reference to different aspects of sound production that many impressionistic labels may have, I would like to go on now to a brief discussion, in semiotic terms, of how such reference is achieved by the labels.

For the purpose of this discussion, we can distinguish between two groups of impressionistic labels: those that make direct reference to some aspect of the voice, and those that make a metaphorical reference. There are many labels with an obvious direct reference to auditory aspects of the voice, for example 'raucous, shrill, hoarse, strident'. In some cases there is additionally an iconic element of similarity of auditory impression, between the sound of the word used as a label, and the sound of the type of voice to which the label refers. Examples of auditorily iconic labels for voices would be 'booming, orotund, rumbling, staccato, tinkling, twittering'.

Some labels plausibly make direct reference to articulatory aspects of voice production, as for example 'breathy, nasal, throaty, whispery', and slightly less directly 'guttural', and 'pectoral'.

whispery'.

whispery'.
Occasionally an iconic aspect enters the actual articulatory characteristics of the pronunciation of the word used to label the articulatory aspect of the voice in question, as in 'orotund', where it might be thought that the rounded, protruded lip posture necessary in some accents for the pronunciation of the beginning of the word resembles the characteristic lip setting of some orators declaiming in an 'orotund' style.

Where the reference of impressionistic labels is metaphorical, and not direct, there are many labels which exploit an implication of iconic similarity between the actual auditory impression of a particular voice and an impression in another sensory modality, in synaesthetic metaphors. Most of the labels concern a transference of impression from the auditory modality to the visual, as in 'bleary, blurred, colourless, dark brown, golden, light blue, liquid (?), pale, pale pink, rich (?), silvery', and perhaps 'velvety'. The tactile sense is another perceptual modality involved in synaesthetic labels, as in 'brittle (?), fluid, gravelly, gritty, hard (?), liquid, rough, silky, smooth (?), soft, velvety'.

3. The impressionistic basis of some phonetic labels

We return now to the point made earlier about the impressionistic tinge that exists in many supposedly scientific phonetic labels for voices. A number of terms are used in phonetic labels whose reference could be interpreted by a trainee phonetician only in the same culturally-grounded way that he would interpret impressionistic labels; he could find no support from his familiarity with the basic taxonomy for phonetic description in other areas, - such as the description of segmental articulation, for example. An example
that comes to mind is the relative ease of interpretation of a label like 'nasalized voice', from experience of the technical reference of the term 'nasalized' in other technical phonetic contexts, compared with the vagueness of interpretation of a label like 'creaky voice', as a descriptive phonetic label new to the trainee phonetician, (where he couldn't call on his previous phonetic experience for help with its interpretation, but would have to fall back on whatever direct or metaphorical reference he could appeal to as a layman). Other examples of impressionistically contaminated terms currently used in phonetic taxonomy are 'falsetto, head-register, chest-register, harsh, breathy'. One may concede that any of these terms can quite properly enter the technical vocabulary of phonetics, once a technical definition has been allocated to which, ideally, all phonetician users then subscribe. But one may also question whether the prior descriptive resources of the overall taxonomy of general phonetic theory could not handle the labelling problem more efficiently, by minimising the number of basic terms in the taxonomy and by revealing principles of vocal control of greater generality than would otherwise be evident. One way to do this would be to follow the suggestions by Catford (1964) that the principle of labelling segmental articulation on the basis of location and type of stricture could be extended to the laryngeal components of many types of voices. Catford's 'location of stricture' terms are: 'normal glottal, ligamental, arytenoidal, ventricular'; his 'type of stricture' terms are: 'voiceless, whisper, voice, creak, stop, breathy voice, whispy voice, whispy creak, voiced creak, whispy voiced creak'. Catford points the way to more precise scientific characterizations of the various phonation types he mentions by giving acoustic, aerodynamic and physiological
specifications for them, but defends his continued use of terms like 'whisper' by saying 'The articulatory terms 'fricative' and 'trill' could be applied to phonatory strictures, but, since two types of each of these categories must be distinguished, it is more convenient to use the imitation-labels (....) 'whisper', 'voice', etc.' (1964: 30). Nevertheless one might look forward to the possibility of replacement of such imitation-labels, when more is known about laryngeal physiology, by appropriate 'articulatory' terms of the sort Catford suggests.

The more impressionistic type of phonetic labels in this area, such as 'whispy falsetto voice', could then perhaps usefully continue to stand proxy for the more precise but more cumbersome phonetic definitions, once those definitions can be scientifically justified. But until a precise phonetic explanation of the mechanisms that produce the voices in question is available, it is important to remember that most of this supposedly 'phonetic' part of our professional vocabulary is phonetic only in a limited sense, and has only a limited claim to scientific status. The terms are phonetic in that they are what phoneticians use: as such, they at least serve to isolate auditorily separable components of composite phonetic events, and attach labels of reasonably-agreed reference to them. The terms will become fully phonetic, and more legitimately scientific, only when the articulatory and acoustic correlates of the currently auditorily identified components can be specified. Until that time, their strongly impressionistic foundation of auditory imitation should not be disregarded.

B. Indexical labels for voices

The second of the two most basic types of label are indexical labels. These, it will be recalled, refer to the characteristics
that the listener ascribes to the speaker producing the voice, rather than directly to aspects of the sounds produced by the speaker.

In a technical semiotic sense, the sounds produced by the speaker, in their different aspects, act as indices to his imputed characteristics. In discussing indexical labels, there are thus two factors to be taken into account, - the type of vocal behaviour involved and the type of speaker-characteristic of which it is taken to be an index. In this section, I shall try to show that speaker-characteristics can be divided into a number of groups, and shall make some suggestions about the aspects of vocal behaviour that indicate particular characteristics. That the reader may well occasionally disagree about the claimed correlation between a particular type of characteristic and a given vocal index emphasizes the need for research in this area: once again, however, my object is not so much to secure exact phonetic agreement about the correlates, but rather to make some plausible suggestions about the typology of the correlations.

1. The distinction between intrinsic and extrinsic features of the voice

An initial distinction is drawn between intrinsic and extrinsic features of human voices. Intrinsic features are those that lie outside the control of the individual speaker. Examples of such uncontrollable features are the maximum and minimum possible length of a given speaker’s vocal tract, the maximum possible area of velopharyngeal opening, the length of the vocal folds, factors of mechanical inertia of muscles of a given mass, and so on. They are not culturally-specific, and are not learnable. These intrinsic features are the reflection of the basic physical foundation of the speaker's
vocal apparatus: they can serve as indices only to physical characteristics of the speaker, such as his sex, age, and health. A speaker's intrinsic physical limitations set the limits to the range of potential performance within which he must make his socially and idiosyncratically specific choices of vocal behaviour. All such choices of vocal activity, over which he can exercise any degree of volitional control, constitute the extrinsic features of his voice. Such extrinsic features will include not only the choices of segmental articulation, but also the choices of manipulation of the voice dynamic strand of speech, and all the potentially controllable habitual muscular settings which characterize the manipulable component of his voice quality. Extrinsic features of a person's voice serve as indices to any and all learnable aspects of the persona he projects, and are thus almost entirely social and psychological indices.

2. **Intrinsic indexical labels**

There are rather few indexical labels for voices which could be called intrinsic indexical labels. Making reference primarily to sex, age and health, they would include such labels as 'a man's voice, a young voice, a senile voice, a robust voice', and possibly 'a weak voice'. The phonetic correlates of labels like these are confined almost entirely to features of voice dynamics such as pitch-range and loudness-range, with occasional implicit reference to uncontrollable aspects of voice quality production such as the very fine detail of the mode of vibration of the vocal cords. One type of intrinsic indexical label with rather specific voice quality correlates is the type that identifies transient states of health, such as 'a cold-in-the-head voice', and 'an adenoidal voice'. 
3. Extrinsic indexical labels

There are, on the other hand, very many extrinsic indexical labels for voices: they can be descriptively divided into two major classes, referring to social aspects of the speaker's identity (mostly permanent factors), and to psychological aspects (both permanent and transient).

In parenthesis, it should be said that many extrinsic indexical labels make implicit reference to levels of spoken language higher than that of phonology alone. Thus, 'an upper-class voice' may be indexically signalled not only by phonetic correlates of phonology and paraphonology, but also by details of syntactic, morphological and lexical choices, - that is to say, not only by phonetic features of accent, but also by linguistic features of dialect (Abercrombie, 1956: 43). Such considerations of levels of linguistic behaviour higher than phonology and paraphonology will be left out of account here, and only directly phonetic features will be considered, in their role as extrinsic indices to characteristics of the speaker's identity.

The discussion above of descriptive labels concentrated on the plausible identification of which of the three strands of voice production was implicitly referred to in the label, segmental pronunciation, voice dynamics, or voice quality. In extrinsic indexical labels, there is less possibility at the current stage of knowledge about voices of such phonetic specificity. I should like therefore to move to a level of analysis which is slightly more abstract.

The main concern here, apart from drawing relevant distinctions
between different types of indexical characteristics, will be to identify phonological versus paraphonological aspects of vocal behaviour, and to try to show something of the typology of their correlation with the indexical characteristics denoted by the labels for the voices.

The notion of a speaker's 'accent', and the definition of 'accent' here is the one given a fuller justification in the chapter on semiotic aspects of speech, namely: an individual speaker's accent is made up of all the phonetic correlates of his phonology and paraphonology, together with all the extrinsic habitual muscular settings which characterize his voice quality. Put simply, a speaker's accent consists of all the extrinsic features of his vocal behaviour. The assumption is made in the following discussion of extrinsic indexical labels that the vocal behaviour which is acting as an index to any given labelled characteristic is (with some specified exceptions) some composite part of the speaker's accent.

1) indexical labels of regional origin

The first social characteristic of speaker, for which extrinsic features of the voice can act as an index, is that of regional origin. There are very many labels such as 'an English voice', where 'English' could be replaced by 'American, countryman's, foreign, Liverpool, Yorkshire', and so on. The one comment I would like to make here, in support of the very wide definition of 'accent' given above, is that a phrase such as 'a Liverpool voice' has implications, phonetically, as much for velarized and denasalized aspects of voice quality (Abercrombie, 1967: 94) as it has for the other phonetic features normally associated with accent, - segmental pronunciation and
voice dynamic aspects of phonology and paraphonology.

ii) **indexical labels of social status**

Social status is often the denotation of indexical labels, and the vocal correlate is usually the whole amalgam of the speaker's accent. Indexical labels of this sort include 'upper-class, middle-class, working-class', and possibly such terms as 'superior'. Educational status labels often carry a connotation of social class, when people speak of 'an educated voice', or 'an illiterate voice'; and there is the further possibility here that the aspect of vocal behaviour to which implicit reference is being made may well be aspects of higher-level dialectal choices as well as of features of accent.

iii) **indexical labels of profession**

Indexical labels denoting a profession, particularly when the imputation of a profession is used as a disparaging comment on the sound of a speaker's voice, seem often to be concerned with features of voice dynamics and with extrinsic aspects of voice quality. Examples are: 'a lecturer's voice, a politician's voice, a school-teacher's voice, a sergeant-major's voice, a newsreader's voice'.

iv) **indexical labels of interactional status**

There are areas where social and psychological indexical labels seem to fall together, where comment is made about the nature of the psychosocial interaction in which the speaker is participating. The first group concerns what one might call the interactional status of the speaker, in his relationship with his listener. Adjectives for voices in this circumstance would include such items as 'condescending, flattering, grovelling, patronising, smarmy, whining', and
perhaps 'superior'.

v) indexical labels of type of interaction

A second category concerns the type of interaction involved between speaker and listener, but still with the emphasis on the speaker, in labels such as 'his telephone voice'.

vi) indexical labels of perlocutionary effect

The third category shifts the emphasis to the effect of the voice on the listener in the interaction. Of course, all indexical labels imply a conclusion being drawn by the listener about the characteristics of the speaker, so in that sense all indexical labels imply 'an effect on the listener'; but I have in mind an effect on the listener by means of vocal behaviour that Austin (1962: 101) has called a 'perlocutionary' effect. Labels for such voices would be 'annoying, boring, calming, frightening, interesting, persuasive, soothing, soporific' and so forth.

The first two categories, indices of 'interactional status' and 'type of interaction', would involve features of paraphonology and sometimes of phonology; the third category, 'perlocutionary' indices, is much more difficult to pin down, and may perhaps be the result of a more holistic effect of the total accent of the speaker.

vii) indexical labels of psychological characteristics

Indexical labels for psychological characteristics of a speaker themselves fall into two chief groupings, - those that concern relatively short-term aspects of psychological states, in
their communication of mood or attitude by paraphonological means, and those that concern longer-term aspects of personality. The vocal indices of personality are a highly controversial field (Kramer, 1964), but it seems likely that they involve not only paraphonological behaviour, but also all other features of accent, (and probably of some dialectal features), together with some of the intrinsic features of speech production. In other words, a speaker's personality may be judged by listeners not only on the basis of choices of vocal behaviour he makes, but also possibly to some extent on physical, biological features over which he has no possibility of volitional control. For example, in our culture, a man with a long vocal tract and large vocal cords, with a correspondingly deep-pitched bass voice with low formants, may well have attributed to him personality characteristics of mature authority that have rather little in common with the actuality of his psychological make-up, purely because of our cultural stereotypes of 'authoritative' voices.

Labels for indicating the speaker's mood constitute a long list of terms like 'amused, angry, bored, cold, dry, excited, irritated, remote, sarcastic, sardonic', and so on. Labels for indicating personality include often rather similar terms, such as 'assured, dominating, effeminate, excitable, irritable, neurotic, nervous, virile'.

Another interesting group of indexical labels falls rather uncomfortably between extrinsic and intrinsic indexical labels. These are labels like 'mannelish, womanish, childish'. To some extent they refer to features which are learned, and culturally-
specific, and are therefore in the extrinsic category. There is an intrinsic element involved, however, in that the labels imply a discrepancy between the normal indexical value of the vocal behaviour concerned and the visibly establishable truth about the physical identity of the speaker. One might think that, faced with an apparent discrepancy between the audible intrinsic evidence, and the visible intrinsic evidence, about the speaker's sex and age, the listener believes his eyes rather than his ears, and concludes that the audible evidence is misleading. This may then lead him to draw unfavourable psychological conclusions about the personality of a speaker who is apparently dissembling about such basic aspects of his physical identity as his sex or age.

C. Labels for voices in speech pathology

Lastly, there is one further group of taxonomies for voices which is of semiotic and phonetic interest. That is, the various systems of terminology for discussing voices used by speech pathologists, which are sometimes descriptive impressionistic, sometimes descriptive phonetic, often indexical, and sometimes hybrid. Confining comment here to the taxonomies for voice quality, some of the labels show a fairly high degree of phonetic specificity, as in the use of terms like 'hyperrhinophonia' (or 'hyperrhinolalia') for 'severely nasalized voice', 'hyporhinophonia' (or 'hyporhinolalia') for 'denasalized voice', and 'dysphonia plicae ventricularis' for 'ventricular voice'. Nevertheless, the limitation of such systems, phonetically specific though some of them may be, is in their emphasis on deviations, for whatever etiological reasons, from socially acceptable norms of voice quality. This is not to disparage
such an emphasis; in very many areas of speech pathology attention is necessarily focused on abnormality. However, these systems are inherently too partial for general applicability to voice qualities of all kinds. The advantage of a descriptive phonetic system is its ability to function without prejudice as to culturally-assessed factors of 'normality' or 'abnormality'.

Speech pathology and speech therapy take a fundamentally semiotic interest in voice quality, in that they are concerned with its use as a diagnostic index for a variety of medical, social and psychological states. In their professional communication about the different types of voice quality which characterize their patients, speech pathologists and therapists use labels which can be classified in exactly the same semiotic terms as the labels discussed earlier: they are either descriptive labels to do with the sounds produced by the speaker, or they are indexical labels to do with the characteristics of the speaker producing the sounds.

1. **Symptomatic labels**

The descriptive labels used in this area, which can be either impressionistic or phonetic, can usefully (in the context of the voice as a diagnostic clue) be called **symptomatic labels**. These symptomatic labels fall into two different classes, depending on whether they are concerned with 'functional', non-pathological qualities, or with 'organic', pathological qualities. In the case of non-pathological qualities, the assumption underlying phonetically-based labels, discussed earlier, that 'the vocal apparatus, and all parameters of vocal control, are standard to all human beings' is maintained. This means that symptomatic phonetic labels for non-pathological voice qualities, such as 'dysphonia plicae ventricularis' and
'Hyperrhinophonia', or the more usual phonetic terms 'ventricular voice' and 'nasalized voice', can be used for accurate written communication (vital in adequate case reports), because they represent qualities that any phonetically-trained reader should be able to reproduce for himself. Unfortunately, many descriptive symptomatic labels lack this phonetic base, and as impressionistic labels, fall back on the need for audible demonstration for successful communication. Labels of this sort include terms like 'harsh, hoarse, metallic, muffled'.

In the case of pathological voice qualities, and those due to abnormal anatomy or physiology, the assumption of a 'standard vocal apparatus ....' cannot be maintained, by definition. This means that, in the absence of a comprehensive acoustic scheme of categorization of all major types of voice qualities, normal and abnormal, all symptomatic labels for pathological voice qualities have to be impressionistic, and cannot be phonetic. These labels, needing audible demonstration for accurate communication, include terms like 'diplophonic voice, harsh voice, hoarse voice'. One complication is immediately apparent, - the occurrence of terms like 'harsh' and 'hoarse' in the lists of both non-pathological and pathological qualities. This of course means that it is not possible to tell from the label itself whether the voice quality referred to reflects an underlying pathology or not. One might argue, reasonably, that the labelling of symptoms should in any case be in quite separate terms from the diagnostic decision that succeeds the prior identification of the presenting symptom. This is certainly a valid argument in general terms, but in this case it conceals an undesirable situation:
since 'hoarse', for instance, is not a phonetic label but an impressionistic one, then because an audible demonstration of the quality referred to is not given, the reader is left with the impression that the use of the same descriptive label in the non-pathological and the pathological cases is meant to refer to identical or very similar qualities. However, since one derives from a normal vocal apparatus and the other from an abnormal one, one can have no confidence that the symptoms are necessarily substantially similar.

2. **Etiological labels**

The indexical labels used by speech pathology and speech therapy are concerned with the cause of the voice quality in question, and can therefore be referred to as etiological labels. These etiological labels can be similarly divided into the two classes 'functional' or non-pathological and 'organic' or pathological. In the case of non-pathological etiological labels, reference is often made both to the etiology and to the symptom, as in 'hysterical falsetto', or 'hyperkinetic dysphonia'. Such terms have the disadvantage mentioned above of amalgamating prejudicial aspects of the diagnosis of the underlying cause with the description of the symptoms.

The same difficulty obtains in the case of 'organic' or pathological etiological labels, though here it is more usual to find the etiology specified without reference to the symptom, in terms like 'smoker's voice, adenoidal voice, deaf voice, cleft palate voice, eunuchoid voice, castrato voice', and also in more technical-sounding terms like 'endocrine dysphonia, myopathic dysphonia, spastic dysphonia'. Labels of this sort have to be
audibly demonstrated, or the professionally-involved reader or listener has no advantage over the man-in-the-street in his ability to interpret them.

Conclusion

Finally, it is necessary to note that, as is usually the case in Procrustean typologies like the one we have been discussing, the analysis is not exhaustive, nor is it completely decisive, even between the major divisions. Occasionally, labels for voices just do not fit quite as easily into the particular categories as has been implied above. Sometimes a particular label's referent is so multi-valued that it has to be entered in very many different categories without our being able to give major weighting to any particular category. For example, 'a sepulchral voice', (if this is meaningful to writer and reader as members of the same culture), does have some implications for the sound of the voice involved, particularly with respect to voice dynamics and perhaps voice quality, as noted earlier. But it also has indexical implications for the perlocutionary effect of the voice on the listener, and perhaps for the indexical impression of the personality and mood of the speaker.

However, the purpose of this discussion has been to show that the majority of labels for voices can be typologically classified in the way suggested, and that some plausible phonetic relevance can be given to the wide variety of lay usages for labelling voices, provided that the major semiotic distinction between labels for the sound of the voice versus labels for the imputed characteristics of the speaker is maintained.
Chapter 2. The history of voice quality analysis

A. Classical writings on the voice

The explicit classification of different qualities of the voice, on a phonetic basis, really began only in the nineteenth century. An interest in voice quality goes back, however, to a very much earlier point in time. Hunt (1858) stated that, in ancient Greece:

"The discipline for the formation and improvement of the voice among the Athenians was so comprehensive that, as we are informed by the Roman writers, not less than three different classes of teachers were employed for this purpose, viz. the vociferarii, phonasi, and vocales. The object of the first class seems to have been to strengthen the voice and extend its compass, the office of the second to improve its quality, so as to render it full, sonorous and agreeable; while the efforts of the third, who, perhaps, were considered as the finishing masters, were directed to the proper intonation and inflection" (quoted by Browne and Behnke, 1884: 8).

Stanford (1967: 148-149) says that:

"Besides teachers and performers, two other professional groups in ancient Greece took a special interest in vocal qualities. The physicians (....) listened to them in making their diagnoses. Also the writers on "physiognomics", the art or science of deducing character from physical qualities, had a good deal to say about the supposed relationship between certain kinds of voices and certain kinds of people. Some believed, for example, that people with deep and tense voices were brave, those with high and slack voices cowardly; if disgruntled, one..."
that rises from low to high. If spiteful and morally lax, you are likely to speak with a nasal quality. Greedy and vain people have high, clangy voices like birds; stupid ones bleat like sheep or goats. If you hear a dry quality in someone's voice, look out - he is probably a wily fellow. And a man's cracked or broken tone should warn you against his gluttony and violence. If he talks with pararhotacism like Alcibiades, he must be haughty, proud and hardhearted. So, at least, the physiognomists thought.

Part of this earliest interest in voice quality sprang from the classical concern with oratory, and a useful commentary on Greek and Latin writings on oratory, including voice quality, is Austin's Chironomia (1806). Like Hunt, he refers to the different sorts of teachers concerned with the voice:

'.... Phonasci, Vociferarii (and) Vocales, were the common appellations of those who taught the exercise and management of the voice. Tertullian called them Edomatores Vocis. Galen says they recommended to their disciples the frequent use of the warm bath. Cresollius mentions other practices of the Phonasci, some of which are curious, and some he considers useful' (Austin, 1806: 557)

Commenting specifically on the role of the Phonascus, Austin (1806: 22-23) writes that:

'.... it was the custom for the Roman youth to recite weekly, chosen passages from the poets (....) They frequently employed (....) Phonasci, whose sole business it was to regulate the modulations of the

1. It should be noted that (Ludovicus) Cresollius, like Galen, was not a Roman writer. He was a French Jesuit priest of the seventeenth century called Cresol, who wrote Vacationes Autumnales, one of the many Jesuit books on rhetoric about that time (Sandford, 1938: 72).
voice, to manage it by peculiar regimen, and to administer remedies when it happened to be deranged'.

In a footnote to this, he quotes a passage from Cresollius' *Vacationes Autumnales*, which can be translated as follows:

'Octavius Augustus (...) paid constant attention to the Phonascus, and by these efforts he achieved the object of pronouncing everything with a pleasant and characteristic sort of tone of voice. Galen, who was a very learned and intelligent man, writes of himself "and I adopted what the Phonasci call voice exercises". The Emperor Nero, who took great care of his voice, and gave remarkable attention to it, neither said nor attempted virtually anything without the help of a Phonascus'.

Austin supports this last comment by Cresollius on Nero with a quotation (Austin, 1806: 24) from Chapter 26 of Suetonius' *Nero*, which is translated in Loeb edition as:

'So far from neglecting or relaxing his practice of the art after this, he never addressed the soldiers except by letter or in a speech delivered by another, to save his voice, and he never did anything for amusement or in earnest, without a phonascus by his side to warn him to spare his vocal organs and hold a handkerchief to his mouth'.

A theme running through much of classical writing on voice delivery in oratory was the avoidance of unpleasant quality. Stanford (1967: 148) comments that:

'Another aspect of the speaking voice which the Greeks often mentioned was its quality in terms of timbre and resonance. Modern writers variously describe unpleasant voices as "throaty" or "nasal" or "guttural" or "hoarse" or "thin" or "harsh" or "chesty" or "breathy" and so on'.

(Stanford refers here to Berry (1962) as saying that 'unpleasant voices are variously described as throaty, nasal, denasal, muffled, guttural, whangy, thin, or heavy'). Stanford continues:
Pleasant voices are "rich", "vibrant", "warm", (....)
The Greeks deployed a rich vocabulary for qualities of this kind. They especially disliked hollowness, coarseness, thickness, roughness, breathiness, throatiness, brokenness.

A final point in Austin's Chironomia which is of interest to the study of voice quality illustrates this 'cosmetic' approach, as it were, to pleasant and unpleasant aspects of voice quality. As an Appendix to his book (1806: 553ff) he gives a long list of labels for voice quality from the Greek writer Julius Pollux. Pollux was a 'professor of literature at Athens under the reign of Commodus' (Sweet, 1899: 212), in the second century A.D., the author of a work called Onomasticum. Austin quotes from the fourth chapter of Book 2, in the 1706 Amsterdam edition. Austin's English translations of the Greek and Latin labels listed by Pollux are as follows (only the Latin version of the original is given here). Firstly, for the 'good' qualities:

"altam, high; excelsam, powerful; splendidam, brilliant; mundatam, pure; suavem, sweet; illecebrosam, attractive; exquisitam, melodious, cultivated; persuasibilem, persuasive; pellacem, tractabilem, engaging, tractable; flexilem, flexible; volubilem, executive; dulcem, sweet; stridulam, sonorous, harmonious; manifestam, distinct; perspicuam, perspicuous, articulate".

Secondly, for the 'bad' qualities:

"nigrum, obscure; fuscam, dull; injucundam, unpleasing; exilem, pusillam, small, feeble; angustam, thin; difficilem auditu, molestam, faint; subsurdam, obscuram, hollow, indistinct; confusam, confused; absonam, discordant; inconcinnam, neglectam, unharmonious, uncultivated; intractabilem, unattractive, unmanageable; impersuasibilem, uninteresting; rigidam, rigid; asperam, harsh; distractam, cracked; tristem, doleful;
infirmam, raucam, unsound, hoarse; aeneam, brassy; acutam, shrill, sharp”.

Pollux' list is very unsatisfactory, and in the light of the typological principles suggested in the previous chapter, can be seen to be based on no single criterion of description. This, however, is no less true of many of the terminologies used for the description of voice quality today, as we shall see below.

Two writers were outstanding in the classical period for analytic comment specifically on voice quality. The first was Cicero, in the first century B.C., in his two treatises De Oratore and Brutus. The second was Quintilian, in the first century A.D., in his Institutiones Oratoriae. Of the two, Quintilian made the more extensive and systematic comment on voice quality, but it is as true of both Quintilian and Cicero as it is of almost all writers on the subject until the nineteenth century that they almost entirely failed to distinguish in their descriptive labelling features of voice quality as such from features of voice dynamics such as pitch, loudness, tempo and continuity (Abercrombie, 1967: 95-110), and from features of segmental pronunciation. Even when writers in this span of two thousand years successfully managed to isolate a feature that could reasonably be allocated to voice quality, the labels they chose to use were most often impressionistic imitation-labels (such as Cicero's use of a label like "acute"), or were indexical labels for some physical, psychological or social attribute of the speaker to which the writer imagined the voice quality was acting as an index (as in Quintilian's use of a label like "effeminate").

The conflation of impressionistic description and indexical comment can be seen in the following extracts from Cicero's
treatises De Oratore and Brutus, (in Watson's translations (1899)):

'... any illiterate Athenian will easily surpass the most learned Asiatics, not in his language, but in sweetness of tone, not so much in speaking well as speaking agreeably. Our citizens pay less attention than the people of Latium, yet among all the people that you know in the city, who have the least tincture of literature, there is not one who would not have a manifest advantage over Quintus Valerius of Sora, the most learned of all Latins, in softness of voice, in conformation of the mouth, and in the general tone of pronunciation' (De Oratore, Book III, c.XI).

(on Antonius) '... his voice was strong and firm, though naturally hoarse ...' (Brutus, c.XXXVIII). (on Cnaeus Pompeius) '... his voice was sonorous and manly' (Brutus, c.LXVIII). (on Catulus) '... his reputed purity of diction was owing chiefly to the sweetness of his voice and the delicacy of his accent' (Brutus, c.LXXIV).

Cicero also discusses the communication of attitude and emotion in speech, in his commentary on 'tones':

'For every emotion of the mind has from nature its own peculiar look, tone, and gesture; and the whole frame of a man, and his whole countenance, and the variations of his voice, sound like strings in a musical instrument, just as they are moved by the affections of the mind. For the tones of the voice, like musical chords, are so wound up as to be responsive to every touch, sharp, flat, quick, slow, loud, gentle; and yet, among all these, each in its kind has its own middle tone. From these tones, too, are derived many other sorts, as the rough, the smooth, the contracted, the broad, the protracted, and interrupted; the broken and divided, with varieties of
modulation; for there is none of these, or those that resemble them, which may not be influenced by art and management; and they are presented to the orator, as colours to the painter, to produce variety' (De Oratore, Book III, c.LCIII).

Cicero's notion that there are certain features of communication of emotional states which are universal to all human beings, on which are superimposed modulations which are (presumably) culturally-relative, recurs repeatedly in writings on the voice up to the present day.

Quintilian, in his (very modern-sounding) Institutiones Oratoriae, was the first writer to try to separate features of voice quality from those of voice dynamics. He had only partial success, but his ideas can be seen to have anticipated major distinctions drawn in the description of voice quality in current phonetic writing. He also distinguished between a speaker's natural voice (the intrinsic features, in the terms of this thesis) and the aspects of voice capable of voluntary control by the speaker.

Quintilian's often quite phonetically-specific ideas can be seen in the following extracts from c.III of Book XI of his Institutiones, (again translated by Watson, 1899):

'.... the first thing to be considered is what sort of voice we have, and the next, how we use it. The natural power of the voice is estimated by its quantity and its quality. Of these, the quantity is the more simple consideration, for it may be said in general that it is either much or little; but between the extremes of these quantities there are many diversities, and many
gradations from the lowest tone to the highest, and from the highest to the lowest. Quality is more varied; for the voice is either clear or husky, full or weak, smooth, or rough, of smaller or larger compass, hard or flexible, sharp or flat (....)

The general tone of the voice, however, ought to be sweet, not grating.

In the management of the voice there are many particulars to be observed: for besides the three main distinctions of acute, grave, and intermediate, there is need of many other kinds of intonation, as the forcible and the gentle, the higher and the lower; and of slower or quicker time. But between these varieties there are other intermediate varieties: and as the face, though it consists of very few features, is infinitely diversified, so the voice, though it has very few variations that can be named, has yet a peculiar tone in each individual; and the voice of a person is as easily distinguished by the ear as the face by the eye.

It would be tempting to see in this last passage an adumbration of a componential approach to voice description, where the 'infinite diversification' Quintilian writes of should come about by the combinations of a small number of basic components, with each component divisible into very many gradations. But it seems more likely that this diversification should properly be thought to mean the variability of the qualities of individual voices due at least in part to the involvement of the variety of intrinsic features.

Like Pollux, mentioned earlier, Quintilian was interested in the 'cosmetic' aspects of voice production, in oratory at least:

'As to rules for delivery, they are precisely the same as those for language.
For as language ought to be correct, clear, elegant, and to the purpose, so delivery will be correct, that is, free from fault, if our pronunciation be easy, clear, agreeable and polished, that is, of such a kind that nothing of the rustic or the foreign be heard in it; for the saying Barbarum Graecumv, that a man is "Barbarian or Greek" is not without good foundation, since we judge of men by their tones as of money by its clink" (Institutiones, Book XI, cIII).

This last evocative phrase, 'we judge of men by their tones as of money by its clink', could hardly put the function of the extrinsic, controllable element of voice quality, as an index of social and psychological characteristics, more concisely. Quintilian continues his comment on 'cosmetic' qualities of the voice, saying that:

'If the voice, too, be naturally, so to speak, sound, it will (....) not be dull sounding, gross, bawling, hard, stiff, inefficient, thick, or on the contrary, thin, weak, squeaking, small, soft, effeminate (....).

That delivery is elegant, which is supported by a voice that is easy, powerful, fine, flexible, firm, sweet, well-sustained, clear, pure, that cuts the air and penetrates the ear' (ibid.).

His discussion of cosmetic aspects leads Quintilian on to comment on indexical factors to do with the effect on the voice of various sorts of vocal abuse, and of fatigue:

'..... the voice must not be strained beyond its natural power, for, by that means, it is often choked, and becomes less clear the greater the effort that is used; and sometimes, if urged too far, it breaks out into the sound to which the Greeks have given a name from the crowing of young cocks'.

'..... But the good qualities of the voice, like those of all
our other faculties, are improved by attention and
deteriorated by neglect. The attention to be paid to the
voice by orators, however, is not the same as that which is
required from singing-masters; though there are many
things equally necessary to both; as strength of body, for
instance, that the voice may not dwindle down to the weak
tone of eunuchs, women, and sick persons (....). It is
necessary that the throat be in good condition, that is,
soft and flexible, for by any defect in it the voice may be
rendered broken, husky, rough or squeaking; (....) the
throat, when swollen, strangles the voice, when not clear,
stifles it, when dry, roughens it, and when affected by
spasms, gives forth a sound like that of broken pipes.
(....) Too much moisture also impedes the voice, and too
little weakens it. As to fatigue, it affects the voice as
it affects the whole body, not for the present merely, but
for sometime afterward' (Quintilian, Institutiones,
Book XI, c.III).

Like Cicero, Quintilian has something to say about the function
of the voice in the communication of emotional states and attitude:

'.... the voice is the index of the mind, and has as many
variations as the mind itself. Hence, in speaking on
cheerful subjects, it flows in a full and clear tone, and
is itself, as it were, cheerful; in argument, it arouses
itself with its full force, and strains, so to speak, every
nerve; in anger, it is fierce, rough, thick ....' (ibid.).

B. **Voice quality and paralinguistic description**

We see thus that extrinsic aspects of speech were of interest
to Cicero and Quintilian in their paralinguistic function of acting
as 'affective indices' (Abercrombie, 1967: 9) to a speaker's
ostensible attitudinal and emotional state. This clearly derives
directly from their concern with spoken delivery as an art that can be cultivated, and we shall repeatedly find that comments on voice quality are often found embedded in a discussion of paralinguistic uses of the voice. There are two facets here which are of immediate interest to the study of voice quality.

The first is that the description of paralinguistic features is almost always relevant to the description of potential extrinsic voice quality features, because the main difference between vocal paralinguistic features and voice quality's extrinsic component is the time scale involved, and not so much the identity of the features used (Laver, 1975). Paralinguistic features are used on a relatively ephemeral, short-term basis, while the same features, when used as extrinsic components in voice quality are quasi-permanent. The way that different writers in the history of phonetics go about the business of constructing a descriptive vocabulary for handling paralinguistic features is thus interesting to a study of voice quality, precisely because many of the same phenomena are involved in the two areas.

The second facet is one which will be discussed more fully in a later chapter: that is, analytic decisions about which phonic events should be deemed linguistically-relevant in a particular accent, which paralinguistically-relevant, and which relevant to voice quality, cannot be taken in isolation. All three strands have to be considered together, and a decision about any one of them has immediate reciprocal implications for the other two. While analytic schemes for describing the phonetic substance of linguistic units are thoroughly well-established, the same cannot be said for the substance of paralinguistic
units; we have an opportunity here to see the description of voice quality against the background of emerging systems for the description of spoken paralanguage.

For both these reasons, therefore, this historical outline of writings on voice quality will include material about paralinguistic communication, found in the literature of phonetics from the sixteenth century onwards.

There is one further qualification that needs to be made before coming to the work of the early British phoneticians. That is, many writers from the time of Wallis (1653) to the present day comment specifically on the hypothesized long-term extrinsic settings of the vocal apparatus which they held to characterize the phonetic aspects of different languages. The concept of an articulatory setting is so central to a phonetic approach to describing voice quality that the account of the historical development of the concept is included in the next chapter, which sets out the descriptive phonetic system, rather than in this chapter.

Leaving detailed discussion of language-characterizing settings aside, this chapter will be concerned with two other broad aspects of the history of voice quality analysis. We shall trace the development of the various comprehensive schemes for classifying voice quality on a general, componential basis, which begin to appear in the nineteenth century; and we shall try to follow a number of threads of development in the earlier, less systematic observations on voice quality and paralinguistic features that are scattered through the phonetic literature.
C. Early British writings on the voice

The first mention of voice quality in early British works seems to be by Roger Ascham, in *The Scholemaster* (1570: 8), when he writes of '.... a voice, not softe, weake, piping, womanishe, but audible, stronge, and manlike ....'

Writers in the seventeenth century, while never approaching even a rudimentary general description of voice quality, quite often noted small details about voices, such as the cause of hoarseness, for example. Wallis (1653: 4, translated by Kemp, 1972), discussing voicing and the larynx, says 'The hoarseness which often accompanies catarrh originates in the same place, hindering this vibration of the larynx and trachea'. Kemp (p.135, op.cit.) points out that the first three editions of Wallis' *Grammatica Linguae Anglicanae* were worded slightly differently from the sixth edition which was the basis for Kemp's translation and commentary, and that the relevant passage in those three editions made the catarrh itself the factor hindering the vibration of the larynx, not the hoarseness accompanying it. Kemp (ibid.) also draws attention to similar comments in Holder (1669: 113), Cooper (1685, translated by Jones, 1911: 2-3), Amman (1700: 29) and Petrus Montanus (1635, Book 2, p.32ff). He says that:

'Amman gives fever and shouting as causes of hoarseness, and explains it as due to an excess of mucus from the glands, which deprives the cartilages of their elasticity. Petrus Montanus (....), on the other hand, attributes it to dryness of the throat' (Kemp, 1972: 135).

It may be remembered that Quintilian, like Amman, attributed impediments in the voice to the presence of too much moisture, as noted above.
Holder (1669: 117) also comments that:

'... it will be painful and irksome to a deaf Person to exercise his Voice, as even those, who have no defect, are apt to be tired and spent with much speaking, and find a hoarseness in their Voice, and weariness in the Lungs and Muscles of the Larynx'.

Sibscota (1670), in a free translation of an essay from a collection by Anthony Deusingen, entitled Fasciculus Dissertationum Selectarum, published in Gröningen (1660: 147-230), continues this commentary on the speech of deaf speakers:

'And sometimes there may be some peculiar defect in the structure of the Ear, extending itself to the Jaws, and Palat of the Mouth, which is the reason that those who are Deaf use to speak through the nose, as Aristotle affirms in his second and fourth Problem (....)' (Sibscota, 1670: 14).

He goes on to say:

'... those that are thick of Hearing, have a kind of hoarse speech' (i.e. 'those, whose deafness ariseth from the ill-affection of the Nerves of the fifth pair') (Sibscota, 1670: 28).

Early in the eighteenth century, Maittaire (1712) wrote about the extrinsic, controllable aspects of voice quality and voice dynamics, in a passage which, without acknowledgment, is mostly an extension of some of Quintilian's comments quoted above:

'The Voice Two things in it are carefully to be observed; what voice you have, and how to use it.

It may be, as to its Quantity, Great or Small.

As to its Quality, Clear or Thick, Full or Slender, Soft
or Harsh, Contracted or Spread, Hard or Easy to be managed, Sharp or Blunt.

The Breath is either Long or Short.

The Good Qualities, as of all other things, so of the Voice are bettered with Care, and impaired through Negligence. Frequent Exercise, Temperance and Frugality conduce much to their improvement.

(....) 'Tis a fault of the voice to be too much stretched or rowling: the mouth is best, when it is Ready, not Precipitate; Moderate, not Slow. (....). Nothing can be worse than a Tone or Cant. A true Pronunciation is ever suited to what we speak. The Affections are either Real and Natural, which need no Art: or else Feigned and Put on; and in these the great Art is to be first moved with them, as if they were Real; then the Voice, as a faithful Interpreter of the Mind, will convey what impressions it has received from our Soul, into those of the Judges or Auditors. It is capable of as many Changes as our Minds; Easy in Cheerful Matters; Erect and Firm, when we strive as for the Mastery; Fierce, Harsh and Thick in Anger; Soft, in Begging; Grave, in Persuading; Short, in Fear; Strong, in Exhortation or Narration, Even, between Acute and Grave; in short, it Riseth or Falleth, as the Affections are Raised or Composed' (Maittaire, 1712: 239-240).

Maittaire also includes an interesting comment on facial expression in speech: 'Small is the motion of the Nostrils and Lips, unless in Scorn and Contempt. We ought always to speak more with the Mouth than the Lips' (op.cit., p.241).

According to R.C. Alston, the Essay on Elocution by Mason (1748), 'has a particular historical significance because it represents a renewed interest in a neglected art. The interest in pronunciation which is so characteristic of
writers on English in the second half of the eighteenth
century, and finds expression in numerous treatises on
pronunciation and elocution as well as in dictionaries of
English, can be traced back to sixteenth century manuals
of pulpit oratory, but the movement which was to reach
fruition in the works of writers such as Sheridan and
Walker has its origins in a complex coincidence of
interests: among which are the improvement of dramatic
speech (to which Garrick contributed so effectively)
and parliamentary debate; the public recitation of
poetry; and a concern to "fix" pronunciation and
formulate a universal standard. Mason's Essay thus
appears at the beginning of an important movement'
(Alston, in editorial note to facsimile reprint (1967)
of Mason, 1748).

Although Mason shared 'a concern to fix pronunciation and formulate
a universal standard', and contributed some impetus to the movement
Alston describes, he can also be seen as continuing a very long-
established tradition of an interest in prescriptive cosmetic aspects
of voice, started in the classical period and notable in the works of
Cicero and Quintilian. Alston acknowledges this, and says that
the Essay's

'Immediate significance can be seen as related to the
re-appraisal of the great classical treatises on oratory
by Longinus, Cicero and Quintilian. William Smith's
translation of Longinus (1739) was to be reprinted
numerous times between 1740 and 1800; William Guthrie's
translation of Cicero's De Oratore appeared in 1742 (...) and his translation of Quintilian followed in 1756(...).

Mason's Essay was reprinted in the same year as the first
edition, and again in 1751, 1757, 1761 and 1787' (Alston,
ibid.).

Mason himself was quite explicit about the value of classical writings:
'Those who desire to be more particularly acquainted with this Subject, and with the several other Branches of Oratory, I would advise not to trust altogether to the Rules of modern Writers, but to repair to the Fountain Head; and converse with the great Masters and Teachers of this Art among the Antients; particularly, Dionysius of Halicarnassus, Cicero, Quintilian, and Longinus' (Mason, 1748: 39).²

Mason begins with some sensible comments on using a conversational tone of voice in public speaking and reading aloud:

'To avoid all kinds of unnatural and disagreeable Tones, the only Rule is to endeavour to speak with the same Ease and Freedom as you would do on the same Subject in private Conversation. You hear no body converse in a Tone; unless they have the Brogue of some other Country, or have got into a Habit (as some have) of altering the natural Key of their Voice when they are talking of some serious Subject in Religion. But I can see no Reason in the World, that when in common Conversation we speak in a natural Voice with proper Accent and Emphasis, yet as soon as we begin to read, or talk of Religion, or speak in Publick, we should immediately assume a stiff, awkward, unnatural Tone. If we are indeed deeply affected with the Subject we read or talk of, the Voice will naturally vary according to the Passion excited; but if we vary it unnaturally, only to seem affected, or with a Design to affect others, it then becomes a Tone and is offensive' (Mason, 1748: 17-18).

The notion of a 'Tone', although rather vague, was widespread in writings on elocution in the eighteenth century. Very often, it seems to be used in the same way that we use the term today in

² It may be necessary to mention here that while Cicero and Quintilian wrote at some length on aspects of spoken delivery relevant to voice quality, neither Dionysius of Halicarnassus nor Longinus gave it any particular attention.
phrases like 'tone of voice'; that is, as a cover term for a wide variety of phonetic phenomena used as paralinguistic cues for conveying emotional and attitudinal information. But there is also a tendency to use 'Tone' to refer to paralinguistic behaviour that the writer condemns as insincere or inappropriate for the particular situation, as in the last sentence of the passage quoted from Mason immediately above, — '.... but if we vary (the Voice) unnaturally, only to seem affected, or with Design to affect others, it then becomes a Tone and is offensive'. Appropriacy of vocal cues to the speaker's attitudinal or emotional state seems often to be judged on the basis of an imagined iconic resemblance between the vocal behaviour and the mental state to which it is taken to be acting as an index. This may be seen in the following passage from Mason (1748).

'The Voice must express, as near as may be, the very Sense or Idea designed to be conveyed by the emphatical word; by a strong, rough and violent, or a soft, smooth, and tender Sound.

Thus the different Passions of the Mind are to be expressed by a different Sound or Tone of Voice. *Love*, by a soft, smooth, languishing Voice; *Anger*, by a strong, vehement, and elevated Voice; *Joy* by a quick, sweet, and clear Voice; *Fear*, by a dejected, tremulous, hesitating Voice; *Courage*, hath a full, bold, and loud Voice; and *Perplexity*, a grave, steady, and earnest one. Briefly, in *Exordiums* the Voice should be low; in *Narrations*, distinct; in *Reasoning*, slow; in *Persuasions*, strong: it should thunder in *Anger*, soften in *Sorrow*, tremble in *Fear*, and melt in *Love*' (Mason, 1748: 25-26).
This attribution of indexical characteristics of the speaker to the speaker's voice is easily understandable if we remember the very strong influence on writers such as Mason (and Maittaire) of Quintilian, who was quoted earlier as saying that

'... the voice is the index of the mind, and has as many variations as the mind itself. Hence, in speaking on cheerful subjects, if flows in a full and clear tone, and is itself, as it were, cheerful; in argument, it arouses itself with its full force, and strains, so to speak, every nerve; in anger, it is fierce, rough, thick ....' (Quintilian, Institutiones Oratoriae, Book XI, c.III).

Mason also mentions one specific type of inappropriate use of a 'Tone':

'It is false Oratory (....), to seek to persuade or affect by mere Vehement of Voice. A Thing that hath been often attempted by Men of mean Furniture, low Genius, or bad Taste, among the Antients as well as the Moderns. A Practice which formerly gave the judicious Quintilian great Offence: Who calls it not only clamouring, but furious Bellowing; not Vehemence, but downright Violence.

Besides, an overstrained Voice is very inconvenient to the Speaker, as well as disgusting to judicious Hearers. It exhausts his Spirits to no Purpose. And takes from him the proper Management and Modulation of his Voice according to the Sense of his Subject. And, what is worst of all, it naturally leads him into a Tone' (Mason, 1748: 7-8).

Bayly (1758), in a phonetically sophisticated book, discusses 'tones', and distinguishes clearly between intrinsic and extrinsic features of the voice:
The voice itself is indeed a gift of nature; but with respect to the tone it is extremely in the power of affectation, or ill habit to hurt it, and of art to improve it. The most remarkable ill tones are perhaps such as arise from what is called speaking through the nose and in the throat. Of guttural tones there is great variety. Some are like the bleating of a sheep, or noise of a raven; some resemble the croaking of a frog, and quacking of a duck: All which seem to be owing to some trick of compressing the wind pipe in such a manner as to confine the tone in the throat instead of letting it pass freely out. The voice is also often hurt by another trick: that of shutting the teeth, and confining the tone within the mouth instead of opening the teeth and lips properly so as to bring it out with fulness and rotundity (Bayly, 1758, Pt.3: 180-181).

Thomas Sheridan (1762) follows Cicero in asserting the universality of vocal means for communicating strong emotions, and he distinguishes very clearly between language and 'tones':

'Every one will at once acknowledge that the terms anger, fear, love, hatred, pity, grief, will not excite in him the sensations of those passions, and make him angry or afraid, compassionate or grieved; nor, should a man declare himself to be under the influence of any of these passions, in the most explicit and strong words that the language can afford, would he in the least affect us, or gain any credit, if he used no other signs but words. If any one should say in the same tone of voice that he uses in delivering indifferent propositions from a cool understanding, "Sure never any mortal was so overwhelmed with grief as I am at this present" (.....). Sure, no one would feel any pity for the distress of the (speaker) (.....). We should either believe that he jested, or if he would be thought
serious, we should be moved to laughter at his absurdity. And why is this? But because he makes use of words only, as the signs of emotion, which it is impossible they can represent; and omits the use of the true signs of the passions, which are the tones, looks and gestures.

This will serve to shew us that the language or sensible marks by which the emotions of the mind are discovered, and communicated from man to man, are entirely different from words, and independent of them. (...) the language of the animal passions of man (...) should be fixed, self-evident, and universally intelligible' (Sheridan, 1762: 100-102).

In 1781, Sheridan gets closer to a modern conception of the coded, arbitrary nature of paralinguistic features, while still maintaining the universality of affective indices to the most basic emotional states:

'Tones may be divided into two kinds; natural, and instituted. The natural, are such as belong to the passions of man in his animal state; which are implanted in his frame, by the hand of nature; and which spontaneously break forth, whenever he is under the influence of any of these passions. These form a universal language, equally used by all the different nations of the world, and equally understood and felt by all. Thus, the tones expressive of sorrow, lamentation, mirth, joy, hatred, anger, love, pity, etc. are the same in all countries, and excite emotions in us analogous to those passions, when accompanying words which we do not understand.

The instituted tones, are those which are settled by compact, to mark the different operations, exertions, and emotions of the intellect and fancy, in producing their ideas; and these in a great measure differ, in different
countries, as the languages do.

The former of these, it is evident, neither require study nor pains, when we are ourselves under the influence of any of those passions, as they are necessarily produced by them; (...).

With respect to the latter, it will require great pains, and much observation, to become master of them' (Sheridan, 1781: 120-121).

Nowadays, we would prefer to believe that many of his 'natural tones' are more culturally-relative, more 'instituted', than Sheridan suggests, but he puts the distinction between learned and unlearnable elements in paralanguage very succinctly.

Fenning (1771: 180-181), without acknowledgement, copies Mason (1748: 25-26), in the passage quoted above ('Love (is expressed) by a soft, smooth, languishing Voice' etc.), for which Fenning has 'Love is expressed by a soft, smooth, languishing tone' etc.

Herries, in his book The Elements of Speech (1773), makes some of the most perceptive and phonetically interesting comments on voice quality to be found before the nineteenth century. He writes, for example, that

'... Others, who are not accustomed to expel their breath with the same freedom through the nostrils as through the mouth, pronounce the three nasals m, n, and ng, very imperfectly, which produces that dull disagreeable sound, which we call sneveling or SPEAKING THROUGH THE NOSE. The latter term is entirely wrong, because it is the defect of NOT speaking thro' the nose which occasions that impropriety in articulation. Sometimes this habit
arises from an excess in taking snuff, which ought always to be avoided by a publick speaker or singer' (Herries, 1773: 55-56).

This comment still applies to current writing in phonetics, where the term 'denasal' is applied to a quality of the voice which is not lacking in nasal resonance, but rather has a special sort of nasal resonance. This is discussed at greater length in the section on velopharyngeal settings.

Herries also has a nice comment on a particular type of voice used in public speaking:

"Many of our public speakers have had their powers of utterance enervated and restrained in their younger years. At home and at school they have been put under a false REGIMEN. They perhaps have been told "that it was not PRETTY, it was not genteel to breathe too strongly, to roar out the words, and bellow like a clown; it was quite vulgar, it grated their ears, it was enough to fright a person". The young gentleman takes the hint. He begins to speak FINE. He minces out his words, and warbles his modulations like an Italian singer. What is the consequence? His voice as he grows up retains the same unmanly quality. He dare not, he cannot exert it. He speaks upon the most important, the most alarming subjects, with the delicate tone of a waiting-gentlewoman.

Let this effeminate mode of education be banished from our land' (Herries, 1773: 99-100).

This theme is continued, in a discussion about the function of the voice as an indexical clue to physical features such as a speaker's physique, and the possibility of drawing wrong indexical conclusions:
"Do we not often behold men of the most robust habit of body, speaking in public, with such a weak, puerile voice, that if we were to trust to the testimony of our ears alone, we should conclude that their constitution was quite enfeebled and decayed. From hence we may infer, that a strong BODY is not always accompanied with a strong VOICE" (Herries, 1773: 106).

Finally, Herries discusses some details of the inter-relationship of factors of pitch with those of voice quality:

'The true criterion of just speaking is, when each of the articulate sounds is uttered forcibly and distinctly. But we find that whenever we go beyond our natural pitch, we lose the command of articulation. Our tones are weak, shrill, and broken. Every excess of passion has a tendency to straiten the glottis, and render the voice more acute. This we may observe in the sharp, hurrying voice of anger, the plaintive wailings of grief, the clear-gliding warblings of joy. If, therefore, a public speaker is deeply animated with his subject, his voice insensibly ascends, and sometimes is carried to such a pitch that he loses all command of it. Cicero informs us, that when Gracchus, an eminent pleader at Rome, was in the vehement parts of his discourse, his voice became too high and squeaking. To remedy this inconvenience, he placed a servant behind him, with a pitch-pipe in his hand, who, at such a time, sounded a note in unison with the medium of his voice, on which he immediately descended to his usual sweetness' (Herries, 1773: 152).

It may be of interest here to note that James Rush, who was a major figure in the later history of voice quality analysis, was very scathing about this anecdote, and about classical writings generally. In his characteristically jaundiced and irascible tone,
he wrote that:

'If one should be disposed to believe in the vocal perfection of the Greeks, through any other than their own testimony, he might well question the authority of their Roman eulogists: since they themselves, the pupils of the Greeks, display no better analysis and system in their institute of elocution. We may fairly estimate their discrimination, when with the same pen that deals out the extravagancies of praise upon the oratorical action of their masters, they gravely give us, as proof of their own nicety in vocal matters, the story of one of their famous orators having occasion for a Pitch-pipe, to enable him to recognize his own voice, and affectedly to govern his melody, through the more acute perceptions of a slave, who now and then blew this little regulating trumpet at his elbow!' (Rush, 1859: 675n (5th edition of Philosophy of the Human Voice (1827)).

D. Writings on voice quality in the 19th and 20th centuries

In the nineteenth century, attempts at explicit, general componential schemes for describing voice quality begin to appear. Interest in the subject of delivery, tone of voice and voice quality was still maintained (particularly in America). We have noted Austin's Chironomia (1806). Although this had only one British edition, and no American edition, Haberman (1954: 117-118) says:

'it exerted an enormous influence upon elocutionists. In England, A.M. Hartley called it "incomparably the ablest treatise on delivery in general that has yet appeared in our language". In America, a lot of writers, among them Caldwell, Bronson, Bacon, Fulton and Trueblood, and, as late as 1916, Joseph A. Mosher, were indebted to this extraordinary book'.
This is to overestimate its value, but it gives us an idea of the general interest during the nineteenth century in the voice.

In Britain during this time, Willis (1829) wrote that 'Thus we say that a man has a clear voice, a nasal voice, a thick voice ....' (as distinct from considerations of the speaker's vowel qualities, that is), and by Sweet's time, attempts to set up a voice quality classification on a phonetic basis were becoming frequent (in America, at least), although none of them began to approach any degree of comprehensiveness. Sweet (1877: 97-99) distinguished between 'clear', 'dull', 'harsh' and 'nasal' qualities, and expanded this list in his Primer of Phonetics (1890b: 69), to the following set of qualities: 'clear quality', 'dull quality', 'nasality', 'wheeziness' and 'gutturality'. Sweet was probably the first writer to assert quite explicitly that voice quality is susceptible of systematic componential description, when he wrote that 'besides the various modifications of stress, tones, etc., the quality of the voice may be modified through whole sentences by various glottal, pharyngeal and oral influences' (1877: 97). This position was re-asserted in The History of Language (1900: 136), when he wrote

'.... the general quality of the voice is likely to be modified by changes in the shape of the throat and mouth passages, which give rise to the various qualities of voice known as clear, dull, muffled, nasal, wheezing, strangled voice ....'.

He also drew a clear distinction between intrinsic features of the voice and extrinsic modifications, when he wrote that modifications of voice quality

'must be carefully distinguished from those which are due
to peculiarities in the organs of speech themselves. Thus defects in the palate may cause permanent nasality (together with a peculiar hollowness of sound), an abnormally large tongue, gutturality, etc. All of these peculiarities are inseparable from the individual (1877: 99).

It is perhaps a little surprising that Sweet stopped short, in considering intrinsic aspects of the voice, at those which arise from some abnormal anatomical factor, and didn't go on to the logical conclusion that nearly all speakers can be distinguished on the basis of differences of intrinsic anatomy even within the range of whatever might be considered normal variations of anatomy.

The interests of voice quality classification in the nineteenth century were probably most advanced, though not always very scientifically, by the efforts of the American elocutionists.

The first and most influential of those, in the study of voice quality, was James Rush. Son of one of the signers of the Declaration of Independence, Dr. Benjamin Rush, James Rush was a physician who was educated at Princeton, and at the University of Pennsylvania, where he was awarded his M.D. degree. He then spent a year at the University of Edinburgh in 1810, as a student of Moral Philosophy under the Scottish philosopher Dugald Stewart. It was during the year he spent in Edinburgh that he became interested in the writings of Bacon (Gray and Hale 1943), and from then on put great emphasis, following Bacon, on the value of objective scientific observations, which had its influence on his approach to the description of speech.

In 1827, he published the first of six editions of The Philosophy
of the Human Voice. Hale (1954: 226) comments that 'In publishing a vocal philosophy which gave a physiological foundation and explanation of vocal theory, Rush gave an entirely new and different emphasis to the study and teaching of speech'. (By 'the study and teaching of speech' Hale is referring to the American tradition of interest in rhetoric elocution).

Rush gives an extensive commentary on the description of voice quality. He deserves to be considered in some detail; the quotations given below are taken from the enlarged fifth edition published in 1859.

Rush, early in his book, gives a very clear enunciation of one of the most basic characteristics of phonetic analysis:

'A description of the different modes and forms of sound in the human voice, without exemplification by actual utterance, is always insufficient and often unintelligible. With a view to facilitate instruction, it is desirable to discover the mechanical movements of the organs, together with action of the air upon them; that a reference to conformations and changes of the organs, and to the impulses of the air, may enable an observer to exemplify to himself, the description of vocal sounds, by using the known physical means which produce them' (Rush, 1859: 131).

Rush begins with his divisions of 'vocal sound':

'All the constituents of the human voice, may be referred to the five following Modes: Quality, Force, Time, Abruptness, Pitch. The details of these five modes, and of the multiplied combination of their several forms, degrees, and varieties, includes the enumeration of all the Articulation and the Expressive powers of speech'
The first mention of qualities of the voice (in the broad sense of 'voice quality' used in this thesis, not in the sense of 'phonatory quality' alone, states that

'The thirty-five elements of speech may be heard under four different kinds of voice; the Natural, the Falsette (sic), the Whispering, and that improved quality, to be presently described under the name of the Orotund' (Rush, 1859: 138).

Rush discusses the 'natural' voice in these terms:

'The natural voice is said to be produced by the vibration of the glottis. This has been inferred, from a supposed analogy between the action of the human organ, and that of the dog, in which the vibration has been observed, and on exposing the glottis during the cries of the animal; and from the vibration of the chords, by blowing through the human larynx, when removed from the body. The conclusion is therefore probable, but until it is seen in the living function of the part, or until there is sufficient approximation to this proof by other means, it cannot be admitted as a portion of exact physiological science' (Rush, 1859: 139).

It is a little surprising that Rush, by 1859, had not heard of Manuel Garcia's success with a dentist's mirror as a laryngoscope in seeing the vocal folds in action, reported in 1854.

'Falsette' is described as follows:

'The Falsette is a peculiar voice, in the higher degrees of pitch, beginning where the natural voice breaks, or outruns its compass. The piercing cry, the scream, and the yell are various forms of the falsette. (....) The striking difference in quality, between the natural and
the falsette voices, has created the idea of a difference in the respective mechanisms, not only of their kind of sound, but likewise of their pitch. It has been supposed, that the falsette is produced at the "upper orifice of the larynx, formed by the summits of the arytenoid cartilages and the epiglottis" (Dodart): and the difficulty of joining it to the natural voice, which is thought to be made by the inferior ligaments of the glottis, is ascribed to the change of mechanism in the transition. On this point I have only to add that the falsette (.....) may be brought downward in pitch, nearly to the lowest degree of the natural voice; (.....) and since the natural voice may by cultivation be carried above the point it instinctively reaches, it suggests the inquiry, whether these voices may have a different agency of mechanism (.. rather than by ..) an extension of the powers of the same organization' (Rush, 1859: 142).

Rush starts his discussion of 'Whispering' by saying that 'The Whispering voice is well known', but goes on to declare that

'We are not acquainted with the mechanical cause of whisper, as distinguished from vocality. It has been ascribed to the operation of a current of air on the sides of the glottis, while its cords are at rest; whereas vocality is said to proceed from the agitation of the air by the vibration of those cords. This however is merely an inference from analogy, and has a claim to possibility, but no more' (Rush, 1859: 146).

A longer section is devoted to the description of 'Orotund' voice:

'The voice now to be described, is not perhaps in its mechanism, different from the natural; but it is rather to be regarded as an eminent degree of fulness, clearness and smoothness in quality; and this may be either native or acquired.
The limited analysis, the vague history of speech by the ancients, and the further confusion of the subject by commentators upon them, leave us in doubt whether the Latin phrase, 'os rotundum'; used more to our purpose in its ablative, 'ore rotundo', by Horace, in complimenting Grecian eloquence; referred to the construction of periods, the predominance or position of vowels, or to quality of voice. Whatever may have been the original signification of the phrase, the English term 'roundness of tone', specifying as we may suppose, the kind or quality, seems to have been derived from it' (Rush, 1859: 151).

Rush goes on to give a specification of what he meant by 'orotund' voice, although he concedes that he knows 'how difficult it is to make such descriptions definite, without audible illustration' (p.152). He writes that

'On the basis of the Latin phrase, I have constructed the term Orotund; to designate that assemblage of attributes which constitutes the highest character of the speaking voice.

By the Orotund, or adjectively, the Orotund voice, I mean a natural, or improved manner of uttering the elements with a fulness, clearness, strength, smoothness, and if I may make the word, a sub-sonorous quality; rarely heard in ordinary speech, and never found in its highest excellence, except through long and careful cultivation.

By Fulness of voice, I mean a grave and hollow volume, resembling the hoarseness of a common cold.

By Clearness, a freedom from aspiration, nasality, and vocal murmur.

By Strength, a satisfactory loudness or audibility.

By Smoothness, a freedom from all reedy or guttural harshness.
By a Sub-sonorous quality, its muffled resemblance to the resonance of certain musical instruments' (Rush, 1859: 151).

By 'vocal murmur', Rush meant 'an obscuring accompaniment of sound, as if the whole of the voice had not been made-up into articulation. It is not an unfrequent cause of indistinctness in speakers' (Rush, ibid.). He also had this to say about his use of the phrase 'guttural harshness':

'There is a harsh quality of voice called Guttural; produced by a vibratory current of the air, between the sides of the pharynx and the base of the tongue, when apparently brought into contact above the glottis. If then the term "voice from the throat" which has been one of the unmeaning or indefinite designations of vocal science, were applied to this guttural quality, it would precisely assign a locality to the mechanism' (Rush, 1859: 153).

It is very difficult to know what he meant, in the absence of an audible demonstration, particularly when many of the aspects of orotund voice that he specifies are only the absence of another quality of the voice (aspiration, nasality, vocal murmur, reedy or guttural harshness). There is also the difficulty that two of the factors he prescribes seem to be mutually contradictory, - 'a satisfactory loudness' and a 'muffled resemblance to the resonance of certain musical instruments'.

There is, however, one possibility of establishing what it was that Rush meant by labels such as 'Orotund voice'. I have quoted Rush at some length because of the remarkable influence he exercised on the field of elocution. Nearly all the terminologies for the description of voice quality published by American elocutionists for
a hundred years after the first appearance of his Philosophy of the Human Voice in 1827, up to and including Woolbert (1927), were based on that of Rush. It would be possible, in principle, to elucidate the meaning of some of Rush's terms, if a reasonably detailed continuity of influence could be established from Rush to writers in the twentieth century. It would be too sanguine to hope for more than the most approximate hint of Rush's conceptualization, but a continuity of teaching and influence can be shown. The following discussion is based in part on two interesting review articles about the history of American elocution in the nineteenth century by Gray (1943), and Robb (1954).

William Russell was the first editor of the American Journal of Education, from 1826. He was a prolific writer, and of his thirty books, sixteen were concerned with elocution. He was a contemporary of Rush, and strongly influenced by him, as well as by Austin's Chironomia (1806), (Robb, 1954: 187-188). With Goldsbury, Russell wrote an American School Reader (1844), in which the following labels were applied to voice qualities: 'harsh', 'smooth', 'aspirated' (whispery or breathy voice, J.L.), 'pectoral', 'guttural', 'oral', 'crotund', and 'pure tone'. They have no specific label for Rush's 'natural' quality, and don't mention his 'falsette' quality.

Russell founded the School of Practical Rhetoric and Oratory in 1884, with James Murdoch, a well-known actor who had been taught the principles underlying the Philosophy of the Human Voice by Rush himself. Robb (1954: 189) describes Murdoch as 'a leader in the elocutionary movement for fifty years'.

Murdoch and Russell collaborated to produce Orthophony, or
Vocal Culture in Elocution (1845) which was phenomenally successful, having eighty four impressions by 1882. They gave four basic qualities, - 'Whispering', 'Pure Tone', 'Orotund' and 'Aspirated'. In a 'cosmetic' approach, they list various potential faults in quality, such as 'the hollow and false pectoral murmur', 'the aspirate', 'the nasal' and 'the oral' (presumably meaning denasal voice, J.L.). Again, they don't include 'falsette'.

Hamill was a student of Murdoch, and in 1882 published his Science of Elocution.

In this he lists 'Pure Tone', 'Orotund', 'Aspirate', 'Pectoral', 'Guttural', 'Oral' and 'Nasal'. In the 1886 edition, he added 'Falsetto' (sic).

Fulton and Trueblood, (mentioned earlier as having been influenced by Austin's Chironomia (1806)), studied under both Murdoch and Hamill, and in 1893 published Practical Elocution. They listed 'Normal', 'Orotund', 'Oral', 'Aspirate', 'Guttural', 'Pectoral', 'Nasal', and 'Falsetto' qualities. One of the most interesting aspects of their work, from the point of view of this thesis, is that they chose to call one of their qualities 'Normal', and said that 'the Normal is the natural basis upon which all the other qualities rest, each of which is some modification of or variation from the Normal' (Fulton and Trueblood, 1893: 41).

Finally, Woolbert (1927) was the last in the continuous succession from Rush, studying under Trueblood. Woolbert is the link with American 'speech science' of the twentieth century, and begins to move towards an acoustic specification, very speculatively and imprecisely, of the qualities he suggests. 'Orotund' is the
'All-around Resonance', 'Pectoral' the 'Chest Resonance', and 'Guttural' the 'Throat Resonance'. He mentions the term 'Falsetto', but without any acoustic description.

It would be pleasant to be able to demonstrate that the evolution of the descriptive system which had its genetic origin in the work of Rush in 1827 had produced a competent and comprehensive structure by the time Woolbert made his contribution in 1927, a century later. Unfortunately, it has to be said that Woolbert's attribution of particular qualities to principal locations where 'resonance' takes place is not a major improvement on the 'metaphorical' system that Rush himself castigates in his discussion of 'that improved quality of the singing-voice, called by vocalists 'Pure Tone'', where he writes that:

'there are several terms used to describe the mechanical causes of its different characters and qualities. Among these, the causations implied by the phrases 'voce di testa', and 'voce di petto', or the voice from the head, and from the chest, must be considered as not yet manifest in physiology; and the notions conveyed by them must be hung up beside those metaphorical pictures, which with their characteristic dimness or misrepresentation have been in all ages, substituted for the unattainable delineations of the real processes of nature' (Rush, 1859: 153).

This is perhaps a little harsh on Woolbert; enough of Rush's system is visible in Woolbert's work, - and more particularly in that of Fulton and Trueblood (1893) - to give support to Hale's suggestion, noted earlier, that the value of Rush's work was in the physiological foundation he offered for the study of speech, (not least for the
Amongst more recent attempts to characterize voice quality, four treatments are of particular interest, those of Trager (1958) and Fairbanks (1960) in America, and Abercrombie (1967) and Crystal (1969) in Britain.

The article by Trager (1958) on paralanguage looks at first sight very relevant to the sort of approach adopted in this thesis, particularly when Trager says, early in the article,

'When language is used it takes place in the setting of an act of speech. Speech (‘talking’) results from activities which create a background of voice set. This background involves the idiosyncratic, including the specific physiology of the speakers and the total physical setting (....). Against this background there takes place three kinds of events employing the vocal apparatus: language(....); variegated other noises, not having the structure of language - vocalizations; and modifications of all the language and other noises. These modifications are the voice qualities. The vocalizations and voice qualities together are being called paralanguage (a term suggested by A.A. Hill (....))' (Trager, 1958; 3-4).

Trager's 'voice qualities' have nothing to do with the sense of 'voice quality' in this thesis. 'Voice qualities', for Trager, refer to the different phonetic parameters that can be used for paralinguistic communication. Some of them, such as 'pitch range' and 'tempo', would fall into Abercrombie's 'voice dynamics' (Abercrombie 1967: 95-102); but some of them are more than a little opaque. For example, the parameter of 'vocal lip control' Trager says 'ranges'
from heavy rasp or hoarseness through slight rasp to various degrees of openness' (Trager, 1958: 5). Although it should be acknowledged that Trager was exploring terrain that was relatively new to phonetics, progress in paralinguistic analysis is not accelerated by the use of auditory-imitation labels of this sort.

The concept which is of interest here, is 'voice set', which constitutes a 'background' against which other elements can be discerned. It is used, however, as a portmanteau label for an uncomfortably wide variety of phenomena. Trager writes that

'

'Voice set' as here delimited (....) involves the physiological and physical peculiarities resulting in the patterned identification of individuals as members of a societal group and as persons of a certain sex, age, state of health, body build, rhythm state, position in a group, mood, bodily condition, location. From the physical and physiological characteristics listed are derived cultural identifications of gender, age grade, health image, body image, rhythmic image, status, mode, condition, locale - and undoubtedly others' (Trager 1958: 4).

Trager's list is thus concerned with indexical information and the vocal features which act as vehicles for them. If we combine the analytic terms used in this thesis with Abercrombie's division of different indices into the three classes that (a) indicate membership of a group, (b) characterize the individual, and (c) reveal changing states of the speaker (Abercrombie, 1967: 7), then we can say that Trager's list includes features of accent as an index of type (a) ('identification of individuals as members of a societal group'); intrinsic features of voice quality, as an index of type (b) ('sex,
age, state of health, body build'); and paralinguistic features as an index of type (c) ('mood'). But it also contains some factors that I am unable to understand, such as 'rhythm state' and 'rhythmic image' (Trager, ibid.).

This is not to say that Trager's article is without relevance to this thesis. It is true that he conflates factors of voice quality and of paralinguistic communication in his concept of voice set, as noted above. But he does approach the concept of paralinguistic phenomena (in 'voice qualities') being identified against a background ('voice set') which has at least some elements of voice quality in it (intrinsic features). He also reaches towards the idea that paralinguistic phenomena and voice quality are procedurally (for the analyst) mutually defining, although he eventually shrinks from the task of carrying the idea through to an explicit definition:

'In analyzing and recording the paralinguistic phenomena to be described, it is necessary to state what the voice set back of them is. Such a statement is at least in part an abstraction going back from the actual observation of the paralanguage. But it is not the intention here to discuss the exact nature of voice set and its relation to paralanguage - this being a large separate task' (Trager, 1958: 4).

Fairbanks (1960) gives a more substantial contribution to the study of voice quality. In his Voice and Articulation Drillbook, he devotes one chapter to voice quality. He discusses four types of quality, - harshness, breathiness, hoarseness and nasality, saying that the first three are 'defects of tone generation' and that 'nasality' is a 'defect of transmission' (Fairbanks, 1960: 170). Fairbanks gives some acoustic and physiological descriptions of the
qualities, together with illustrative spectrograms of a single speaker producing all four qualities on sustained examples of the same phonetic vowel.

'Harshness' is defined as 'irregular, aperiodic noise in the vocal fold spectrum. A common cause is excessive laryngeal tension (....) Harsh speakers tend to initiate phonations abruptly, with obtrusive glottal attacks' (Fairbanks, 1960: 175-176).

'Breathy quality' is described as an inefficient laryngeal vibration:

'In the coordination of normal voice quality the vibrating vocal folds approximate in the midline once per cycle, closing the glottis and interrupting the airflow. In breathy quality the vocal folds vibrate, but the intermittent closure fails and airflow is continuous. The firmness of the basic glottal closure is insufficient for a given airflow (or the force of the airflow is excessive for a given closure). Breathy quality is almost invariably accompanied by limited vocal intensity (.. and ..) low pitch. Vocal attacks tend to be aspirate, in contrast to the glottal attacks of harshness' (Fairbanks, 1960: 179).

While harshness and breathiness are described in terms which emphasize their controllable, extrinsic nature, Fairbanks describes 'hoarse quality' in a way which suggests that he believed that hoarseness can have, superimposed on its basic intrinsic condition, a controllable extrinsic element as well. He writes:

'Universally familiar as a symptom of acute laryngitis, hoarseness combines the features of harshness and breathiness (....) The harsh element predominates in
some hoarse voices, the breathy element in others and the same kind of variations may be heard within a given voice (....) If persistent hoarseness is your problem (....), if you have not already done so, seek medical advice' (Fairbanks, 1960: 182-183).

He notes that a hoarse voice quality in which the breathy element predominates over the harsh element is sometimes referred to as a 'husky' voice (p.182). We shall see later, in the section in Chapter 3 on types of phonations, that 'husky' voices are probably better thought of as the combination of harshness and whispery voice, rather than breathy voice.

On nasality, Fairbanks states that

'Excessive nasality, or hypernasality, is one of the most common voice problems, but mild nasality is heard in many good voices. It may be a virtue, in fact, although the evidence is inconclusive. (....) Nasality is imparted to the vowel spectrum by lowering the velum and coupling the nasal cavity into the system (....) Constriction of the oral channel (above or behind the tongue, between teeth and lips, etc.) tends to increase the relative prominence of nasality in the spectrum. Excessive nasality usually accompanies such organic conditions as velar insufficiency, velar paralysis, cleft velum or palate, and anterior nasal obstruction' (Fairbanks, 1960: 172).

So here too, Fairbanks treats a particular voice quality as one which can arise from either intrinsic or extrinsic causes, with the remedial practice which is the objective of his book being applicable of course only to the extrinsic condition.

Abercrombie (1967: 89-95) gives a short account of a phonetic framework for the description of voice quality, which is by far the
most adequate so far published, even though he warns that much about voice quality is not understood, and that his account is 'necessarily no more than tentative'. This thesis takes Abercrombie's model as its chief point of departure, and owes much of its descriptive structure to his suggestions: the extent of the debt will become obvious in the course of the next chapter.

Abercrombie was quoted earlier as saying that

'the term "voice quality" refers to those characteristics which are present more or less all the time that a person is talking: it is a quasi-permanent quality running through all the sounds that issue from his mouth' (Abercrombie, 1967: 91).

He comments that the term 'voice quality' is possibly somewhat misleading, in that 'in this particular context "voice" has a much broader meaning than its technical phonetic sense', - it has a 'much more general meaning' than the quality of the sound resulting from the vibration of the vocal folds (Abercrombie, ibid.). The term 'voice quality' has been used in this wider sense by many authors (Abercrombie refers to it as a 'traditional term'(p.91)), and instead of adopting some such usage as 'personal quality' or inventing a new one, voice quality will be retained throughout this thesis in Abercrombie's meaning. To avoid confusion, the term used for the quality of the sound resulting from phonatory activity of the vocal folds will be phonatory quality.

Abercrombie's description of voice quality, although it was finally published in 1967, has been the subject of lectures from
1947 onwards. He distinguishes first between uncontrollable and potentially controllable components (the intrinsic and extrinsic features respectively of this thesis). The uncontrollable components are divided into anatomically-derived permanent features and more ephemeral aspects such as result from laryngitis. The controllable features are then attributed to the effect of 'various muscular tensions which are maintained by a speaker the whole time he is talking, and which keep certain of the organs of speech adjusted in a way that is not their relaxed position of rest. These adjustments give a kind of general "set" or configuration of the vocal tract, which inevitably affects the quality of sound which issues from it' (Abercrombie, 1967: 92-93).

A number of different types of muscular adjustments are then briefly discussed: firstly, adjustments of the lips and tongue, which are said to give, in effect, a continually maintained secondary articulation; then, adjustments of the velum and pharynx; next, adjustments of the larynx, giving different phonation types, citing Catford (1964); and finally, adjustments of the vertical position of the larynx (Abercrombie, 1967: 93).

The most recent contribution to the description of voice quality comes from Crystal (1969: 97-125). Crystal takes up the theme of voice quality acting as a perceptual background against which linguistic and paralinguistic vocal acts can be discerned, one of the strands in the meaning of Trager's 'voice set' discussed briefly above. Crystal describes voice quality in a number of different ways, without much analytic detail. Firstly, voice quality is

'that relatively permanent, non-institutionalized,
idiosyncratic, background voice quality which accompanies a person when he speaks and is the main source of our ability to recognize personal identity vocally. This (....) is normally completely uncontrolled (....), and we learn to discount a speaker's voice quality as contributing nothing to the meaning of language as soon as we have recognized it for what it is - something idiosyncratic to the speaker, as opposed to something shared by other speakers in the speech community' (Crystal, 1969: 100).

I would disagree with Crystal on the desirability of maintaining one particular aspect of this description of voice quality, and that is his position that voice quality must be non-institutionalized, and not shared with other speakers in the speech community. The velarization with continuing velic closure that Abercrombie (1967: 94-95) says characterizes the Liverpool accent (as a learnt quality, copying the adenoidal speech of other speakers in the community) is surely best treated as part of the voice quality of an individual speaker using that accent. Crystal discusses this point, briefly, in a later section (p.124), and decides to 'leave the question of group qualities open'. This is a point to which we shall return in a later chapter.

Another related and major disagreement with Crystal's description concerns his comment that voice quality is 'normally completely uncontrolled'. This either leads to the exclusion of any extrinsic component from voice quality, and so removes voice quality from strictly phonetic relevance, which is a position diametrically opposed to the whole conception of voice quality in this thesis; or 'normally' in Crystal's description is given far too wide a latitude of interpretation. Crystal presumably means that the control of voice quality is normally outside a speaker's awareness,
and in that circumstance beyond his volitional control, - but such an interpretation is not very clearly expressed by a phrase such as 'normally completely uncontrolled'.

Crystal elsewhere describes voice quality as

'the permanent "background" speaking characteristic of the voice against which conventional linguistic patterns are identified' (Crystal, 1969: 104),

which is unexceptionable from the point of view adopted in this thesis.

Another description Crystal gives of voice quality is that it is 'the permanent non-segmental idiocyncratic factor in a person's speech' (1969: 126). He offers a more extended definition when he concludes his chapter on voice quality with:

'Voice quality is thus the permanently present person-identifying "background" vocal effect in speech, constituted by the same set of acoustic-physiological parameters as constitute speech, but being distinguished from speech by a different set of parametric values which are never utilised for purposes of communication. Voice quality is readily distinguishable from linguistic contrasts on almost all occasions by its being (a) contextually random, i.e. the occurrence of voice quality does not correlate with non-physiologically determined categories (linguistic or non-linguistic), and consequently (b) wholly statistically predictable - a fact which would hold for the component parameters of voice quality as well as the total phenomenon' (Crystal, 1969: 124-125).

There are a number of points one could take issue with here, but the most important disagreement concerns Crystal's decision that voice quality is 'distinguished from speech by a different set of
parametric values which are never utilized for purposes of communication' (my emphasis). This formulation runs into difficulties immediately one starts to consider the relation between segmental articulation and voice quality. For example, in the case of a speaker characterized by nasalization as a factor in his voice quality, the 'acoustic-physiological' parametric values for that quality may well be exactly the same as those for a nasalized linguistic segment, say a vowel of a particular tongue position preceding a nasal consonant. In such a situation, one could perhaps say that voice quality is neutralized and that therefore voice quality has to be considered 'quasi-permanent' and not 'permanent'. We shall discuss in the next chapter a way of coping with this problem of neutralization, by distinguishing between segments which are susceptible to the influence of a particular voice quality component, versus those which are non-susceptible precisely because they do already share the same parametric values as the voice quality component. There are many possibilities of such sharing of parametric values by segmental articulation and voice quality, each of which would cause difficulty if Crystal's definition were to be maintained.

The final item of interest in this summary historical survey of the analysis of voice quality comes from a discussion not of voice quality as such, but from a study of the perception of vowel quality. Ladefoged (1962: 14-15) writes about the distinction between phonetic quality and what he calls 'personal quality', and suggests that the distinction is 'one of the basic assumptions of phonetics' (p.14). Voice quality, in the elaborated definition which will emerge from the descriptive account to be set out in the next chapter, would be included in Ladefoged's 'personal quality'. Of particular relevance
to this thesis is the support that Ladefoged gives to the contention noted briefly above in the Introduction, that considerations of personal quality, as well as being of interest to general phonetic theory, are crucial for the specifically linguistic study of speech, in that

'Until we know which of the acoustic features which are due to the shape of the vocal tract can be correlated with personal quality and which with phonetic quality, we cannot equate judgments about phonetic quality (which involve sociolinguistic criteria) with statements about the phonic data' (Ladefoged, 1962: 15).

It is in the posing of this question, about their mutual relationship, that the study of the substance of spoken language and the study of voice quality have to come together.
Part II

Chapter 3. The phonetic description of voice quality

Introduction

The essence of the phonetic description of voice quality to be put forward in this chapter is that the phenomena of voice quality, like those of segmental performance, can be effectively described in terms of contributory components.

As briefly discussed in the first chapter, the quality of every speaker's voice is the product of two sorts of features, intrinsic features and extrinsic features. Intrinsic features are those whose quality is directly due to the invariant anatomy and physiology of the speaker's vocal apparatus. Extrinsic features derive their quality from long-term muscular adjustments, or settings of the intrinsic vocal apparatus, which were once acquired perhaps by social imitation, or idiosyncratically, and are now habitual, and outside the speaker's normal awareness (Abercrombie, 1967; Honikman, 1964; Laver, 1964, 1967, 1968). Intrinsic and extrinsic features are the 'features' in the title of this thesis.

While intrinsic factors of basic anatomy are beyond the speaker's control, and the extrinsic muscular settings to a certain extent within his voluntary control, both these sources of voice quality can transmit indexical information, although of different sorts, as we shall see in the following chapter on voice quality and indexical information.

Comments will be made on intrinsic aspects of the basic anatomy and physiology of the vocal apparatus at the relevant points in the
text, but the major theoretical framework of the descriptive model will be based on the extrinsic, habitual, muscular settings that are available to all anatomically and physiologically normal speakers. This assumption of a standard vocal apparatus, as discussed briefly above in the chapter on principles of labelling, is the most fundamental of all assumptions in a universal, general phonetic theory. Consequently, here, as in many other areas in general phonetics, a primary concern is to demonstrate the general human potentials in the area of study, and to set up a descriptive structure applicable to those activities which human beings normally have in common. For this reason, although some discussion on indexical clues in voice quality to a number of pathological states will be given in the next chapter, the essence of the communicability and utility of the descriptive scheme set out in this chapter is that the muscular settings to be commented on can all be voluntarily imitated, while the inherently invariant features of anatomy cannot.

We have seen that there is an interaction between the intrinsic and extrinsic features in a speaker's voice quality, in that the intrinsic constraints 'set the limits to the range of potential performance within which (the speaker) must make his socially and idiosyncratically specific choices of extrinsic vocal behaviour'. Extreme extrinsic choices at the very limits of the intrinsic range are not often made, and most speakers select muscular settings well within the extreme possibilities allowed. However, that extreme adjustments in voice quality are possible, and sometimes made, is illustrated by Abercrombie (1967: 94), who writes that

'There are many professional mimics on stage, radio and
television who are able to give convincing imitations in which the performer's own voice quality characteristics are effectively submerged. The ventriloquist, also, has to have command of several voice qualities. The extreme of virtuosity, probably, was reached by a certain music-hall performer, a large middle-aged man, who had learnt to produce, completely convincingly, the voice quality of a seven-year old girl, showing that it is possible to compensate, by muscular adjustments, for extreme anatomical differences.

The question of what features in voice quality are being imitated in producing convincing imitations of another person's voice is a very difficult one to answer, but it seems not implausible that it is probably a matter of imitating the muscular settings of the larynx and supralaryngeal vocal tract, together with features of pitch-range and loudness-range, when trying to compensate for extreme anatomical differences. In this case cited by Abercrombie, clearly the large performer would not have been able to produce an acoustically perfect imitation of the voice quality produced by a vocal tract of considerably different dimensions and resonatory characteristics: which makes it all the more interesting, from the point of view of perceptual aspects of voice quality, to speculate as to what particular muscular adjustments the performer selected as giving the perceptually most convincing cues for the overall quality he was trying to imitate. They may have been features such as extreme elevation of the larynx, overall fronting and raising of the tongue, narrowing of the space between the lips, and high-pitched, soft, falsetto phonation. It seems unlikely, though, that even these features of adjustment would result in the performer's vocal tract taking on the actual dimensions of the seven-year old girl's, so that one would have to say that the performer was compensating 'by muscular adjustments, for extreme
anatomical differences only in the sense of reproducing a perceptually acceptable similarity, on criteria of acceptability yet to be established rather than truly duplicating the actual anatomical dimensions of the other speaker.

A. The history of the concept of an articulatory setting

1. The concept of an articulatory setting in early writings on the voice

While Honikman's term 'articulatory setting' is new (1964), the general concept of extrinsic settings is not. Settings in voice quality, (as overall, long-term tendencies underlying and colouring, as it were, the moment-to-moment segmental articulations), and settings in voice dynamics, (as constraining tendencies in longer-term aspects of features such as pitch-range, loudness-range and tempo), have been the subject of phonetic comment for a very long time. Writers on phonetics have made appeal to the concept of settings most often in the discussion of the ways in which general tendencies in pronunciation characterize different languages. A very early forerunner was Isadore of Seville, in the seventh century A.D., quoted by Fromkin and Rodman (1974: 230), to the effect that

'All the Oriental nations jam tongue and words together in the throat, like the Hebrews and Syrians. All the Mediterranean peoples push their enunciation forward to the palate, like the Greeks and the Asians. All the Occidentals break their words on their teeth, like the Italians and Spaniards ...

Wallis, in his *Tractatus de Loquela*, prefatory to his *Grammatica Linguae Anglicanae*, (1653), was the first in what might be thought to be the phonetic tradition in Britain to discuss the topic, when he
It is worth noting (...) that differences in pronunciation occur in various languages which are not attributable so much to the individual letters, as to the whole style of speech of the community. For instance, the English as it were push forward the whole of their pronunciation into the front part of the mouth, speaking with a wide mouth cavity, so that their sounds are more distinct. The Germans, on the other hand, retract their pronunciation to the back of the mouth and the bottom of the throat, so that they have a stronger and more forceful pronunciation. The French articulate all their sounds nearer the palate, and the mouth cavity is not so wide; so their pronunciation is less distinct, muffled as it were by an accompanying murmur. The Italians, and the Spaniards even more, speak with a slow tempo, the French, and the Scots equally, raise or sharpen the pitch of the last syllables of sentences and clauses, while the English lower or deepen it; this is a characteristic not of individual words but of the sentence taken as a continuous whole. I leave it to others to observe differences of this kind among other peoples, as the opportunity presents itself' (translated by Kemp, 1972: 209-211).

Kemp also points out that 'in describing the English pronunciation as at the front of the mouth Wallis is not referring to the place of articulation, but rather to the comparative lack of interference with the airstream at this point' (op. cit. p.62).

Kemp notes that 'Wallis's account of the speech habits of different communities is imitated by Wilkins (1668: 380-381), who also mentions variations in style of speech between individual members of one speech community' (Kemp, 1972: 62). This is certainly the case, but Wilkins
in his comments about settings, does more than just suggest that both languages and individuals can be characterized by them, as can be seen in the following quotation:

'Though each of the Letters have their distinct powers naturally fixed, yet that difference which in the various manner of Pronunciation, both somewhat alter the sound of them. And there are no two Nations in the world that do exactly agree in the same way of pronouncing any one Language (suppose the Latin) Amongst persons of the same Nation, some pronounce more fully and strongly, others more slightly, some more flatly, others more broadly, others more mincingly. (...) 'Tis obvious to anyone to observe, what great difference there will be in the same words, when spoken slowly and treatably, and when tumbled out in a rapid precipitate manner. And this is one kind of difference in the pronunciation of several Nations; The Spaniards and Italians pronouncing more slowly and Majestically, the French more volubly and hastily, the English in a middle way betwixt both. Another different mode of Pronunciation betwixt several Nations may be in regard of strength and distinctness of pronouncing, which will specially appear in those kind of Letters which do most abound in a Language. Some pronounce more deeply Guttural, as the Welsh, and the Eastern people, the Hebrews, and Arabians, etc. Others seem to thrust their words more forwards, towards the outward parts of the mouth, as the English; others more inward towards the palate, as the French; some speak with stronger collisions, and more vehement aspirations, as the Northern people generally, by reason of their abundance of spirits and inward heat; others more lightly and softly, as the Southern Nations, their internal spirits being weak, by reason of the outward heat. (...)'.

I cannot omit here the Censure which an ingenious person gives (R.C. of Anthony in Cambridge, in margin)
concerning the difference in many of our European Languages, in respect of their pronunciation. The Italian (saith he) is in pronunciation pleasant, but without Sinews, as a still flowing water; the French delicate, but inward and nice, like a woman that dares scarce open her mouth, for fear of marring her Countenance. The Spanish, Majestical, but withal somewhat terrible and fulsom, by the too much affectation of the Letter O. The Dutch manly, but withal harsh and quarrelsom. Whereas our English (saith he) hath what is comely and Euphonical in each of these, without any of their Inconveniences. 'Tis usual for men to be most favourable towards the Language unto which they have been most accustomed. 'Tis likely that Forreigners may be as apt to complain of several Defects in our Language as we of theirs' (Wilkins, 1668: 380-381).

What it is important to note here, for the study of the contribution of articulatory settings to voice quality, is Wilkins' originality in being the first writer on phonetics, as far as I am aware, to approach the idea that the setting which characterizes a speech-community can be seen as the product of the nature and distributional frequency of the segmental stock of their phonology: in other words, the characterizing setting represents the long-term highest-common-factor of all the articulations of the speech-community, weighted for frequency of occurrence. The relevant part of the quotation from Wilkins is where he suggests that 'Another (....) mode of pronunciation betwixt several Nations may be in regard of strength and distinctness of pronouncing, which will specially appear in those kind of Letters which do most abound in a Language ....'. That Wilkins was thinking chiefly in articulatory terms here is supported by the fact that it is only at this point in his discussion that he deals in unequivocally articulatory descriptions of the shape of the vocal tract ('thrust
their words more forwards, towards the outward parts of the mouth ... more inward towards the palate ...'), where elsewhere his descriptions rather concern voice dynamic aspects such as tempo, or are couched in terms of auditorily-based imitation labels.

Wilkins' 'physiognomic' thoughts about a connection between speech and internal spirits and internal and external heat are now amusing, but one can still admire his laconic and firm assertion of the cultural relativity of phonaesthetic notions such as that English is 'comely and Euphonical', expressed by 'R.C. of Anthony in Cambridge' (who was Richard Carew (Abercrombie, personal communication)).

Holder (1669) also discusses, briefly, the various ways in which languages, and individual speakers of the same language, can be differentiated. He deals first with lexical, phonological and phonetic differentiation:

'There are other differences of Sound in Speaking, by which the Tones of several Nations, and oft of several persons in the same Nation, is rendred distinct, which are partly to be referred to their Alphabets, and partly to their Words and manner of Pronunciation, and Accent' (Holder, 1669: 74-75).

He later discusses differentiation by means of some extrinsic settings and by intrinsic features:

'Many more Observations of these kinds might easily be made, and are to be found in different Languages, all over the habitable world. And in general, the Freedom or Apertness and vigour of pronouncing (as is particularly observed in the Bocca Romana) and giving somewhat more of Aspiration; and on the other side, the closeness and Nufling, and (as I may say) Laziness of Speaking (which
varieties are found in several Nations comparatively, and by the different natural shapes of the Mouth, and several conformations of the Organs of speech in those of the same Language) render the sound of their Speech considerably different, though they all should use the same Alphabet' (Holder, 1669: 78-79).

Cooper (1687: 10-11), in his English translation of a passage from his Grammatica Linguæ Anglicanae (1685: 18), gives clearer expression to an idea already sketched in by Wallis, Wilkins and Holder, and which has to be considered in any phonetic study of voice quality:

'Each of the vowels may be pronounced in a slender, middle, or gross sound; as the Organs are greater or less; the Appulse of the Active Instruments, Tongue, Lips, to the passive, Teeth, Palate, stronger or weaker. Or as the passage of the Breath from the Lungs thorow the Wind-pipe, Larynx or Mouth is longer or shorter, broader or narrower, emitted with greater or less force. Whereby the same notion of the breath, made by the same specifical Organs, causing the same specifical sound, may be spoken in divers Tones: as appears from the speech of persons of both Sexes, young and old, healthful and sickly, and men of several countries, some speak very broad and openly, others fine and inwardly, and some others in a mean. The German and Welsh speak roughly and difficulty (sic) in the Throat; so the Hebrew Tongue abounds in Gutturals. The French quick and hasty, full of Vowels, as the Greek, the English speak their syllables and words distinctly and outwardly from the Lips'.

The crucial section is 'Whereby the same motion of the breath, made by
the same specifical Organs, causing the same specifical sound, may be spoken in divers Tones'. The implication is that two sounds may be judged to be the same, (phonetically the same) even though they may have been produced by speakers who differ in both intrinsic and extrinsic aspects. That is, phonetic quality is independent of both intrinsic and extrinsic aspects of voice quality. This implication seems also to underlie comments by Wallis ('differences in pronunciation occur in various languages which are not attributable so much to the individual letters, as to the whole style of speech of the community'); by Wilkins ('Though each of the Letters have their distinct powers naturally fixed, yet that difference which is in the various manner of Pronunciation, doth somewhat alter them'); and by Holder ((intrinsic features and extrinsic settings) 'render the sound of their Speech considerably different, though they all should use the same Alphabet').

The independence of phonetic quality from intrinsic contributions to phonic quality is clear-cut; but we shall see in the chapter on semiotic aspects of speech production that the question of whether it is profitable to exclude the extrinsic settings of voice quality from contributing to linguistically-relevant phonetic quality is problematic. What seems unquestionable is that phonetic quality is not synonymous with phonic quality, the perceptual correlate of the totality of the acoustic events of speech. Just how much of an abstraction phonetic quality is from the phonic material is debatable, but writers who reject the abstract nature of phonetic quality, in suggesting that phonetic quality is 'raw data', are surely misguided.
Aickin (1693) offers a fairly free translation of Wallis's comments on language-characterizing settings, while Greenwood (1711) follows Wallis's original rather more closely.

Bayly (1758), as we have seen in the preceding chapter, mentions that

'The voice is (....) often hurt by (....) shutting the teeth, and confining the tone within the mouth instead of opening the teeth and lips properly so as to bring it out with fulness and rotundity' (Bayly, 1758: Part 3, p.181).

Bayly's approval of articulatory settings which give a voice 'rotundity' is echoed by Herries (1773),

'Endeavour to acquire a ROUNDEESS and openness in your speech. Let there be nothing shrill or squeaking in it. Some speakers pronounce with great distinctness, and yet there is a smallness and a puerility in their tone, which is very unsuitable to the grandeur and dignity of publick Eloquence. The Greeks who carried the fine arts to greater perfection than any nation whatsoever, were remarkable for this OSROTONUM, this full and flowing articulation. Roundness of voice may be acquired by depressing the tongue and jaw, and enlarging the cavity of the mouth in the pronunciation of vowels' (Herries, 1773: 117).

Webster (1789), like Bayly and Herries, remarks in the passage quoted in the preceding chapter on the effect on speech of jaw settings, in discussing the 'drawling nasal manner of speaking in New England' (p.106). He refers to

'.... local peculiarities in pronunciation, which prevail
among the country people in New England, and which, to foreigners, are the object of ridicule. The great error in their manner of speaking proceeds immediately from not opening the mouth sufficiently. Hence the words are drawled out in a careless lazy manner, or the sound finds a passage thro the nose.

Nothing can be so disagreeable as that drawling, whining cant that distinguishes a certain class of people; and too much pains cannot be taken to reform the practice. Great efforts should be made by teachers of schools, to make their pupils open the teeth, and give a full clear sound to every syllable. The beauty of speaking consists in giving each letter and syllable its due proportion of sound, with a prompt articulation' (Webster, 1789: 108-109)

2. The concept of an articulatory setting in the phonetic literature of the 19th and 20th centuries

We have seen that the general concept of an overall articulatory setting was quite clearly established over three hundred years ago by Wallis (1653); the general concept had to wait for a specific name until the second half of the nineteenth century, however, when phoneticians such as Sweet, Sievers, Storm, Jespersen and Vidtor became interested in the topic. Kelz (1971) gives a good account of the widespread adoption of Franke's term 'Artikulationsbasis' (coined earlier but published posthumously in an article in 1889), replacing such terms as Sievers' 'Operationsbasis', Storm's 'Mundlage' and Sweet's 'organic basis'. Discussion here will be confined mainly to Sweet's contributions, because of his importance in the more general field of voice quality.

Sweet first wrote about this area in 1877, in a discussion of
'varieties of voice quality', which he said were 'mostly individual or national peculiarities' (Sweet, 1877: 98). He gave the following examples:

'Narrowing of the upper glottis (...) gives an effect of strangulation. It is common among Scotchmen, and combined with high key gives the pronunciation of the Saxon Germans its peculiarly harsh character.

Partial closure of the mouth is a common English peculiarity. It has a tendency to labialize back vowels, and even when there is not actual labialization it gives the vowels generally a muffled sound, so that \( \alpha \), for instance, is not easily distinguished from \( \mathcal{O} \). It also tends to make the general speech nasal, for the breath being impeded in its passage through the mouth, naturally seeks another through the nose. Germans sometimes say of the English, with humorous exaggeration, that they speak, not with their mouth like other people, but with their nose and throat'.

In 1884, Vöctor published his *Elemente der Phonetik* (translated into English by Ripman in 1899), in which the term 'artikulationsbasis' first appeared. Sweet used this term in 1885, in his *Elementarbuch des gesprochenen Englisch*, the German version of his *A Primer of Spoken English*, which was not published in English until 1890. In the English version, Sweet rendered 'artikulationsbasis' as 'organic basis', saying that

'The general character of English speech depends on the following peculiarities of its organic basis: The tongue is broadened and flattened, and drawn back from the teeth (which it scarcely ever touches), and the forepart of it is hollowed out, which gives a dull sound, especially noticeable in \( \mathcal{U} \).

Rounded (labial) sounds, such as \( \mathcal{U}, \mathcal{W} \), in 'who'.
are formed without any pouting of the lips.

In unrounded vowels, such as \( a, i \), the lips have a passive, neutral position. In the formation of front vowels such as \( i, e \), there is no 'chinking' or spreading out of the corners of the mouth, by which in other languages their sound is made clearer' (1890a: 4).

In his *A Primer of Phonetics* (1890b) Sweet expands the comments he made on voice quality in his *Handbook* (1877), to suggest that, in the terms of this thesis, while intrinsic 'peculiarities are inseparable from the individual', voice quality modifications 'which are the result of controllable organic positions (...) may - and often do - characterize the speech of whole communities' (1890b: 73).

He then goes on to give a more extended description of his idea of an 'organic basis' than in his *Primer of Spoken English*, explicitly equating 'organic basis' with 'basis of articulation':

'Eve language has certain general tendencies which control its organic movements and positions, constituting its organic basis or basis of articulation. A knowledge of the organic basis is a great help in acquiring the pronunciation of a language.

In English we flatten and lower the tongue, hollow the front of it, and draw it back from the teeth, keeping the lips as much as possible in a neutral position. The flattening of the tongue widens our vowels, its lowering makes the second element of our diphthongs indistinct, front-hollowing gives a dull resonance which is particularly noticeable in our \( i \), its retraction is unfavourable to the formation of teeth-sounds, and favours the development of mixed vowels, while the neutrality of the lips eliminates front-round vowels. Our neutral tongue-position is the low-mixed or mid-mixed one of the vowels in "further" (...).
In French everything is reversed. The tongue is arched and raised and advanced as much as possible, and the lips articulate with energy. French therefore favours narrowness both in vowels and in consonants, its point-consonants tend to dentality, and, compared with the English ones, have a front-modified character, which is most noticeable in the \( \rightleftharpoons \), while the rounded vowels are very distinct.

The German basis is a compromise between the English and the French, standard North German approaching more to the French. No language, however, carries out the tendencies of its basis with perfect consistency.

Thus in English we have the point-teeth \( \rightleftharpoons \); and mixed vowels occur in French and German, etc." (Sweet, 1890b: 74-75).

This description of 'organic basis' is then repeated in its essentials in Sweet's *The Sounds of English* (1908: 57-58), and in his 'Phonetics', (*Encyclopaedia Britannica*, 11th edition, 1911: 466).

The term that is mostly used nowadays to refer to the concept of an 'organic basis' is 'basis of articulation'. Heffner (1950), for example, uses this term in discussing the 'Indifferenzlage' (a term also used by Franke), or 'habitual attitude' of the vocal organs. The ambiguity of phrases like 'habitual attitude' has been the source of a certain amount of confusion, and it is therefore worth looking at Heffner's treatment in detail. He writes that:

'It is clear that there are certain general peculiarities of utterance which characterize the speech movements of whole groups of speakers. One of the most important of these is sometimes spoken of as the general basis of articulation. The matter deserves much more careful
treatment than it has yet received. It has, however, been observed that in some dialects or languages the position of the tongue from which its articulatory movements start and to which it tends to return is relatively low and retracted, with the tongue surface quite broad. All the movements of the tongue are made to harmonize with this habitual attitude, or "Indifferenzlage" of the organ, and the vowels particularly are then marked by dullness rather than by brightness of timbre. Some groups of speakers join with this generally low level tongue position a tendency to allow the lips to remain largely passive. In other regions the whole level of the tongue in its speech movements is higher and more in the front of the mouth. The tongue is more tense and the movements are quicker. Often in such cases the lips are notably active. French is a language which is usually spoken from a high and tense forward basis of articulation. German is intermediate between French and English in this respect, for English is spoken from a comparatively low and relaxed basis of articulation. On the whole, American English is even less energetically enunciated than British English. These are broad generalizations to which there are many individual exceptions. They are, moreover, supported only by the observations of men like Storm, Sievers and Jespersen. No method of measurement has been devised which would permit the mathematical description of a basis of articulation' (Heffner, 1950: 98-99).

This is unkind to scholars such as Storm, Sievers and Jespersen, and it also shifts analytic attention to the configuration of the vocal organs before and after speech, from the 'long-term highest-common-factor of all the articulations of the speech-community, with individual articulations weighted for frequency of occurrence', which I suggested earlier was the tenor of Wilkins' (1668) description of
settings in different languages, and which seems to be the general sense adopted by phoneticians in Britain who have written on the topic since that time. I don't believe that this is what Heffner meant to imply, but there is considerable ambiguity in some crucial phrases, such as 'from which its articulatory movements start and to which it tends to return'. Does this mean 'start at the beginning of continuous speech and return at the end of speaking', or 'start at the beginning of continuous speech and tend to return repeatedly during speech, and return finally at the end of speaking'? Presumably something more like the latter than the former, on the evidence of the later phrase 'All the movements of the tongue are made to harmonize with this habitual attitude', although how exactly the harmonizing occurs is not very clear from the context.

That this sort of formulation reflects an underlying ambiguity can be seen in entries in linguistic glossaries such as Pei and Gaynor (1954), who say that the term 'basis of articulation' means 'the neutral position of the organs of speech and their various parts which is peculiar to or characteristic of a given speech-community or of the native speakers of a given language', which hedges its bets about what 'neutral' might mean, and Pei (1966), who calls 'basis of articulation' 'the over-all neutral position of the speech organs and their various parts when not speaking which is characteristic of a speech community of the native speakers of a language', which could hardly help to characterize any aspect of the phonetic performance of the group except the moments of starting and stopping speaking.
Chomsky and Halle (1968: 295) quote a definition by Marouzeau which is rather difficult to penetrate. They write

'phonetic transcriptions consistently disregard many overt physical properties of speech. Among these are phonetic effects that are not locatable in particular segments but rather extend over entire utterances, such as the voice pitch and quality of the speaker and also such socially determined aspects of speech as the normal rate of utterance and what has been called the "articulation base":

"the system of characteristic articulatory movements of a given language that confer upon it its general phonetic aspect; in French the mobility of the lips and forward position of the tongue" (Marouzeau, 1943: 38).

Leaving aside the syntactic ambiguity of the term 'general', the definition is either too vague to be useful, or its two halves seem incompatible. The first half seems to include all aspects of typical articulation, rather than solely the 'highest-common-factor' of the different articulations of the language, - the second half then exemplifies long-term articulatory tendencies.

The best extended discussion of the concept of a setting, and from whom the term 'setting' is borrowed in this thesis, is by Honikman (1964). She writes:

'Articulatory setting does not imply simply the particular articulations of the individual speech sounds of a language, but is rather the nexus of these isolated facts and their assemblage, based on their common, rather than their distinguishing, components. The isolated articulations are mutually related parts of the whole
utterance; they are clues, as it were, to the articulatory plan of the whole; the conception of articulatory setting seeks to incorporate the clues or to see them as incorporated in the whole. Thus an articulatory setting is the gross oral posture and mechanics, both external and internal, requisite as a framework for the comfortable, economic, and fluent merging and integrating of the isolated sounds into that harmonious, cognizable whole which constitutes the established pronunciation of a language' (Honikman, 1964: 73).

Honikman distinguishes between 'external' and 'internal' articulatory settings. External settings are visible ones, mostly concerning the lips and the jaw, while the internal settings, 'intimately bound up with, and to a large extent governing' them, are factors to do with 'the over-all positioning of the internal mobile organs of the mouth for natural utterance' (Honikman, 1964: 75). External settings are thus basically those of the lips and the jaw, and the internal ones are those of the tongue, of general muscle tension, and of pressure of tongue-contact with the palate (p.79).

Honikman's approach to the concept of a setting characterizing the pronunciation of a language emphasizes two particular aspects, echoing the comments of Wilkins (1668), reported above, to do firstly with the highest-common-factor in the various segmental articulations of a language, and secondly with the need to give a statistical weighting to the contribution of individual segments to a setting on the basis of their frequency of occurrence in the spoken language. The first aspect is seen in the first sentence of her account of an articulatory setting, cited above:

'Articulatory setting does not imply simply the particular
articulations of the individual speech sounds of a language, but is rather the nexus of these isolated facts and their assemblage, based on their common, rather than their distinguishing, components' (p. 73) (my emphasis)

The second aspect is stated just as explicitly:

'The internal articulatory setting of a language is determined, to a great extent, by the most frequently occurring sounds and sound combinations in that language. Since it is the articulation for consonants that interrupts or impedes the free flow of the air stream through the mouth, the setting required for the most frequent consonants has been an important bearing on the articulatory setting as a whole - no less important than that required for the most frequent vowels' (p. 76) (my emphasis).

Honikman gives a very graphic account of the tongue-setting for English (Received Pronunciation), which shows the articulatory relationship between a long-term setting and the momentary articulations of segments, and which is therefore worth extended quotation:

'Almost throughout English, the tongue is tethered laterally to the roof of the mouth by allowing the sides to rest along the inner surface of the upper lateral gums and teeth; the lateral rims of the tongue very seldom entirely leave this part of the roof of the mouth, whereas the tip constantly (or some other part of the dorsum, occasionally) moves up and down, periodically touching the central part of the roof of the mouth, but generally not very long at a time, before it comes away. Thus, one might regard the tethered part - in this case, the lateral contact - as the anchorage, and the untethered part as the free or operative part of the tongue-setting.

By anchoring the tongue we, naturally, lessen its freedom of movement. Therefore it is important to note
the extent of the anchorage, for this prescribes the range of play of the free part as well as of the tongue as a whole. The forward limit(s) of tethering might well serve as points of reference in describing the anchorage.

Thus, the alveolar consonants of English (.....) require lateral anchorage as far forward as the upper posterior pre-molars (but never beyond the anterior pre-molars). These teeth on either side of the roof of the mouth serve, as it were, as forward mooring-posts for the tongue, allowing the part of the dorsum between them to operate as a hinge which enables the tip and blade to swing comfortably up and down, towards, to, and away from the alveolar ridge, but preventing the blade and tip from ranging much further forward without strain. This anterior lateral contact is released for a further back consonant or open or back vowel, and very slightly extended forward for dental sounds.

Since this anchorage is not tensely held, but is rather a pliable cushioning of the tongue-rim, adjustments to it such as lowering, retracting, and advancing are comfortably and smoothly made when required' (Honikman, 1964: 76-77).

This is exactly how the relationship between a setting and the segments affected by it will be assumed in this thesis. One slight interpretive difference is that Honikman often writes of the articulatory setting for a language, implicitly grouping all contributory sub-settings of different parts of the vocal tract together. This usage will be avoided here, and each separable contributory setting will be given equal analytic status, because of the explicit theme, in this model, of describing voice quality in terms of relatively independent physiological components.

To conclude this brief account of the development of the concept
of a setting as a factor in voice quality, we come again to
Abercrombie (1967). It will be recalled from the discussion of the
history of voice quality analysis in the previous chapter that
Abercrombie (1967: 93) uses the idea of certain muscular tensions,
or adjustments, held throughout speech, to explain the phonetic
basis of voice quality. The notion of a long-held muscular
adjustment is essentially the same as Honikman's 'setting', and
Abercrombie cites Honikman's article for 'a very full and clear
account of adjustments of this sort' (Abercrombie, 1967: 171n).
No further elaboration is needed here, therefore, for the idea of
'muscular adjustments', and the reader is referred to the latter
part of the previous chapter for an account of the types of muscular
adjustments applied in different parts of the vocal tract described
by Abercrombie.

B. A phonetic description of voice quality

We have seen that the concept of a setting is fairly abstract;
in particular, its 'statistical' aspect has been emphasized in the
erlier discussion, concerned as it is with characteristic common
denominators underlying the rapidly fluctuating and spasmodic
muscular activities which constitute the speech of a given individual
or group of individuals. In taking into account these long-term
aspects of muscular activity in speech, the analytic process may
here parallel the perceptual process applied to voice quality by the
human listener. When meeting a speaker for the first time, a
listener seems likely to perceive his voice quality by some process
of abstraction of relevant features gathered over a period of time
from spasmodic and ephemeral clues in the moment-to-moment fluctuations
of segmental and suprasegmental articulations. The most obvious
eexample of this is that perceptual clues to voice quality lie
chiefly, though not exclusively, in the intermittently occurring
voiced segments, where such features as the acoustic resonatory
characteristics of the vocal tract are most easily audible.

This comment about the intermittent, spasmodic nature of clues
in segmental performance to a speaker's voice quality leads on to an
important reservation about the relationship between segments and
settings. That is, no articulatory setting normally applies to
every single segment a speaker utters. The applicability of a given
setting is neutralized by any segment which shares articulatory
parameters with that setting. The segment may over-ride the
parametric values of a setting, as in the case of a velar stop in
the speech of someone generally characterized by velarization, or
a nasal stop in the speech of a speaker with a nasal voice quality.
In these examples, the contribution of the setting to segmental
articulation is redundant. Alternatively, the performance of a
particular segment may momentarily reverse the value of a parameter
normally exploited by a setting, as in oral stops performed by a
speaker with a nasal voice quality.

In order to be able to discuss the relationship between segments
and settings more clearly, I shall draw a distinction between segments,
which are susceptible to the influence of a given setting, and those
which are not susceptible, because of their pre-emptive articulatory
requirements. The terms susceptible and non-susceptible will be
given technical status and used only in these senses, when applied to
segments, from here on.
The assumption is made that each setting discussed will apply to all segments susceptible to its influence in the speech of the speakers concerned. It is partly the intermittency of occurrence in the stream of speech of these susceptible segments that constrains definitions of voice quality such as Abercrombie's (1967: 91) to qualify the permanence of voice quality by mitigators such as 'quasi-permanent quality', and 'characteristics which are present more or less all the time that a person is talking'.

One final aspect of the idea that the perception of voice quality is mediated by clues in segmental articulation is that, as noted earlier, for the linguistic phonetician trying to analyse the phonology and paraphonology of a language from the speech of a native informant, the allocations of phonic features to voice quality and to phonology and paraphonology have to go hand in hand as mutually entailing hypotheses. We shall return to this point in the chapter on the semiotic analysis of speech, to argue that linguistically-relevant aspects of speech can only be assessed against the background of speaker-characterizing voice quality, and vice versa.

In the description of extrinsic settings, comment is directed principally towards articulatory, physiological specifications and then, where possible, towards acoustic details, trying always to follow the fundamental phonetic assumption of an anatomically and physiologically standard vocal apparatus. Not surprisingly, it has to be acknowledged that not all writers agree on the physiology of some of the mechanisms; and that too little is yet known about certain areas to allow a detailed and comprehensive statement. This is especially true of the mechanisms controlling the movements of the
tongue, and of the vocal cords in some phonation types, and also of the fine details of the mechanisms of lip-control. However, a broad outline is feasible, and the physiological descriptions offered here will wherever possible include a statement of at least the identity of the muscles which are the prime movers or protagonists in achieving a particular setting. Less often, comment will also be offered on the identity of the antagonist, bracing and neutral muscles that respectively resist, brace against, and are passively affected by, the action of the prime movers (Van Riper and Irwin, 1958: 329).

The acoustic specifications of the effects of the various settings will be in terms appropriate for producing an auditorily similar effect on an electronic speech synthesiser. An account will be given at the end of this chapter of an attempt to synthesize some of the voice qualities discussed, on PAT, the resonance analogue speech synthesiser designed by Walter Lawrence (Anthony and Lawrence, 1962; Lawrence, 1953). The great value of speech synthesis to phonetics is that a machine such as PAT allows the exploration of a very wide variety of (i.e. the acoustic correlates of) programmed articulations AS IF pronounced by a speaker with a standard vocal tract length of standard resonatory characteristics, - precisely the same qualification demanded by phonetic theory for its descriptive schemata. In the normal experimental situation, using speech synthesis, the experimental input to PAT is a parametric program corresponding to some sequence of segmental articulations. Each parameter is constrained to fluctuate between standard limiting values, with the result that the audible output is a stretch of 'speech' as if pronounced by an individual maintaining an unvarying voice quality throughout, - indeed PAT's voice quality is by now well enough known
to be immediately recognizable to most researchers using speech
synthesis. It is possible, however, to reverse this normal
situation, and instead of using different segmental programs
produced with a standard voice quality, use one single segmental
program a number of times, each time with changed limiting values
of the ranges of one or more acoustic parameters. This has the
auditory effect of a single speaker saying the same sentence
repeatedly, but changing the extrinsic settings of his voice quality
for each repetition. If the limiting values of the acoustic
parameters are changed more substantially, the effect is of a
change of intrinsic features, making PAT sound like a different
speaker. The experiment reported later (Laver, 1964, 1967) is an
attempt to duplicate the acoustic correlates of a variety of
extrinsic settings as if spoken by the same individual. It has
been shown (Fant, 1968) that the acoustic parameters used by
synthesisers such as PAT allow an acoustic specification of segmental
articulations, which is usually phonetically adequate in terms of
intelligibility; since both segmental articulation and extrinsic
settings in voice quality depend upon and exploit many of the same
sorts of vocal features, it seems reasonable to suppose that a
specification of extrinsic settings in terms of PAT's acoustic
parameters may have the same sort of relevance to the phonetic
study of voice quality.

The settings will therefore be described in terms of the
following acoustic factors: the frequency and intensity range of
the fundamental; the frequency, amplitude and bandwidth character-
istics of the first three formants, and sometimes of the fourth;
the spectral characteristics of noise components; and sometimes
details of the shaping of the larynx pulse.
Finally, it should be noted that it is a matter of convenience of description and presentation that the different settings are being treated as analytically separate, as if they were mutually independent. In fact, as we shall see in the discussion of various settings, there is often a quite considerable influence exercised by different settings on each other. Changes of setting in a given location can cause simultaneous changes in settings in many other locations. This is partly because of the anatomical interpenetration of the physiological systems involved (Kenyon, 1927), and partly because the different parts of the vocal tract and the larynx act acoustically as a single, unified complex (Curry, 1940; Flanagan and Meinhardt, 1964; Gray and Wise, 1959; Miller, 1964; Travis, Bender and Buchanan, 1934; Trenschel, 1969). It is important to note, though, that although there are definite interactive acoustic effects between the vocal tract and the larynx, it has not yet been possible to show that the variation of the phonatory source spectrum with different configurations of the supralaryngeal tract is systematic, except within the performance of individual speakers (Carr and Trill, 1964; Kártorj, 1965b).

With especial regard to the resonatory system of the supralaryngeal vocal tract, comment on acoustic correlates of given settings will sometimes be framed as if changes in a particular section of the tract could result in changes of the characteristics of a single formant. The over-simplification of this approach to the theory of formant-cavity affiliation is emphasized by Chiba and Kajiyama (1958: 93), Fant (1960: 113) and Stevens and House (1961: 307). It is acknowledged that no such isolated acoustic changes are possible, and when isolating comment is made here it is in the interests of
simplicity and brevity of exposition, in sketching the outline of a potentially very large field.

We come now to the details of the proposed scheme for the phonetic description of voice quality. The scheme has its foundations in the suggestions of Abercrombie (1967), Catford (1964) and Honikman (1964), and puts forward three principal groupings of extrinsic settings. Firstly, configurational settings of the supralaryngeal vocal tract; secondly, settings of the phonatory mechanism of the larynx; and thirdly, settings of overall degrees of muscular tension exerted throughout the whole vocal system.

1. Configurational settings of the supralaryngeal vocal tract

The three sub-groups of settings of the supralaryngeal vocal tract, the longitudinal, latitudinal and velopharyngeal settings, refer to different sorts of modifications of the long-term shape of the tract and its acoustic characteristics. It should be noted that only settings capable of being produced by all speakers with normal vocal systems are included in the descriptive phonetic scheme, in accordance with the definition offered earlier of the term 'extrinsic'. A discussion of 'modifications' presupposes the definition of some standard reference, or norm, from which deviation can be measured.

It is essential to a general phonetic approach to settings in voice quality that the 'normal' vocal tract setting which is taken as the datum from which deviation is assessed should have no linguistic, culturally-relative prejudice of 'normality'. To avoid this, one solution is not to define some linguistically-relevant configuration of the tract as the norm, but rather to define the norm in terms...
relevant to the standard vocal tract itself. This can be done by specifying that the neutral configurational setting of the vocal tract is that constituted by the tract when it most nearly assumes the shape of a tube of uniform cross-section throughout its length. This is to adopt, with slight rewording, Fant's formulation of the 'ideal neutral vowel', which he says is defined as 'the sound produced from an idealized vocal tract of constant cross-sectional area and an effective length of approximately 17.6 cm' (Fant, 1960: 292). In this condition, the centre-frequency values for the ascending formants will be the odd multiples of the centre-frequency of the lowest formant. If we take as a standard the anatomy of an adult male with a vocal tract length of 17 cm. (as is represented by the acoustic make-up of PAT), then the actual frequencies of the formants will be 1000 cycles apart, with the first formant at 500, the second at 1500, the third at 2500, and so on upwards (Stevens and House, 1961: 308).

The auditory quality of the vowel sound which is produced by this configuration is very close to that of the so-called 'neutral' vowel, schwa - (ə).

The following assumptions will be made about the neutral condition of the vocal tract and larynx: the highest point of the regularly-curved tongue is central in the vowel area; the lips are not protruded, and are neither rounded nor spread; there is velic closure; the larynx is neither raised nor lowered; the type of phonation of the true vocal folds is periodic, efficient and without audible friction, giving 'modal' voice; and the overall degree of muscular tension in the vocal system is neither very tense nor very
lax. Nearly all these assumptions will have to be modified in later discussion, but they will hold for the present.

It is important not to equate this neutral setting of the tract, as many writers have done (including Sweet, 1877: 13), with the habitual posture of the tongue at rest during silence. This may perhaps be the case with some speakers, but it is certainly not true of all. Many speakers, when not speaking, hold the tongue against or near to the palate, even maintaining a slight negative velaric pressure behind an alveolar or palato-alveolar closure. In such speakers, speech is often prefaced by a weak alveolar or palato-alveolar click, as the forward contact is released.

a) **Longitudinal settings**

Modifications of the longitudinal axis of the vocal tract can result from at least three different types of displacement of vocal organs from their neutral position. The first two involve vertical displacements of the larynx and its associated supporting framework, upwards or downwards from its neutral location, giving **raised larynx voice** and **lowered larynx voice** respectively. The third type of modification concerns the forward protrusion of the lips.

i) **raised larynx voice**

The key to vertical movements of the larynx is the hyoid bone, from which one could say the larynx is suspended. The hyoid is a 'U'-shaped bone with the open end pointing backwards, itself nearly horizontally suspended above the larynx by a triple sling system of muscles. The first sling pulls the hyoid upwards and backwards towards the skull and the middle pharynx; the second sling pulls the
The hyoid bone is unique in being the only bone in the body which is not articulated with any other bone, and its muscular suspension from the larynx, pharynx, tongue and jaw, with the muscular tensions of the different sling systems having to be appropriately balanced for the accurate production of almost every single act of the vocal apparatus, makes the hyoid complex the prime example of mutually-influencing interaction of different muscular systems in speech.

Figure 1 is a schematic diagram of the three muscular hyoid slings. There are two chief possibilities of using these slings to raise the larynx from its neutral position. The first is to immobilize the hyoid and pull the larynx up towards it. This is done by using mainly the hyoglossus, geniohyoid, mylohyoid and middle pharyngeal constrictor muscles to fix the hyoid in position, and pulling the larynx upwards using the laryngeal muscle connecting the larynx to the hyoid, the thyrohyoid (Kaplan, 1960: 117).

The second possibility is to raise both the hyoid and the larynx together. The hyoid can be raised by contracting the geniohyoid, genioglossus, mylohyoid and anterior belly of the digastricus, which act to pull the hyoid upwards and forwards, while simultaneously contracting the stylohyoid, posterior belly of the digastricus, palatopharyngeus, and the middle pharyngeal constrictor, which pull the hyoid upwards and backwards (Van Riper and Irwin, 1958: 366). The forwards and backwards components of these mechanisms are made to
Figure 1. A schematic diagram of the location and action of the muscles involved in the hyoid complex.

1. Hyoglossus m.
2. Geniohyoid m.
3. Mylohyoid m.
4. Middle pharyngeal constrictor m.
5. Thyrohyoid m.
6. Geniohyoid m.
7. Anterior belly of the digastricus m.
8. Stylohyoid m.
9. Posterior belly of the digastricus m.
10. Palatopharyngeus m.
11. Thyroid cartilage
12. Hyoid bone
13. Stylopharyngeus m.
14. Salpingopharyngeus m.
15. Sternothyroid m.
16. Sternohyoid m.
17. Omohyoid m.

(after Hardcastle)
balance each other, giving the overall result of raising the hyoid. The larynx can then be raised in two ways: either by actively contracting the thyrohyoid and shortening the distance between the hyoid and the larynx; or passively, by allowing the rising hyoid to carry the larynx up with it by means of the thyrohyoid membrane, whose median section thickens into the median thyrohyoid ligament (Kaplan, 1960: 121), which acts as a mechanical link between the thyroid and the hyoid. The thyrohyoid membrane and ligament have been described in this passive action as 'a checkrein to limit the distance separating these structures' (Saunders, 1964: 76).

Any of these possibilities allow the larynx to be kept in a raised position throughout continuous speech, with momentary positional fluctuations caused by the movements of the muscles directly involved in, or passively affected by, the production of segmental articulation.

There is also the possibility of pharyngeal muscles such as the stylopharyngeus and the salpingopharyngeus being directly involved in the physiology of raising the larynx (Greene, 1964: 48).

Kaplan (1960: 147) comments that the elevation of the larynx tends to decrease the length and caliber of the laryngopharynx. This affects resonance'. The chief resonatory difference from the characteristics of the neutral vocal tract setting is a tendency to raise the frequency range of the first formant, and to compress the frequency ranges of the second and third formants. There is also a tendency for laryngeal elevation to be accompanied by a rise in the fundamental frequency. This, however, is not inevitable, and can be compensated by adjustments of the pitch-control mechanism of the larynx. Since the laryngeal mechanism for achieving higher pitch
has as one of its contributors a raising of the larynx, (presumably 'because the larynx is braced up to the hyoid bone to withstand the strong pull of the cricothyroid muscle acting to stretch the vocal folds' (Catford, 1964: 314)), then to raise the larynx and leave the pitch undisturbed necessarily requires a compensatory adjustment of some sort. An example of a well-known speaker with raised larynx voice is the actor Derek Nimmo; another example is the television personality, James Burke.

Conversely, it is possible to compensate for a naturally low pitch range, in singing, for example, by raising the larynx as a continuously-held setting. In this way, many would-be tenor singers, whose intrinsic vocal equipment endows them with an optimum natural pitch range slightly lower than that appropriate for a tenor voice, manage nevertheless to achieve a (slightly strained) tenor-like pitch range.

Raised larynx often sounds to have a particular mode of phonation associated with it. This can be attributed to two factors: firstly, the change in resonatory characteristics implied in Kaplan's remark quoted above, about the 'decrease in length and caliber of the laryngopharynx', would affect the fine detail of the vibratory pattern of the vocal cords, because of the acoustic coupling of the resonatory system of the vocal tract and the phonatory system of the larynx source; and secondly, because of the mechanical interaction of the different complex muscle systems involved, as noted immediately above.

Auditorily, and with an implication of an empathetically-sensed physiological state, raised larynx voice often sounds rather strained. Van Riper and Irwin (1958: 310) offer this physiological explanation:
'When, in phonation, (the larynx) is raised, and the thyroid is tilted in the direction of the position used in swallowing, many abnormal patterns of muscular contraction take place. Certain muscles, which in normal phonation need only make movements of fixation, anchoring with their antagonists any of the laryngeal structures, must now make strong contractions. Other muscles, normally not employed, must be brought into play to operate the displaced larynx. The whole activity is productive of localized tension'.

See also the section below on settings of overall muscular tension, for the effect of this sort of tension on the auditory and acoustic characteristics of voice quality.

ii) **lowered larynx voice**

The larynx can be depressed by the action of the infrahyoid group of muscles, schematically represented in Figure 1. Together with the sternothyroid, which runs from the thyroid, the cartilage shielding the front of the larynx, to the breastbone, the infrahyoid muscles form a third muscular sling system connecting the hyoid to the breastbone (*sternohyoid*), and to the shoulder blades (*omohyoid*, which comes horizontally forward from the shoulder blades before travelling upwards to insert on the lower border of the hyoid bone *(Kaplan, 1960: 150)*). Kaplan (p.149) also suggests that the sternothyroid is 'perhaps the most significant in such activity' as depressing the larynx.

Depression of the larynx increases the length of the laryngopharynx, and one acoustic result of this is to lower the frequency ranges of the lower formants, most of all the first formant *(Fant, 1960: 64)*. There is a concomitant tendency, (again, not inevitably, and which,
as in the case of raised larynx voice, can be compensated for), to lower the fundamental frequency. Fundamental frequency compensation in lowered larynx voice seems a less common phenomenon than in raised larynx voice, however, and a lowered fundamental seems to be an almost ubiquitous accompaniment to lowered larynx voice. Harold Macmillan and Edward Heath have more in common than membership of the Conservative Party, - they both have, as a component of their voice qualities, settings of lowered larynx, with low pitch ranges.

A lowering of the larynx was said by Passy (1914: 20-21) to be a component of what he called 'sepulchral voice'. Sweet, on the other hand, thought that 'sepulchral tone' was merely the combination of low pitch and an exaggeration of what he assumed to be a typical English tendency towards a 'muffled', 'dull' quality, due to habitual 'slight separation of the jaws and neutral lip position' (Sweet, 1890b: 72). Sweet had earlier attributed the 'exaggerated dulling' in 'sepulchral tone' to the effect of what he called 'cheek and lip rounding' (1877: 97-98), rather than to the neutral lip position referred to in his later work.

If Sweet and Passy were using 'sepulchral' as an impressionistic term to refer to the same sort of quality, and if we concede that a lowered larynx setting is genuinely a component of such a quality, which Sweet's comment (1890b: 72) on its characteristic low pitch might support, then perhaps Sweet's earlier comments (1877: 97) on the lip rounding component are not so unlikely as they may seem at first sight. Passy might have agreed that lip rounding is also a possible feature of the holistically-labelled 'sepulchral voice', but in any case it is striking that lip rounding, and especially
lip protrusion, as a component of many sorts of lip rounding, have much in common with lowering of the larynx, acoustically, and hence auditorily. In all three cases, lowered larynx, lip rounding and lip protrusion, the acoustic effect (Fant, 1960: 641) is to lower the centre frequencies of the lower formants, (though the effects are not completely identical, in that in lowered larynx voice it is the first formant which is most affected, and in labial protrusion and lip rounding it is the higher formants which are most changed).
The configurational similarity of lowered larynx voice and lip protrusion is particularly interesting, in that they both effectively lengthen the longitudinal axis of the vocal tract.

It may be, then, that the concept of a 'sepulchral voice', for Sweet and Passy, was basically an auditory concept. In seeking for a physiological correlate for the auditory quality, they individually arrived at articulatorily very different conclusions, but ones which gave acoustically and auditorily very similar results.

As in the case of raised larynx voice, lowered larynx voice, also, seems to have a special phonatory quality associated with it, by virtue of the same principles of interdependence of the muscle systems involved and of acoustic coupling in the vocal system. At the present stage of research, there is inadequate knowledge of the fine details of such small phonatory adjustments. Only auditory imitation labels are available for the description of the phonatory quality associated with supralaryngeal settings such as these. For this reason, comment on the fine details of phonatory quality will be restricted to those occasions when some acoustic or physiological (rather than solely auditory) characteristics of phonation
can be specified.

iii) **labial protrusion**

In this third type of longitudinal setting, voices with labial protrusion increase the longitudinal axis of the vocal tract.

The physiology of protruding the lips forwards from their neutral position touching the central incisors is chiefly a function of the orbicularis oris muscle, which acts as an oral sphincter, and is composed, primarily, of the fibers of other muscles - particularly the incisive, buccinator, caninus, and triangularis - that attach into the lips. The orbicularis oris fibers blend rather freely at the corners of the mouth and thus form an almost continuous sphincter' (Van Riper and Irwin, 1958: 374).

As well as serving to close the lips and to pull the upper lip down and the lower lip up, it protrudes the lips, in conjunction particularly with the mentalis muscle, which runs from the upper part of the front of the lower jaw down to the skin at the central point of the chin, and in contraction can evert the lower lip, a usual component of lip protrusion (Kaplan, 1960: 274-275; Van Riper and Irwin, 1958: 375). See Figure 2 for a schematic diagram of the muscles involved in labial protrusion.

The acoustic effect of protruding the lips is to lower the frequencies of all formants, with the higher formants more affected, as discussed immediately above (Fant, 1960: 64).

There is usually an interaction between the longitudinal
Figure 2. A schematic diagram of the muscles involved in labial protrusion.

1. Orbicularis oris m.  
2. Mentalis m.
protruded setting and a latitudinal factor if more than the most moderate degree of labial protrusion takes place. Anything more than slight protrusion usually involves a certain amount of horizontal constriction of the space between the lips, (Sweet's 'inner rounding' (1890: 17)). This latitudinal action tied to protrusion is not a mechanical inevitability, since it is physiologically possible to have substantial protrusion without any such horizontal compression, but protrusion without lip-rounding of this sort seems rare.

Protrusion of the lips is occasionally asymmetrical, in either the vertical or horizontal plane or both, but discussion of such idiosyncratic factors is beyond the scope of this thesis.

b) Latitudinal settings

Latitudinal settings of the supralaryngeal vocal tract involve quasi-permanent tendencies to maintain a particular constrictive or expansive effect on the coronal cross-sectional area at some given location along the length of the tract, relative to the cross-sectional area appropriate to the neutral vocal tract as defined earlier.

These latitudinal constrictive and expansive tendencies can be brought about by the action of a number of the vocal organs, and the different settings will be discussed in five groups, according to the organ principally responsible. The five groups thus relate to the activities of the lips, the tongue, the faucal pillars, the pharynx and the jaw.

i) Labial settings

A description of the habitual extrinsic settings of the lips in
voice quality entails a discussion of the well-established but somewhat under-differentiated phonetic concept of labialization. However, a consideration of the concept of labialization does not exhaust the field of potential settings of the lips; and in any case, as we shall see later, the adoption of such terms as 'labialization' from the traditional terminology of segmental description relating to 'secondary articulation' can only be done with an explicit reservation. It may lead to a clearer exposition, therefore, if we initially approach the area of labial settings without reference to labialization as such.

The latitudinal muscular adjustments involved in labial settings consist of tendencies to constrict or to expand the space between the lips (the 'rima labiorum'), which will be called here the interlabial space, in two dimensions of the coronal plane of the lips, horizontal and vertical, compared with the neutral setting defined earlier. The interlabial space is defined not by the vermillion border of the outer anatomical edge of the lips, but by the maximum horizontal and vertical dimensions of the aperture through which the airstream can pass. The assumption is made that these horizontal and the vertical dimensions lie on exactly the same coronal plane. This is in fact not quite true, - for example, both in extreme protrusion of the lips and in extreme lip spreading with the angles of the mouth pulled laterally and backwards, the coronal plane of the maximum vertical dimension of the interlabial space is further forward than the coronal plane of the maximum horizontal dimension. Nevertheless, for descriptive convenience the assumption will normally be held to apply.
All possible states of the horizontal and vertical parameters, leaving aside for the time being scalar variations within a parameter, give a total of eight latitudinal settings which deviate from the neutral: horizontal expansion of the vertically neutral interlabial space; vertical expansion of the horizontally neutral interlabial space; horizontal constriction (vertically neutral); vertical constriction (horizontally neutral); horizontal expansion with vertical expansion; horizontal constriction with vertical constriction; horizontal expansion with vertical constriction; and horizontal constriction with vertical expansion.

We have already discussed labial protrusion as a longitudinal setting: combining protrusion and the neutral non-protrusion settings with the eight conditions listed above for the non-neutral latitudinal settings of the lips, we have sixteen different types of labial settings. Together with the protruded and non-protruded neutral latitudinal labial setting, this gives a total of eighteen categories of labial settings available for potential use in voice quality. All are physiologically possible settings, but some of the eighteen seem to be used very much more frequently than others (which is itself an interesting observation). Figure 3 is a tabulation of the eighteen settings, with an indication of the settings which, as a tentative impression at least, seem most and least common in their occurrence among speakers of English.

Figure 4 is a schematic diagram of the front view of the nine possible latitudinal settings.

A common labial setting deviating from the neutral is the one involving (moderate) horizontal expansion of the interlabial space.
<table>
<thead>
<tr>
<th>PARAMETRIC DESCRIPTION</th>
<th>NON-PROTRUDED</th>
<th>PROTRUDED</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>very common</td>
<td>rare</td>
</tr>
<tr>
<td>horizontal expansion</td>
<td>common</td>
<td>rare</td>
</tr>
<tr>
<td>vertical expansion</td>
<td>rare</td>
<td>fairly rare</td>
</tr>
<tr>
<td>horizontal constriction</td>
<td>common</td>
<td>fairly common</td>
</tr>
<tr>
<td>vertical constriction</td>
<td>fairly rare</td>
<td>fairly rare</td>
</tr>
<tr>
<td>horizontal expansion &amp;</td>
<td>common</td>
<td>rare</td>
</tr>
<tr>
<td>vertical expansion</td>
<td>common</td>
<td>common</td>
</tr>
<tr>
<td>horizontal constriction &amp; vertical constriction</td>
<td>fairly rare</td>
<td>fairly rare</td>
</tr>
<tr>
<td>horizontal constriction &amp; vertical expansion</td>
<td>rare</td>
<td>very common</td>
</tr>
<tr>
<td>PARAMETRIC DESCRIPTION</td>
<td>CONFIGURATION</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-----------------------</td>
<td></td>
</tr>
<tr>
<td>neutral</td>
<td><img src="neutral" alt="Diagram" /></td>
<td></td>
</tr>
<tr>
<td>horizontal expansion</td>
<td>![Diagram](horizontal expansion)</td>
<td></td>
</tr>
<tr>
<td>vertical expansion</td>
<td>![Diagram](vertical expansion)</td>
<td></td>
</tr>
<tr>
<td>horizontal constriction</td>
<td>![Diagram](horizontal constriction)</td>
<td></td>
</tr>
<tr>
<td>vertical constriction</td>
<td>![Diagram](vertical constriction)</td>
<td></td>
</tr>
<tr>
<td>horizontal expansion &amp; vertical expansion</td>
<td>![Diagram](horizontal expansion &amp; vertical expansion)</td>
<td></td>
</tr>
<tr>
<td>horizontal constriction &amp; vertical constriction</td>
<td>![Diagram](horizontal constriction &amp; vertical constriction)</td>
<td></td>
</tr>
<tr>
<td>horizontal constriction &amp; vertical expansion</td>
<td>![Diagram](horizontal constriction &amp; vertical expansion)</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.** Schematic diagram of the front view of labial settings (all with protrusion and non-protrusion)
with a neutral vertical component and non-protrusion, which resembles a fixed, slight grin. If the amount of horizontal expansion is increased, and/or vertical expansion is added to this setting, the 'fixed grin' is changed to a quasi-permanent 'smile'.

Another common labial setting is horizontal constriction with vertical expansion and protrusion. In varying degrees, a number of people in public life use this, - for example, Frank Phillips, the recently retired BBC announcer; Norman St. John Stevas, the M.P. and Catholic apologist; Eric Heffer, the M.P.

Another fairly common setting is horizontal and vertical constriction with protrusion, which characterizes Jack Jones, the trade union leader and Anthony Barber, the former Conservative Chancellor. This setting is sometimes popularly described as showing 'pursed lips', although the label is applied more frequently to the same setting with non-protruded lips.

A fairly rare setting is the one with vertical constriction, a neutral horizontal component and non-protrusion; this is typical of Nicholas Fairbairn, the Scottish advocate and M.P.; another is the one with horizontal expansion and vertical constriction, with non-protrusion, seen in Michael Bentine, the Russian-Peruvian Old Etonian comedian. Another interesting setting is the one with vertical expansion, with a neutral horizontal component, but, unusually, protruded lips, of which Edward Heath, the Conservative ex-Prime Minister is a rare example.

Perhaps the most easily identifiable settings, from auditory clues alone, are the 'lip-rounded' Frank Phillips type of setting of horizontal constriction and vertical expansion with protruded lips,
and the 'smiling' setting of horizontal and vertical expansion without protrusion. The second of these often involves a characteristic modification of the fine details of the mode of vibration of the vocal cords, and a tendency to replace bilabial segments with labiodental ones, giving what one might call 'smiling voice', so that the identification by auditory means of 'smiling voice', (as characterizes the BBC 'disc jockey', Tony Blackburn), relies on more than the effect of the labial setting alone.

It is now possible to give some indications of the equivalence between this approach to the three-dimensional description of labial settings and the traditional phonetic terminology developed for the description of lip positions in segmental articulation. The use of terms referring to 'secondary' segmental articulations for the purpose of describing features of voice quality is discussed below in the section on lingual settings, but some translations of orthodox phonetic labels, including 'labialization', can be attempted here. 'Labialization' as a term has been used in such a variety of ways that it is probably safe to suggest that the only articulatory action to which the various usages usually have any reference in common is horizontal constriction of the interlabial space. The same applies to the general term 'lip rounding'. It is only when the phonetic labels become more specific in their reference that detailed translation is feasible. Thus the terms 'close rounding' and 'open rounding' can be translated as follows: 'close rounding' is used usually to correspond to horizontal and vertical constriction of the interlabial space, with the degree of protrusion very marked for an articulation such as Cardinal Vowel No. 8, for example (see
Jones, 1962, Fig. 22, p.33), but less marked for Cardinal Vowel No. 7 (ibid. Fig. 21).

'Open rounding' would correspond to horizontal constriction and vertical expansion of the interlabial space, with marked protrusion. The vertical expansion gives way to vertical constriction in a progression to 'close rounding'.

The term 'lip spreading' seems to refer uniquely to horizontal expansion of the interlabial space, with no protrusion, although there is a tacit convention that a spread lip position is limited to postures without much if any vertical expansion of the interlabial space.

Sweet was one of the first phoneticians to begin to distinguish between the three dimensions of labial articulation suggested here. In his _Handbook of Phonetics_ (1877: 13-14) he distinguishes the different parameters in this way: firstly, 'Projecting (pouting) the lips (....) of course practically lengthens the mouth channel by adding a resonance-chamber beyond the teeth'; secondly, 'inner rounding' is his label for 'lateral compression of the cheek passage'; and thirdly, he sees 'lip-narrowing' as being the result of constrictive effort in the vertical dimension, and whose degree of aperture normally varies with the height of the tongue, 'high vowels having the narrowest, low the widest lip-aperture'.

In his _Primer of Phonetics_ (taking quotations from the 3rd edition, in 1906), Sweet makes this classification rather clearer. He continues to speak of 'pouting' (1906: 18); he implies that vertical constriction is the most important component of what he now calls 'outer rounding' - i.e. his previous 'lip-narrowing',
in saying 'In outer rounding - with which front vowels are rounded - the lips are brought together vertically' (1906: 17); he then suggests in effect that horizontal constriction is the most important component of his 'inner rounding' - 'Back vowels, on the other hand, are rounded by lateral compression of the corners of the mouth, and, apparently of the cheeks as well' (1906: 17).

Another phonetician who explicitly distinguishes between the three labial dimensions is Heffner (1950). He put forward a descriptive scheme for labial articulation in terms not dissimilar to the ones used in this thesis. He said that, apart from the 'spread lip position', there are

'two types of lip rounding. (a) A long narrow slit is produced between the two lips by bringing the lips vertically nearer each other. This is called vertical lip rounding. (b) A horizontally short, more or less oval opening is produced by closing the lips from the corners towards the center until only a small aperture remains. This is called horizontal lip rounding' (Heffner, 1950: 98).

He also noted that

'protrusion of the lips is often a concomitant of horizontal lip rounding. It is much less frequently found with vertical lip rounding' (ibid.)

Like most other writers on phonetics, Heffner pays no attention to the possibility of vertical expansion of the interlabial space, as opposed to its constriction. In fact he quite explicitly excludes factors to do with expansion from the consideration of lip rounding, in saying

'Rounded vowels are produced when the area of the aperture
of the mouth cavity is reduced by contraction of the muscles of the lips' (1950: 98).

But if one examines the photographs of Daniel Jones's lip posture in pronouncing the Cardinal Vowels and looks at the overall area of the 'aperture of the mouth cavity' as related to rounded versus spread vowels, then it is precisely the expanded vertical component which gives Cardinal Vowel No. 6, with open rounding, an aperture area scarcely less than that for the unrounded Cardinal Vowel No. 3 (Jones, 1962, Fig. 20 and Fig. 17, p.33).

No account is taken here of idiosyncratic asymmetries of latitudinal labial settings.

The physiology of horizontal and vertical expansion and constriction of the interlabial space has been described by a number of writers, and the following comments rely chiefly on the accounts of Hardcastle (1975, Forthcoming), Fromkin (1965: 93-109), Kaplan (1960: 272-274), and Van Riper and Irwin (1958: 374-375). Figure 5 is a schematic diagram of the action of the labial muscles discussed below.

Horizontal expansion of the interlabial space is achieved principally by contraction of the zygomaticus, risorius and buccinator muscles. The zygomaticus runs from the temporal bone diagonally forward and down to the angle of the mouth, and acts to lift the corners of the mouth upward and outward; the risorius runs from near the ear across the cheek to the corner of the mouth as well, and widens the mouth laterally; and the buccinator is the large, strong muscle which makes up most of the cheek wall, coming from the jaw near the ear to the corner of the mouth once again, pulling
Figure 5. A schematic diagram of labial muscles involved in latitudinal control of the interlabial space.

1. Zygomaticus m.
2. Risorius m.
3. Buccinator m.
4. Quadratus labii superioris m.
5. Quadratus labii inferioris m.
6. Orbicularis oris m.
the corner laterally and to some degree backwards.

Vertical expansion of the interlabial space, where this is done with no contribution from lowering the jaw, is almost entirely brought about by two muscles: the *quadratus labii superioris*, which runs from three points of origin (the side of the nose near the bridge, the lower edge of the eye-socket, and the edge of the temporal bone) down to the upper lip, pulls the upper lip upwards and slightly laterally; and the *quadratus labii inferioris*, which has its point of origin on the chin just to the side of the point of the chin, and runs up to the skin of the lower lip, contracting to pull the lower lip downwards.

Horizontal constriction makes use of the *orbicularis oris* muscle, whose sphincteric effect was described in the earlier section on labial protrusion.

Vertical constriction also employs sphincteric action of the *orbicularis oris*, with appropriate antagonistic tension from the muscles specified above for horizontal expansion, and from the *quadratus labii superioris* helping to position the upper lip, and the *quadratus labii inferioris* the lower lip.

The acoustic effect of horizontal expansion of the interlabial space, 'lip-spreading', is chiefly to raise the centre frequencies of the formants (Fant, 1957: 19); the effect of the settings discussed above as contributing to the various sorts of 'lip-rounding' (depending on the closeness of the rounding) is similar to that of protruding the lips, - a lowering of the formant frequencies, with the higher formants more affected (Fant, 1960: 64).
It is difficult to be more specific about the acoustic effects of the detailed variations of labial settings. The descriptive scheme has been discussed so far as if it were largely one for accounting for static, postural configurations. The factors suggested as relevant for contributing to the acoustic effect of the settings have been stated only as comparative categories of dimensions relative to the neutral setting without quantifiable values attached to the dimensions. In particular, the cross-sectional latitudinal specification of the interlabial space has been discussed only as a shape and not as a quantified area. But cross-sectional measurements are relevant to acoustic quality only in terms of cross-sectional area, and not cross-sectional shape. Difficulties immediately arise from this position, since it seems to follow that if two settings have exactly the same degree of protrusion, and different latitudinal shapes, but exactly the same cross-sectional areas, then the descriptive model would assert that the settings, being of different categorial values, are therefore also auditorily different, while acoustic theory would require that the settings should produce exactly the same acoustic effect. Clearly, in such a situation, an acoustically acceptable explanation has to be found, if the descriptive system is not to be undercut.

Two possible explanations come to mind. First, as briefly mentioned earlier, although the descriptive scheme treats the interlabial space, for analytic convenience, as being two-dimensional, it is in fact three-dimensional, even in the case of non-protrusion, with a depth corresponding at least to the front-to-back thickness of the lips. This thickness varies with the different conditions of
tension and contraction in the many interdigitating facial muscles involved in both longer-term extrinsic labial settings (which can also be affected by different mandibular settings), and in momentary segmental articulations. This will result in variations in the fine detail of the rapidly-changing cross-sectional area-function along the short, variable length of the lip-cavity. So two settings of different shapes of the interlabial space in two dimensions, but of equal cross-sectional area at that plane of maximum constriction, may, perhaps, almost never have identical cross-sectional area functions along the whole length of the labial section. There is also the related point, explored more fully in the section on settings of overall muscular tension, that since different shapes of the lip-aperture are created essentially by the action of different muscles or by different degrees of tension in the same muscles, then the acoustic properties of the cheek and lip section of the vocal tract may well be different for the different settings (in terms of radiation and absorption of sound through the cavity-walls, acoustically reflected as changes in the bandwidths of the formants most affected), and therefore give different auditory impressions.

Secondly, and perhaps more importantly, if, instead of looking at the problem as if it concerned a steady-state, long-held posture, we consider the situation of normal, continuous speech, it is clear that the muscular adjustments of different labial settings impose differing amounts and types of constraint on the ability of the lips to execute their segmental articulations, and of course it is precisely this constraining factor that is abstracted as the setting underlying speech. This constraint operates both in the
relatively steady-state aspects of articulations and in the dynamic transitions to and from such states. It is assumed that as listeners we receive our auditory clues to labial settings from both these aspects.

In considerations of the acoustic effect of lip settings in any particular individual, account may have to be taken, depending on individual anatomy, of intrinsic aspects of the acoustic influence of the spatial relation between the lips and the front teeth. Such speaker-specific details would not, of course, enter the description of the extrinsic settings, which are speaker-neutral.

iii) lingual settings

There are five different groups of lingual settings. Four major ones involve constraints on articulation in the sagittal plane of the vocal tract (that is, in the vertical, central, front-to-back plane). Three groups of sagittal settings concern the body of the tongue, the tip and blade of the tongue, and the tongue-root. The fourth sagittal setting is to do with limitations on the movements of the tongue radially away from its neutral position; the minor coronal setting is also concerned with factors of such radial movements, and the resultant degree of lateral curvature of the tongue-surface. These last two settings are among the many aspects that have to be considered in an account of overall muscular tension, and will therefore be discussed later, in the section on tension settings.

If we consider here only the sagittal settings of the body, tip and blade, and root of the tongue, they all involve the adoption of a setting which tends to maintain the relevant part of the tongue displaced
from its neutral position in the vocal tract, either towards or away from the upper surface or the back surface of the tract. The result is thus to constrict or expand the oral cavity or the pharynx. The problem arises of how to describe these long-term tendencies of lingual displacement, in articulatory terms. Phonetic theory of course already has, in the notion of 'secondary articulation', a means of describing displacements of the bulk of the tongue which colour, as it were, the auditory quality of types of segmental articulation of a greater degree of stricture. We can thus call a long-term tendency to keep the body of the tongue slightly raised and retracted from its neutral position in the mouth 'velarized voice', and a tendency to keep it lowered and retracted 'pharyngalized voice', and so forth. Abercrombie (1967: 93) discusses this application of the concept of secondary articulation to the description of voice quality, and it is necessary here only to note that, while an analysis of the segmental make-up of a voice quality such as velarized voice will show segments with velarization as a secondary articulation, 'secondary' applies only to that segmental level of analysis, and not to the level of voice quality analysis as such, where quasi-permanent velarization is in no sense analytically 'secondary'.

The use of labels of secondary articulation to classify voice qualities whose constituent, susceptible segments show the particular secondary articulation can be extended to apply not only to lingual settings, but also to labial, faucal, pharyngeal and velopharyngeal settings.

Returning to the analysis of lingual settings, both in the mouth and in the pharynx, we are accustomed, in the phonetic description
of segmental articulation, to think chiefly about the location and
degree of the maximum constriction of the vocal tract. We often
discuss the articulation of a vocoid segment in terms of the
location along the length of the vocal tract of the 'highest point'
of the tongue surface. This highest point is then taken to
characterize the configuration of the rest of the tongue surface,
which is assumed to be regularly curved, convexly, both sagittally
and coronally. Using this point as the descriptive datum is valid
and economical for describing segmental articulation, particularly
as one is quite often concerned to be specific about the location of
the maximum articulatory stricture. In describing voice quality
settings, however, it is less a matter of specifying relatively fine
distinctions of place of articulation than of stating fairly gross
tendencies for the positioning of the bulk of the tongue. We can
dispense with the attempted precision of using the 'highest point of
the tongue' as a datum point, by choosing a quite different datum
point. We are concerned with the positioning of the mass of the
tongue, and therefore can profitably take as our reference point the
long-term average speech position of the approximate centre of gravity
of the tongue (Laver, 1968).

Figure 6 shows the centre of gravity of the tongue lying
vertically beneath the junction of the hard and soft palates, in
the neutral configuration. The vocal tract, in a stylized version,
can be seen as an arc, of about 270 degrees, curving at relatively
constant distance round the centre of gravity of the tongue, from the
larynx to the lips. Settings of the body of the tongue can now be
visualized as resulting from the shift of the centre of gravity along
Figure 6. Stylized diagram of the sagittal section of the vocal tract in the neutral configuration, showing radial directions of movement of the centre of gravity of the tongue in extrinsic settings of the body of the tongue.

1. Centre of gravity of the tongue
2. Dentalization
3. Alveolarization
4. Palato-alveolarization
5. Palatalization
6. Velarization
7. Uvularization
8. Pharyngalization
9. Laryngo-pharyngalization
10. Junction of the hard and soft palates
any radius of the circle of which, in the neutral configuration, it is the centre.

The customary phonetic labels for place of articulation can be appropriately used, remembering that implications of secondary status apply only to the analysis of the segments which can be thought to be superimposed on the tongue setting, and do not apply to the setting as such. So various constrictive settings would result from the following radial movements of the location of the centre of gravity of the tongue: upwards towards the hard palate, and slightly forwards, to give a voice 'with palatalization', or palatalized voice; forwards and upwards for 'palato-alveolarization' or palato-alveolarized voice; forwards and slightly upwards for 'alveolarization' or alveolarized voice; forwards for 'dentalization' or dentalized voice; backwards and upwards for 'velarization' or velarized voice; backwards and slightly upwards for 'uvularization' or uvularized voice; backwards for 'pharyngalization' or pharyngalized voice (where the constriction of the pharyngeal part of the vocal tract is achieved by retraction of the body of the tongue into the pharynx, rather than by the sphincteric action of the muscles of the pharynx bringing the back wall of the pharynx forward as well, which is discussed under 'Pharyngeal settings').

It is probably unrealistic to hope to be able to distinguish auditorily between finer subdivisions than these, such as 'pre-velarized voice', 'alveolo-palatalized voice', and 'post-alveolarized voice', when one is gathering the necessary diagnostic information over quite a long stretch of speech. On the other hand, it may be possible to sub-divide the lingual settings that achieve
pharyngalization into pharyngealized voice versus laryngo-pharyngealized voice, where the former might be reserved for constriction chiefly of the middle pharynx, and the latter for constriction of the lower pharynx and upper larynx, with the centre of gravity of the tongue moving backwards and slightly downwards. A difficulty in drawing an auditory distinction between these last two is that constriction of the pharynx almost always involves other physiologically and acoustically linked phenomena, such as a vertical shift of larynx position downwards, giving a lowered larynx component (and a changed mode of the fine detail of vibration of the vocal cords), a tendency to pull the velum downwards (because the velum and tongue are attached to each other by the palatalo-glossus muscle), giving some nasalization. All of which makes pharyngealized voice into a perceptual complex in which the specific contribution of the pharyngalization is particularly hard to isolate and sub-divide.

Examples of a constrictive lingual setting are found in the television personalities Aimeé Macdonald, and Hattie Jacques, who play comedy roles appropriate to the 'little girl' voice they affect. In the terms just discussed, their voices are the result of a raising and perhaps a slight fronting of the centre of gravity of the tongue, to give palatalization or palato-alveolarization, (and in the case of Aimeé Macdonald raising the larynx and using breathy or whispery phonation). Something of the same effect, without the breathy voice, can be heard in the voice quality that people in our culture use (more usually older women) when talking affectionately to other people's babies (this also often involves protrusion of the open-rounded lips). A good example of velarized voice is the present Prime Minister, Harold Wilson. He also has a setting of denasalization, to be
discussed below. Both velarization and denasalization are indices of his regional origin, in that they characterize speakers from certain parts of Lancashire (Abercrombie, 1967: 95).

There is also an example from one of the nationally-representative articulatory settings discussed earlier, said to characterize French. The overall tendency of French speakers to raise and front the centre of gravity of the tongue is reflected by comments from Wallis (1953):

'The French articulate all their sounds nearer the palate, and the mouth cavity is not so wide ....'

(Kemp, trans., 1972: 211);

Wilkins (1668: 380)

'Others seem to thrust their words more (....) inward towards the palate, as the French';

Sweet (1906: 74)

'In French (....) the tongue is arched and raised and advanced as much as possible, and the lips articulate with energy. French therefore favours narrowness both in vowels and in consonants, its point-consonants tend to dentality, and, compared with the English ones, have a front-modified character ....';

Heffner (1950: 99)

'French is a language which is usually spoken with a high and tense forward basis of articulation';

and Honikman (1964: 78-79), where she writes that the tongue 'is anchored medianly, albeit lightly, to the floor of the mouth by the tip tethering to the lower front teeth (....) Of the free, i.e. untethered part of the tongue, the blade (or tip and blade) and the front are
the dominant articulators',

and

'In French, there is no lateral tension of the lingual muscles but strong thrust is given to the convexed dorsum, especially in articulating the front vowels. French people with whom this has been discussed say it feels as if they were "pushing the words forward out of the mouth".'

Settings where the centre of gravity of the tongue moves downward, or downward and forward, as a continual tendency, are rather rare. One possibility is that the voice which is impressionistically labelled as 'hot potato voice', or 'plummy voice', as if the speaker literally had a hot or uncomfortable object in his mouth, is produced on a tongue-lowered setting. When I have asked people to demonstrate what they thought to be a 'hot potato voice', I have always had the auditory impression of a certain amount of pharyngalization. This applies also to 'plummy voice', together with a tendency to keep the larynx in a lowered position. The voice qualities of Harold Macmillan and Edward Heath, commented on earlier as having settings of lowered larynx with low pitch ranges, also show a lowered and backed setting of the body of the tongue.

Because the tongue is of relatively fixed volume (Paget, 1930: 36), it follows that, unless there are strong, compensating, antagonistic tensions, any lingual constrictive setting will tend to have a corresponding and co-acting expansive tendency in one or more other parts of the vocal tract. Pharyngalization, for example, enlarges the front part of the oral cavity, as does velarization, to a smaller extent. It should be mentioned that the volume of the tongue is not absolutely fixed, however. Lindblom and Sundberg (1972), in
comparing the tongue contour lengths in the production of spoken
and sung vowels, conclude that

'Our investigation seems to suggest that the tongue contour
between the tongue tip and the glosso-epiglottic fold
exhibits some elasticity. The tongue is somewhat
stretched when the tongue tissue is raised and/or when
the larynx is depressed. The site of this spring-
like property cannot be determined from our measure-
ments but anatomical considerations make it seem likely
that at least some of it should be attributed to the
posterior portions. Also the tissues joining the
glosso-epiglottic fold with the larynx can be compared
to a spring. However, it appears to be stiffer than
that of the tongue tissues' (Lindblom and Sundberg,
1972: 1).

They establish that the line along the surface of the tongue in the
mid-sagittal plane from the tip to the glosso-epiglottic fold can
vary by up to 20% (11.4 cm. for a spoken open back vowel to 13.7 cm.
for a sung close front vowel, with spoken vowels having shorter
lengths than their sung counterparts). Variability within the every-
day speech of a given individual, however, is likely to be less than
the maximum given by Lindblom and Sundberg (1972: 2), and the statement
that the tongue is of relatively fixed volume, in the gross differences
between the different settings of the tongue that we are considering
here, can stand.

There is one type of setting of the body of the tongue which is
of a different kind from the ones just discussed. This is a setting
in which the centre of gravity of the tongue remains more or less in
its neutral position, and the segmental articulations tend not to
depart very far from the centre of the articulatory space.
centring setting is one usually found in so-called 'muffled voice', and it will be discussed under **lax voice** in the section on settings of overall muscular tension. Similarly, **tense voice**, the other setting in the section on overall tension, usually involves a lingual setting where the centre of gravity of the tongue remains more or less in the neutral position, but with the segmental articulations being characterized by movements of greater radial distance than in 'lax voice'. (Tense voice is often impressionistically labelled 'metallic voice').

After the group of lingual settings which can be seen as involving primarily a shift in position of the centre of gravity, there is a small group of settings where the centre of gravity is displaced from the neutral position merely as an 'enabling' factor. This is the group of settings of the tip and blade of the tongue.

One example would be the consistent use of the tip of the tongue as the principal active articulator in the front of the mouth instead of the blade, and another would be the reverse situation, of the blade instead of the tip. Honikman (1964) gives many suggestions about differences of shape and posture characteristically distinguishing the settings of the tip and blade of the tongue in English, French, Iranian, Russian and Turkish.

One further example is the tendency to maintain a hollowed, slightly retroflex setting. Both Sweet (1906: 74) and Honikman (1964: 77) suggest that this setting characterizes English speech (Received Pronunciation, rather, explicitly in Honikman, and one assumes that this was also the case with Sweet's comment).
For all these tongue tip and blade settings, the body of the tongue has to conform to a setting which enables the tip/blade to execute the appropriate movements. This would be particularly necessary in the case of the retroflex setting, where to allow the tongue-tip to curl up at relevant moments throughout the speaker's utterances, the body of the tongue seems likely to be set in a slightly lowered and possibly slightly backed position, compared with its neutral configuration. It therefore seems reasonable to maintain that the tip/blade system, though capable of affecting the quality of speech to an audible degree, is only relatively and not entirely an independent system, depending to some extent on the major tongue body system.

The limited independence of the tip/blade system is reflected in the perceptual nature of the effects of such settings. While the tongue-body settings can usually be heard underlying nearly every segment in continuous speech, the tip/blade settings exercise their auditory influence only when the tip or the blade is the active articulator and even then they are in competition, as it were, with the auditory colouring of the ever-present tongue-body setting on which any given tip/blade setting is superimposed.

The same can probably be said of a third lingual system, that of the root of the tongue. Not enough is yet known to be able to assert with confidence that the root of the tongue can be adjusted in habitual muscular settings which expand or constrict the pharynx sufficiently to produce different qualities in a speaker's voice. But it may well be, on the evidence of the contributions that sizeable movements of the root of the tongue make to segmental
quality, shown in X-ray records (Ladefoged, 1964), that settings of the root of the tongue would have to be included in a comprehensive model of voice quality, with the same semi-independent status as the tongue tip/blade system, with its actions needing to be facilitated by corresponding adjustments of the tongue body system. A descriptive statement of the physiology and acoustics of tongue-root adjustments would be very speculative at this time, and will therefore be omitted.

The muscular physiology of tongue control is complicated and, in some details, controversial. The identity of the prime movers, or protagonist muscles, is reasonably well agreed, however, for the major parameters of movement of the body of the tongue (although the situation is made more complex by the fact that moving the tongue in any given direction can be achieved in a variety of ways by the synergistic action of different co-acting muscles).

The following brief outline is distilled from Hardcastle (1975, forthcoming), Heffner (1950), Kaplan (1960), Luchsinger and Arnold (1965), Van Riper and Irwin (1958), and Zemlin (1964).

The body of the tongue can be lifted upwards primarily by the contraction of the styloglossus and palatoglossus muscles, sometimes with some assistance of the superior and inferior longitudinal muscles. Figure 7 shows the location and direction of pull exerted by these and other lingual muscles involved in settings of the tongue.

The styloglossus runs downwards and slightly forwards from the temporal bone to insert into the sides of the tongue; its upper fibres travel along the tongue, near the margins, almost to the tip of the tongue. The effect of contracting the styloglossus is to pull
Figure 7. A schematic diagram of the action and location of the lingual muscles.

1. Styloglossus m. 6. Hyoglossus m.
2. Palatoglossus m. 7. Genioglossus m.
3. Superior longitudinal m. 8. Middle pharyngeal constrictor m.
4. Inferior longitudinal m. 9. Geniohyoid m.
5. Transverse lingual n. (partly after Ladefoged)
the tongue up and back. It can be assisted in this by the palatoglossus, which makes up the forward faucal arch, and which starts in the forward part of the soft palate, curving laterally forward and down, to insert into the sides and upper part of the back of the tongue, with its fibres blending with those of the styloglossus (Van Riper and Irwin, 1958: 379), and with those of the transverse lingual muscle and the hyoglossus (Zemlin, 1964: 219). The action of the palatoglossus normally is to act as one of the depressors of the soft palate, but when the velum is fixed in position by the palatal levators, contraction of the palatoglossus pulls the body of the tongue upwards. (The palatoglossus will be discussed more fully in the section on velopharyngeal settings). The lifting action of the styloglossus and the palatoglossus may be helped by the superior and inferior longitudinal muscles, which, located entirely within the tongue and not attached to any part of the skeleton, serve to bunch the tongue from front to back.

The muscles which are antagonistic to the upwards and backwards tendency of the styloglossus and palatoglossus are the hyoglossus, which attaches the hyoid bone to the sides of the tongue, and the genioglossus. The hyoglossus exerts a pull on the body of the tongue in a downwards and backwards direction, if the hyoid bone is immobilized by the infra-hyoid muscles. It is chiefly responsible, with the styloglossus, for adjusting the vertical positioning of the tongue body in the production of vowels (and hence for the vertical component of the radial displacement of the centre of gravity of the tongue in lingual settings).

With appropriate vertical co-operation between the styloglossus and palatoglossus pulling the tongue up and back and the hyoglossus pulling it down and back, another antagonistic element is needed to counter
the combined tendency to pull the body backwards, and this can be supplied by the genioglossus. The genioglossus is the big, vertically fan-shaped muscle which (with the internal transverse and vertical muscles) makes up the bulk of the body of the tongue, running from the inner surface of the jaw, at the chin, backwards to the hyoid bone and upwards to the dorsum of the tongue. Its action pulls the body of the tongue forward, when the jaw is in a fixed position.

The body of the tongue is pulled downwards chiefly by the hyoglossus, with the hyoid braced against its upward pull by the synergistically acting infrahyoids.

The settings with an element of retraction from the neutral position are achieved by the contraction of all the muscles mentioned so far in this section, together with a contribution from the middle pharyngeal constrictor pulling the body and root of the tongue towards the back wall of the pharynx. The styloglossus, with the palatoglossus acting from a fixed velum, and the hyoglossus from a fixed hyoid, equalizing their vertical mutual counteraction, allow their common retracting tendency to help to pull the tongue body backwards. The backwards tendency is checked and held to the necessary degree by the forwards pull of the geniohyoid from a fixed jaw position.

The overall fronting component of settings of the body of the tongue is produced by the genioglossus as the prime mover, with antagonistic counteracting tension from the styloglossus, and hyoglossus from a fixed hyoid.

It is more difficult to be confident about the physiology of settings of the tongue tip/blade, because of the multiple possibilities
of different synergistic muscle systems producing the same articulatory result. The following comments should be regarded as plausible hypotheses, rather than statements of established fact.

The setting of the tongue tip/blade, where the tip is the principal articulator in segmental sounds with primary articulation in the front of the mouth, relies mostly on the contraction of the superior longitudinal muscle. This runs from the root of the tongue to the tip, immediately underneath the mucous membrane, and in protagonist contraction pulls the tip of the tongue upwards. In the retroflex setting, the superior longitudinal muscle contracts to a greater degree, and the tip is pulled both up and back. To be able to curl back in this way, the tongue tip has to be lengthened, and this is done by the transverse muscle. This is a horizontal muscle whose fibres cross the tongue from side to side, separating the superior longitudinal muscle from the inferior longitudinal. The transverse muscle doesn't reach the tip of the tongue, nor the surface, and its contraction narrows the tongue, thus protruding the tip.

The setting where the blade is the primary articulator in the front of the mouth has to have the tongue tip lowered, and this is brought about by the joint action of the inferior longitudinal muscle, which runs the length of the tongue from root to tip low down in the body of the tongue, and the upper fibres of the genioglossus.

In all three tongue tip/blade settings, the body of the tongue has to be in a position which enables the tip and blade to carry out their actions appropriately, i.e. the body is slightly retracted for the 'tip' setting, slightly fronted for the 'blade' setting, and
slightly backed and lowered for the 'retroflex' setting (as discussed above).

We come now to the acoustic correlates of the various settings of the tongue. There are various ways of characterizing the acoustic consequences of changing the position of the centre of gravity of the tongue. One way would be to specify the range of frequencies within which the centre-frequency of a given formant is constrained to fall as a result of the setting. This is probably the best way, but specifying the formant ranges for all the settings discussed above to this degree of precision is a very complex business, and is usually beyond the scope of this thesis. The alternative to specifying the formant ranges is to indicate in which direction the centre-frequencies of the formants tend to be changed, relatively to those in the neutral configuration, and to attach crude scalar estimations of the degree of deviation from the neutral norms. Thus, sometimes it is possible to say that the 'range of formant two is raised and compressed' relative to neutral values, but more often it will be necessary to use such phrases as 'formant two is raised slightly'.

It may be helpful in attaching a certain degree of absolute quantification to the comments below to list the values for the settings of formant ranges on PAT that conform to the neutral setting of a vocal tract 17 cm. in length: the range within which the frequency-values of the first formant are constrained to fall in the neutral setting is 100 - 1,000 cs.; the second formant 500 - 2,500 cs.; and the third formant 1,400 - 3,400 cs. The fourth formant on PAT is normally fixed in value at 3,800 cs., and is weighted to represent the spectral contribution of all higher formants.
Within these ranges, the values for the neutral vowel itself are:
F1 - 500 cs.; F2 - 1,500 cs.; F3 - 2,500 cs. - as mentioned earlier. It is against this acoustic datum that the shifts of formant values due to changes of extrinsic lingual settings are specified here.

The following commentary on acoustic characteristics incorporates data partly from the experiment on synthesising different voice qualities (Laver, 1964, 1967) reported below, but mostly from articles and books on acoustic phonetics by Arnold, Denes, Gimson, O'Connor and Trim (1958), Fant (1957, 1960, 1962, 1964, 1968), Peterson and Barney (1952) and Stevens and House (1963).

In palatalized voice, with the body of the tongue raised vertically towards the hard palate, the range of the first formant is lowered, relative to the neutral setting; the second formant is raised quite considerably; and the third is raised slightly.

In settings of the body of the tongue which involve a fronting component, as in palato-alveolarized, alveolarized and dentalized voices, the first formant is lower than in the neutral setting; the second formant is high, but drops in frequency progressively as the setting moves from palato-alveolarization through alveolarization to dentalization; and the third formant is kept relatively high.

In velarized voice, with the body of the tongue raised and retracted, the first formant is slightly higher than for the neutral setting; the range of the second formant is raised and compressed; and the range of the third formant is also raised and compressed.
In pharyngalized voice, with the body of the tongue retracted towards the back wall of the pharynx, and possibly slightly lowered also, the first formant rises and the second formant drops, causing an $F_1 - F_2$ proximity.

'Not until the tongue comes very close to the back wall will $F_1$ be lowered again' (Fant, 1957: 19).

The third formant is also rather low.

In the rather rare setting of the body of the tongue which involves a lowering of the centre of gravity of the tongue without any backward movement of the centre of gravity itself, the first formant is raised, and the second and third lowered.

The acoustic characteristics of the settings of the tip and blade of the tongue are difficult to specify, because they are coloured by the effects of the 'enabling' settings of the body of the tongue. The one setting that can be specified quite easily is the retroflex setting. When the degree of retroflexion is slight, the fourth formant is lowered and close in frequency to the third formant. When retroflexion is more severe, the third formant is lowered to values close to those of the second formant.

iii) faucal settings

Another group of configurational settings of the supralaryngeal vocal tract are settings of the faucal arches, or pillars, which can constrict the vocal tract in an approximately coronal cross-section at the back of the mouth.

The faucal pillars are the two sets of muscular arches, one
behind the other, formed by the palatoglossus and the palatopharyngeus muscles, located at the junction of the mouth and the pharynx, connecting the soft palate to the tongue, the side walls of the pharynx, and the larynx.

The front set of faucal pillars, which are visible at the back of the mouth supporting the pendent uvula centrally between them, are usually called the glossopalatal arches. They are made up of the paired palatoglossus muscle, described earlier in the section on lingual settings. The rear set of pillars are made up of the paired palatopharyngeal muscle, which runs from the velum downwards in the walls of the pharynx to be inserted in the posterior border of the thyroid cartilage of the larynx. There is a triangular space between the two sets of arches, which contains the palatine tonsils.

The major function of the faucal muscles in speech is to pull the velum downwards. The palatoglossus, as we have seen immediately above in the section on lingual settings, can also function as a lingual elevator, when the soft palate is braced by the palatal elevators (which will be discussed below in the section on velopharyngeal settings). The palatopharyngeus, similarly, when the soft palate is fixed in position, can help to raise the larynx and lower the pharynx.

When the palatoglossus and the palatopharyngeus contract against the resistance of a fixed velum and larynx, the effect is to approximate the sides of the arches, constricting the vocal tract at the mouth-pharynx junction. Figure 8 is a schematic diagram of the location and action of the faucal muscles.
Figure 8. A schematic diagram of the location and action of the faucal muscles.

1. Palatoglossus m. 4. Tongue 
2. Palatopharyngeus m. 5. Soft palate 
3. Thyroid cartilage
Latitudinal approximation of the faucal pillars has been said by a number of writers to affect the quality of the voice. Alexander Graham Bell (1908: 19-21) traced a type of voice quality that he heard in a school for the deaf (a quality he described as being 'decidedly unpleasant, the voice resembling somewhat the cry of a peacock', and as having 'a peculiar metallic ring, somewhat like the tone of a brass musical instrument') to the effect of approximating the faucal pillars. He was able to train these deaf speakers to relax the muscles of the pillars, 'and at once the voice became natural and pleasant in quality'. He also suggested, in an appendix to the same book (p.123), that

>'When the posterior pillars of the soft palate approximate so closely as almost to touch, a very disagreeable reedy quality of voice results, which can be best described as a sort of "Punch and Judy" effect.

Pike (1943: 123-124) tentatively associates the act of approximation of the faucal pillars with a possible 'lower pharyngeal constriction, glottal tension, and usually a raising of the larynx', and calls the result 'faucalization'.

Greene (1964: 48-49) says that

>'The palatoglossus and palatopharyngeus may be considered in relation to the pharyngeal resonator. Their muscular arches form a flexible and variable arch of communication between the oropharynx and oral cavity, and if over-tensed are capable of materially decreasing the dimensions of the oropharyngeal outlet and creating a "cul-de-sac" resonator as West calls it'.
She is referring to West, Ansberry and Carr (1957), and the 'cul-de-sac' concept is one which has been put forward to explain the acoustic-articulatory basis of nasality. This 'cul-de-sac' theory was originally suggested by Russell (1931) and West (1936), and will be discussed in detail in the section on velopharyngeal settings.

The metallic quality referred to above by Bell may be partially explained by Pike's hint that a degree of lower pharyngeal constriction and glottal tension is present in faucalized voice, suggesting a general state of marked muscular tension. Given that the palato-pharyngeous is connected to the thyroid cartilage, it is not surprising that its contraction should be associated with a disturbance of the fine mode of vibration of the vocal cords, and with constriction of the lower pharynx and upper larynx. The section below on tension-dependent settings of the whole vocal apparatus suggests some acoustic correlates for the tensed condition of the walls of the vocal tract which may be relevant to Bell's and Pike's comments, and for the component of nasality implied by Greene's statement. The acoustic effect of constricting the vocal tract at the junction of the pharynx and the oral cavity is to raise the value for the frequency of the first formant, compared with that for the neutral configuration, and to lower the value for the second formant (Fant, 1960: 210). The most frequent involvement of the faucal muscles in voice quality is of course in nasal voice, discussed later, where they contribute to the complex acoustic characteristics of nasality.

iv) pharyngeal settings

We have seen that constriction and expansion of the pharynx
can be due to the effect of lingual settings, where the body of the tongue can be retracted into the pharynx, constricting it, or fronted, expanding it, or where, perhaps, the root of the tongue might act semi-independently of the body, and constrict or expand the lower part of the middle pharynx.

There still remains, however, the possibility of constriction and expansion of the pharynx by means of the muscles of the pharynx walls themselves. Apart from their function in speech, these muscles play a large part in the physiology of swallowing, together with the muscles of the tongue, and the details of the physiology of the pharyngeal muscles as part of the overall swallow mechanism have been commented on at length by Bosma (1953, 1957a, 1957b, 1961a, and 1961b), and Bosma and Fletcher (1961, 1962), Greene (1964), Kaplan (1960), and Merritt, Nielsen, Bosma, Goates, Haskins, Ramsell and Lamb (1957).

Zemlin (1964: 225) describes the pharynx as follows:

'The pharynx is a cone-shaped tube about 12 cm. in length, and wider at the top than at the bottom. It is about 4 cm. wide at its extreme width superiorly and about 2 cm. from front to back. It narrows considerably until, at the level of the larynx in front and the sixth cervical vertebra behind, it is about 2.3 cm. wide. At its lowest extreme the pharynx is continuous with the esophagus, and at this level the front and back walls of the pharynx are in direct contact with one another and separate only to permit the passage of food into the esophagus'.

Settings of the upper part of the pharynx (the nasopharynx) will be commented on below, in the section on velopharyngeal settings, and
Some remarks have already been made on settings of the laryngo-pharynx in raised and lowered larynx voices. Discussion here will concentrate on settings of the middle pharynx, or oropharynx, which extends from the soft palate to the hyoid bone (Zemlin, 1964: 226).

Emphasis has been placed, throughout, on the complex mutual interdependence of the muscle systems used to achieve particular settings, which have been analyzed as if they were independent of each other. Together with the hyoid complex, (with which the oropharyngeal complex itself interacts), the settings of the oropharynx furnish one of the best examples of the intricate interlocking of muscle systems which affect, and are affected by, settings in various parts of the vocal tract and larynx. Laryngeal, velo-pharyngeal, tongue-body, tongue-root and hyoid-positioning settings all potentially interact with oropharyngeal settings.

The oropharynx is affected by all the pharyngeal muscles, which fall into two major groups, an

'outer and inner layer which are not readily separable throughout. The outer layer is arranged circularly and is comprised of three constrictor muscles'

(i.e. the superior, middle and inferior (pharyngeal constrictors)) (Kaplan, 1960: 203).

'These muscles change the diameter of the tube. The inner layer is roughly longitudinal, and it includes the palatopharyngeus, salpingopharyngeus, stylopharyngeus muscles, and other irregular muscle bundles. The inner muscles function in elevation, depression, expansion and contraction of the pharynx' (Kaplan, 1960: 204).

The complexity of function of some of these inner muscles is shown
by Greene (1964: 48):

'The stylopharyngeus and salpingopharyngeus muscles reinforce the lateral pharyngeal walls and upon contraction raise and shorten them, decreasing the transverse and longitudinal measurements of the pharynx. At the same time, by reason of their attachment to the larynx, they assist in elevation of the larynx'.

Figure 9 is a schematic diagram of the pharyngeal muscles and their action. The superior pharyngeal constrictor, the weakest of the pharyngeal constrictors, acts in a sphincteric role in helping to lift the velum to a closed position, and will be discussed in the section on velopharyngeal settings. But it should be noted here that some of the fibres of this muscle have their origin fairly low down on the sides of the tongue, and in contraction tend to lift the tongue root upwards and backwards, constricting the oropharynx to some degree.

The middle pharyngeal constrictor is the principal muscle involved in constricting the oropharynx. It has its point of origin on the horns of the hyoid bone, and from there it runs upward and backward round the sides of the pharynx, forming a U-shaped sling with the arms pointing forwards and down. The fibres of this muscle spread out in a fan-like shape, with only the middle fibres running horizontally, from the area of the root of the tongue round to the back of the oropharynx meeting in the midline pharyngeal raphe (Zenlin, 1964: 229); the lower fibres run downwards and backwards beneath the inferior constrictor, and the upper fibres rise to cover the lower fibres of the superior constrictor. In contraction, the effect is mainly to narrow the oropharynx; the hyoid tends to be
Figure 9. A schematic diagram of the location and action of the pharyngeal muscles

1. Superior pharyngeal constrictor m.
2. Middle pharyngeal constrictor m.
3. Inferior pharyngeal constrictor m.
4. Stylopharyngeus m.
5. Salpingopharyngeus m.
6. Palatopharyngeus m.
lifted, but only slightly because of the nearly horizontal plane of the fibres in that area (Hardeastle, 1975, forthcoming; Kaplan, 1960: 204).

The inferior pharyngeal constrictor rises in a backwards and upwards fan of fibres from its origin, which is on the cricoid and thyroid cartilages. These fibres also meet at the back of the pharynx in the midline pharyngeal raphe (Zemlin, 1964: 228). The lowest fibres run horizontally and the others rise upwards to cover most of the middle constrictor. In contraction it can pull the larynx upwards, but when the larynx is fixed by the infrahyoids, the inferior constrictor narrows the upper larynx and lower pharynx in a sphincter fashion. It is the broadest, thickest, and strongest of the pharyngeal constrictors (Hardeastle, 1975, forthcoming; Kaplan, 1960: 204; Zemlin, 1964: 229).

The stylopharyngeus muscle, also paired, runs down each side of the pharynx from its origin on the temporal bone to its triple insertion in the constrictor musculature, the palatopharyngeus and the back edge of the thyroid cartilage. Its contraction either pulls the larynx and the walls of the pharynx upwards, or, if the larynx is fixed by the infrahyoid muscles, widens the pharynx laterally (Kaplan, 1960: 205). This is opposed to Greene's view, quoted above, that

'The stylopharyngeus and salpingopharyngeus muscles reinforce the lateral pharyngeal walls and upon contraction raise and shorten them, decreasing the transverse and longitudinal measurements of the pharynx'.

Greene doesn't change her position in the latest (1972) edition of
her book and doesn't cite any source. The question hinges on the exact anatomy of the course of the stylopharyngeus muscle vertically, and Zemlin (1964: 229) supports Kaplan's description.

The paired salpingopharyngeus runs vertically from the cartilage of the Eustachian tube to the back and side of the pharynx, blending with the palatopharyngeus to insert on the thyroid cartilage. Contraction of the muscle 'opens the Eustachian tube, elevates the pharyngeal musculature, causes the bulging of a vertical fold in the lateral pharyngeal wall, and thus narrows the pharynx from the sides' (Luchsinger and Arnold, 1965: 448).

The lateral narrowing by this means and by contraction of the palatopharyngeus (the posterior faucal pillars) is substantial. Luchsinger and Arnold (op.cit., p.450) cite Harrington (1944) as saying that the salpingopharyngeus muscle is the most essential component of lateral narrowing; Greene (1964: 46) says that the contraction of the salpingopharyngeus bunches the lateral pharyngeal mucosa to the extent of 'diminishing the transverse diameter of the pharynx by one third'.

The acoustic effect of constricting the pharynx by means of either the pharyngeal muscles or by retraction of the body or the root of the tongue is likely to be similar, as far as formant frequencies are concerned. The effect, varying in degree with different pharyngeal configurations, would be a rise in the first formant and a lowering of the second, tending to cause a first formant - second formant proximity (Fant, 1957: 19).

The acoustic effect on formant frequencies of expanding the
pharynx would depend partly on what correlated constrictive tendencies were simultaneously exerted elsewhere in the vocal tract, but one component would be a lowering of the frequency of the first formant.

Another acoustic factor than formant frequencies would be the feature of formant bandwidths, deriving from the tension, and therefore acoustic characteristics of absorption, reflection and radiation of sound energy, of the muscular walls of the pharynx. Hardcastle (1975, forthcoming) suggests that isotonic contraction of the pharyngeal muscles narrow the pharynx, but that isometric contraction of the pharyngeal muscles, (where the configuration of the pharynx remains unchanged because the mutually-antagonistic tensions of all the local muscle systems exactly balance each other), would 'have considerable effect on the resonance quality of the laryngeal tone giving it a metallic, strident quality'. This would be because the tensed walls of the pharynx would damp the sound wave less than relaxed walls, and the resultant decrease in damping is reflected in narrower formant bandwidths.

This contribution of formant bandwidth factors to overall auditory qualities of the voice will be discussed more fully in the section on tension dependent settings.

v) mandibular settings

In continuous speech, most speakers visibly move their lower jaw vertically upwards and downwards, moving slightly for almost each segment. This perhaps is to be anticipated, given the intimate anatomical linkage that we have been discussing above between the jaw and the tongue and hyoid bone. This linkage is reflected in the way
that the jaw can be seen to move in sympathy with the articulations of the body of the tongue. Underlying these moment-to-moment variations of jaw position, a general tendency can usually be discerned towards keeping the jaw in a particular setting.

The jaw has four dimensions of movement, all of which can potentially be involved in extrinsic settings in voice quality. These dimensions are: **vertical** (close and open); **horizontal** (protruded and retracted); **lateral** (left and right); and **rotational**, in the coronal plane, with one side higher than the other. All these dimensions of movement are biologically important in biting and chewing, but speech doesn't usually exploit all four. The principal speech-related dimension is the vertical, and most mandibular settings in voice quality are of this sort. It is a nice support for the view that most extrinsic settings are usually derivative of the articulatory tendencies of spoken language, that while many speakers are characterized by vertical settings of the jaw, the use of the other dimensions is much less frequently seen.

There are two chief exceptions to this, although neither case is particularly common. The first is the use of a slightly protruded setting, usually coupled with a slightly open vertical setting. This can often be heard in idiosyncratic segmental pronunciations of the voiceless alveolar fricative in English; as an audible segmental quality, it is very distinctive, and even when the posture is generalized to an overall mandibular setting, the fricative remains auditorily very prominent.

The other case is one typified by the late Robert Newton in his film role of Long John Silver in *Treasure Island*, and much imitated.
for humorous effect. It consists of a laterally offset position of the jaw, with a slight coronal rotation, lowering the same side as the lateral shift. The visual effect of the setting is often exaggerated, in such imitations, by the lip opening being similarly laterally distorted.

A third, minor, case which might be mentioned here, though it will be discussed below in a moment, is the open and retracted jaw setting that is sometimes seen accompanying a lowered larynx setting.

Given that it is very much more common to see variations of vertical adjustments of the jaw, it will be these that will continue to be discussed here, mostly leaving the other possibilities aside.

The muscles that serve to close the jaw are very strong, and can contract very fast: Zemlin (1964: 240) states that

'although the jaw is a rather massive structure, it is (surpassed) in mobility only by the tip of the tongue. That is, the maximum rate of movement for the tip of the tongue is approximately 8.2 per second'.

Van Riper and Irwin (1958: 360) also point out that

'all the muscles that act strongly on the jaw act to close it. The jaw can be opened or dropped with much less strength than it can be closed. We must work to keep our mouth shut. It opens quite easily'.

This last quotation has implications for the muscular setting that we have adopted as the neutral, datum configuration of the vocal tract. Acknowledging that the tongue is to some degree independent of the mandibular setting, nevertheless it is necessary to emphasize that the position of the jaw envisaged in the specification of [the]
the neutral configuration is not one where the muscles which serve to close the jaw are completely relaxed, allowing the jaw to fall open under the force of gravity. Rather, the neutral mandibular setting is seen as the one which interferes least with the tongue and lips achieving a configuration of the vocal tract which most nearly approximates to an equal cross-sectional area along the full length of the vocal tract. This means that the neutral setting of the jaw has to be achieved by a positive degree of contraction of the muscles that lift the jaw, and that this setting lies between the two extreme possible settings, maximally open, and completely closed, with clenched teeth. Many intermediate possibilities exist between these two extremes: the voice qualities of Harold Macmillan and Edward Heath, discussed at a number of points above, while similar in other extrinsic respects such as lowered larynx with low pitch range, and a lowered and backed setting of the body of the tongue, are different to some extent in their mandibular settings. Edward Heath tends to keep his jaw setting either neutral or slightly open, while Harold Macmillan seems habitually to allow his jaw to drop open nearly to its maximum: it may well be that the relaxed state of the jaw muscles in such a setting gives him his air of continual, wondering astonishment at the foolishness of the world, particularly when combined with his habit of keeping his eyebrows raised for an exaggerated length of time. This is rather a frivolous comment on Mr. Macmillan, but it serves to make the point that indexical judgements about speakers' characteristics gather their diagnostic clues just as much from components of voice quality (which often have, as in this case, a visible element), as from other areas of evidence.

There are three main paired muscles which serve to raise the jaw,
the internal pterygoid, the masseter and the temporalis muscles. Figure 10 shows the location and direction of action of these muscles.

The internal pterygoid muscle has its origin on the skull (on the perpendicular plate of the palatine bone and on the lateral pterygoid plate (Zemlin, 1964: 243)). It runs backwards, downwards and laterally, to insert on the upper middle surface of the jaw bone at its angle and on its vertical section. Its chief function is to raise the jaw, assisting the masseter muscle, but it can also be used to protrude the jaw, and used unilaterally, to pull the jaw to one side (Kaplan, 1960: 265).

The masseter muscle is the most powerful jaw muscle (Hardcastle, 1975, forthcoming), and runs from the lower edge of the cheekbone more or less vertically downwards to the angle of the lower jaw (Van Riper and Irwin, 1958: 359). The more superficial part of the muscle lifts the jaw (and is used also in grinding the teeth together), and the deeper fibres can be used to protrude the jaw (Kaplan, 1960: 265).

The last of the muscles which act to close the jaw is the temporalis muscle. This is a wide, thin, fan-shaped muscle which starts on the skull, at the wide end of the fan, attached to a wide arc stretching from the temple to a point on the skull above and behind the ear. From this origin, the fibres converge rapidly as they run forward and down under the cheekbone to insert on the forward surface of the whole of the vertical part of the jaw bone, down as far as the angle of the jaw (Zemlin, 1964: 243). Because the fibres of the temporalis run forwards as well as downwards, it can be used to help to retract the jaw as well as to close it (Kaplan, 1960: 265).
Figure 10. A schematic diagram of the location and action of the mandibular muscles.

1. Internal pterygoid m.
2. Masseter m.
3. Temporalis m.
4. External pterygoid m.
5. Geniohyoid m.
6. Anterior belly of the digastricus m.
7. Mylohyoid m.
8. Genioglossus m.

(after Hardcastle)
Lowering of the jaw necessarily involves the co-operation of the three powerful muscles just described, even when the jaw is being allowed to fall open mostly under the force of gravity. The muscles which actively open the jaw are: the external pterygoid, the geniohyoid, the anterior belly of the digastricus, the mylohvoid and the genioglossus. Figure 10 shows the location of these muscles. The positions and directions of action of all the muscles have been discussed in earlier sections, except for the external pterygoid. This is a paired muscle with its origins on the skull in front of the hinge of the jaw - specifically, it has two origins; one on 'the lateral portion of the greater wing of the sphenoid bone, and the other (.... on ....) the lateral surface of the lateral pterygoid plate' (Zemlin, 1964: 241). The fibres of the muscle run horizontally backwards to the temperomandibular joint of the jaw, just in front of the ear, and exercising a forward pull on the topmost extension of the jawbone just above the joint and on the capsule of the joint, both rotate the jaw downwards and pull the jaw forwards. So, like many of the muscles associated with the jaw, the direction of the main pull is not just unidimensional. Also, being a paired muscle, it can contract unilaterally. When this is done, the effect is to rotate the jaw slightly to the non-contracting side (Van Riper and Irwin, 1958: 360).

Heffner (1950) gives a concise description of the actions of the muscles which lower the jaw, although he relegates the external pterygoid to the role merely of assisting the internal pterygoid to protrude the jaw and to move it laterally. He writes:

'The downward movement of the jaw can be assisted by the
contraction of a number of muscles connecting it with the hyoid bone, provided the latter is not allowed to rise. The digastric muscles, from the chin to the side of the hyoid, the mylohyoid muscles, which are broad sheets of fibres from the sides of the mandible to the body of the hyoid bone, and the geniohyoids from chin to hyoid, when contracted, tend to bring the jaw to the hyoid bone. If at the same time the latter is pulled upon by the sternohyoid and omohyoid muscles, the jaw will be drawn down vigorously. Whether or not the whole head comes with it depends in part upon the action of the muscles which close the jaws' (Heffner, 1950: 35).

This last comment, about the head bending forward and down, is interesting. Speakers who use a lowered larynx voice quality, with a low pitch range (deeper pitch being facilitated by the relaxing effect on the pitch mechanism of the larynx by the mechanical downwards pull of the infrahyoids), assume a jaw setting which often pulls their chin inwards and down, together with a slight rotation downwards of the head. This then would be an example of a constellation of extrinsic settings, with each component setting facilitating or being facilitated by the others.

In referring to different vertical mandibular settings, we can use the conventional phonetic terms close and open, when there is a deviation from the neutral setting. We could say of a speaker who habitually keeps his teeth clenched during speech, or nearly clenched, that he had (say) a 'velarized voice with raised larynx and a close jaw setting'. If it is needed to give labels to jaw settings which involve other dimensions of movement than the vertical, terms such as protruded, retracted, and offset could be used.

The acoustic effect of mandibular settings is similar to that
of the longitudinal and latitudinal labial settings. Stevens and House (1954: 35-36) point out that in the region of the vocal tract about 15 cm. from the glottis, the cross-sectional area and the length of the front 3 cm. or so of the tract are controlled primarily by the mandible and the lips. A close mandibular setting is thus likely to have an acoustic effect like that of a protruded, lip-rounded setting, - a lowering of the formant frequencies, with the higher formants more affected. An open mandibular setting will have an acoustic effect like that of a labial setting with spread lips, - formant frequencies being raised. In specifying the acoustic effect of any particular mandibular setting, the speaker's labial setting also has to be taken into account, since the relationship between mandibular and labial settings is such that each can magnify or diminish the other's effect.

(c) Velopharyngeal settings

Apart from the neutral velopharyngeal setting, which will be defined in a moment, there are two settings, giving nasal voice and denasal voice. More has been written on the subject of nasality than about any other aspect of voice quality. It nevertheless remains an area characterized in phonetic writings by misconceptions and vagueness. Part of the vagueness is due to the lack of explicit distinction between some of the terms used to refer to the phenomena of nasality. One quite common term, that might suggest itself as referring to a single phenomenon, is 'nasal twang'. It is probably fair to say that it is normally restricted to the description of a voice quality setting (Heffner, 1950: 31), as opposed, for instance, to nasality which occurs on an individual segment. But 'nasal twang' is not consistently used, as far as I can establish, for any particular given type of nasality in voice quality. We shall explore some of the
different types of nasality in a moment, but given that different
types do exist, we cannot be sure that 'nasal twang' is not being
used for different phenomena by the various writers who use the term,
such as Abercrombie (1967), Appaix, Sprecher, Hénin and Favot (1963),
Berry and Eisenson (1942, 1956), Bullen (1942), Davis (1941), Heffner
(1950), Greene (1972), Luchsinger (1968), Meader and Muyskens (1962),
Paget (1930), and Sweet (1877, 1890b).

Direct comment about nasality as a feature of voice quality
doesn't seem to begin in writings on phonetics until the seventeenth
century, when Sibscota (1670: 14), translating Deusingen (1660)
written

'.... those that are Deaf .... speak thorough the nose,
as Aristotle affirms in his second and fourth Problem ....'.

We have noted that Bayly (1758: 180) commented that

'The most remarkable ill tones arise from what is called
speaking through the nose and in the throat',

and that Herries (1773: 55) wrote about

'.... that dull, disagreeable sound, which we call
sneveling or SPEAKING THROUGH THE NOSE. The
latter term is entirely wrong, because it is the
defect of NOT speaking thro' the nose, which occasions
that impropriety of articulation'.

Webster (1789: 106-109) referred to nasality in the passage quoted
earlier about

'the drawling, nasal manner of speaking in New England
(....), the great error in their manner of speaking
proceeds immediately from not opening the mouth sufficiently.
Hence the words are drawled out in a careless lazy manner, or
the sound finds passage through the nose'.
By the first part of the nineteenth century, the label 'nasal', was being used for a quality of the voice, by writers such as Rush (1827) and Willis (1829) without any need for explanation, and by the time Sweet (1877, 1890b) and Bell (1908) were writing about voice quality, it was fairly firmly established.

We can mention in passing that a variety of terms are used for 'nasal voice' and 'denasal voice' in the technical vocabulary of speech pathology, where there is no established single usage. They include not only 'nasality' and 'denasality' (Moore, 1957), but also such labels as: 'rhinophonia' (Travis, 1931); 'rhinolalia (clausa, aperta, mixta)' (Luchsinger and Arnold, 1965); 'hyperrhinophonia', 'hyporhinophonia', 'hyperrhinolalia', 'hyporhinolalia' (Greene, 1972); and 'hypernasality' and 'hyponasality' (Van Riper and Irwin, 1958).

An alternative to the phrase 'velopharyngeal settings' used here could be 'velic settings', and 'velic setting' will often be used as an informal paraphrase throughout this section. 'Velopharyngeal' is preferred to 'velic' in the formal vocabulary of the descriptive system for a good reason, however. That is, there is a good deal more to consider in the physiology of nasality and denasality than just the position of the velum; and it is in keeping with the theme running through the whole of the descriptive model, of interdependence of muscle systems in different parts of the vocal tract, that 'velopharyngeal' is used as a reflection of the effect of velic activity on the activities of the pharynx (and hence of the tongue and the larynx). We shall see that some of the vagueness and misconceptions about nasality could well be attributed to the over-simplistic view of velopharyngeal action as involving only the positioning of the velum that is continually repeated in much of the phonetic literature.
The definition of the neutral setting of the velopharyngeal valve system was described earlier, in the specification of the neutral configuration of the whole vocal tract, as being one of velic closure. That was adequate for the earlier discussion of longitudinal and latitudinal settings of the tract, where analytic attention was focused on articulatory details which are relatively independent of velic involvement; but this will now have to be modified. To define velic settings as single, static postures of the velopharyngeal system, as it were, is to disregard the special function of the velum in continuous speech, where it is said to change dynamically from velic closure to velic opening depending on the type of segment involved. We might say that the neutral velic setting is one where the velum is closed for all but nasal segments. But this is superficial because it is not only fully nasal, non-oral, segments that are made with velic opening, but also, in English and presumably in all the other languages of the world, the anticipatorily-nasalized stretch of segmental production immediately preceding the fully nasal sounds. If one then allows the neutral setting to require velic opening on all 'fully nasal' segments and on the anticipatorily-nasalized stretch of segmental production immediately preceding those segments, the difficulty then emerges of how to specify the permissible (durational and qualitative) amount of anticipatory nasalization in a way that is adequately general to all languages rather than more narrowly language-specific. One way out of this difficulty is to adopt as a convenient fiction an interim definition that is based partly on physiological constraints on articulation. In this (articulatory) definition, the neutral velic setting has velic closure on all segments except those where velic opening is criterial for their phonological identity in a given language, and on those
segments immediately preceding them where anticipatory opening of the velum is controlled by universal constraints applying to the temporal integration of the activity of different muscle systems of the vocal tract. This approach has the apparent disadvantage of bringing phonological considerations to bear on what is commonly supposed to be the independent field of general phonetic description; but this might not be the disadvantage that it seems at first sight, because a reasonable case can be argued for the position that any linguistically-motivated theory of general phonetics must have its foundation in phonological assumptions of one sort or another. This is not the best place to explore the argument, though, and it will be discussed in the chapter dealing with semiotic aspects of spoken communication. For the moment, then, the neutral setting of the velum will be taken to be the one where the velum is closed except for fully nasal segments and the minimum possible stretch of preceding articulation. Not enough is known at the moment about unavoidable constraints on the operation of the various muscular systems of the vocal tract to be able to assert that English, for example, characteristically requires more than the minimum amount of anticipatory nasalization, or only the minimal amount, so it is in this sense that the definition of the neutral position is something of a convenient fiction.

It will be understood from the earlier comments about the 'phenomena of nasality', that the concept of 'nasality' is taken to include a number of auditorily distinguishable voice qualities, which can reasonably be grouped together as variations of the single major category of a nasal voice. The acoustic criteria by which this can be justified will be set out after an account of the physiology of the velopharyngeal system and a discussion of conventional phonetic
treatments of the area of nasality.

The physiology of the velopharyngeal system has been the subject of research by many workers, though largely from other disciplines than general phonetics. The facts about the action of the groups of muscles that serve to open and close the velum are reasonably well established, with not many areas of controversy, and the account given here represents the consensus of opinion amongst the summary reports of Hardcastle (1975, forthcoming), Luchsinger and Arnold (1965), Greene (1964), Zenlin (1964), Kaplan (1960) and Van Riper and Irwin (1958).

The mechanism which lowers the velum consists basically of two inverted muscular slings, made by the palatoglossus and the palatopharyngeus, both of them paired muscles. The palatoglossus has its point of origin in the forward part of the body of the soft palate, and curves laterally forward and down to insert into the sides and upper part of the back of the tongue, where its fibres blend with the styloglossus, the transverse lingual muscle and the hyoglossus. The palatoglossus forms the forward arch of the faucal pillars, and was discussed briefly in the earlier section on faucal settings. In contraction, its effect is normally to approximate the sides of the forward faucal arch, and to pull the velum downwards, when the body of the tongue is braced against its tension. When the tongue is not braced, however, palatoglossal contraction tends to pull the body of the tongue slightly upwards and backwards.

The palatopharyngeus also has its points of insertion in the body of the soft palate, and curves laterally downward, through the side walls of the pharynx, to insert in the back border of the thyroid
cartilage of the larynx. The upper part of the muscle forms the back arch of the faucal pillars, and was discussed briefly in the earlier section on faucal settings. Contraction of the palatopharyngeus approximates the sides of the back faucal arch, and pulls the velum downwards, if the larynx is braced by the infrahyoid system. If it is not braced, then the larynx and the lower pharynx tend to be pulled slightly upwards.

The palatoglossus and the palatopharyngeus combine synergistically to lower the velum, acting as a pair of slings with their concavity directed downwards. It is important to note that lowering the velum is not merely a passive relaxation of the muscles which lift the velum, allowing the velum to open by gravity alone. It has been established by the use of electromyography and cineradiography (Fritzell, 1969; Lübker, Lindqvist and Fritzell, 1972) that the palatoglossus in particular is active in pulling the velum downwards.

A schematic diagram of the location and action of the two muscles was given earlier in Figure 8, in the section on faucal settings. This information is included for convenience of reference in Figure 11, which is a combined schematic diagram of the action of the muscle systems which raise and lower the velum.

The action of the palatoglossus, when tensed, affects other settings of the supralaryngeal vocal tract. Diamond (1952) suggests that not only is the palatoglossus inserted into the sides of the back of the tongue, but that some fibres of the palatoglossus itself continue transversely through the body of the tongue to form a sphincter system rather than merely a sling system. Most writers
Figure 11. A schematic diagram of the location and action of the velopharyngeal muscles.

1. Palatal tensor m.  
2. Palatal levator m.  
3. Azygos uvulae m.  
4. Palatopharyngeus m.  
5. Palatoglossus m.  
6. Tongue  
7. Thyroid cartilage  
8. Skull  
9. Hamular process of the pterygoid bone

(partly after Van Riper and Irwin)
support the sling interpretation of palatoglossal anatomy and physiology, but in either case, sling or sphincter, the articulatory action of the body of the tongue is likely to be slightly constrained when the palatoglossus contracts. In nasal voice, with the velum mostly held in its lowered position, there often seems to be an auditory component similar to the effect of slight velarization, that is, to the effect of having a lingual setting in which the centre of gravity of the tongue is moved slightly radially upwards and backwards. One explanation for this impression, if it is an accurate auditory judgment, is this linkage between the lowered velum and the tongue. This is supported by a finding by Hixon (1949) that nasal speakers seem to retract and raise their tongues more than normal non-nasal speakers. An alternative or possibly complementary explanation is that the auditory effect labelled here as 'velarization' is the result, acoustically, of constricting the vocal tract at the velar region, and that this may be achieved not only by the movement of the body of the tongue, but also by the downward movement of the soft palate and the inward, coronal movement of the faucal pillars, particularly the palatoglossal arch.

Lowering of the velum in nasal voice will have an effect on a number of other types of settings. The palatopharyngeus, connecting the velum and the thyroid cartilage of the larynx, has an influence, when it contracts, on the fine detail of the mode of vibration of the vocal cords in laryngeal settings, both when the larynx is free to be pulled upwards and when the infrahyoid complex resists the upward pull. Nasal voice has been shown to have a special mode of phonation associated with it by Fletcher (1947), who is cited by Van Riper and Irwin (1958: 244) for his high speed cinefilms of the action of the
vocal cords in various voice qualities. They report that he found that 'in his subject, the vocal folds opened more abruptly and assumed a different shape in nasal phonation than they did in normal phonation'. Apart from this muscular linkage, the effect of nasality on the fine detail of phonation can also be partially ascribed to the changed acoustic coupling between the resonatory system of the tract and the larynx source.

There is another aspect of the interaction of the palatopharyngeus and the larynx. When the larynx is not braced by the infrahyoid musculature, and is pulled slightly upwards, the resulting change of the longitudinal axis of the vocal tract is reflected in changes of the resonatory characteristics, and hence also in the source - system acoustic coupling. Van Riper and Irwin (1958: 244) give a personal finding that supports this: 'Clinically, we have found many cases of hypernasality who showed a marked elevation of the thyroid on their highly nasalized vowels which did not appear on those less nasalized'. Phonetically, it is easy to test this tendency for the thyroid cartilage to rise slightly during the production of nasality, by lightly resting a finger tip in the central notch in the top of the thyroid while changing an oral vowel to its nasalized counterpart.

The mechanism which acts to raise the soft palate to give velic closure is more complex, and to some degree more controversial. Van Riper and Irwin (1958: 308-395) give a very clear account of differences of opinion that exist in this area, and the comments incorporated here of Bloomer, Buck, Calnan, Harrington and Podvinec are found in their outline.

The difficulty in giving a precise account of the velopharyngeal
closing mechanism lies partly in the fact that not only is its action different in speech from the biologically primary function of swallowing, according to Calnan (1954), Bloomer (1953) and Lübker, Fritzell and Lindqvist (1972), but also that the action differs between different individuals (Bloomer, 1953; Buck, 1954), and between different sounds in the speech of the same individual (Calnan, 1953; Harrington, 1944). Another aspect is that not only do the structure and condition of the nasal tract differ intrinsically from person to person, just as the rest of the vocal apparatus does, but the resonatory condition of the nasal tract of any given speaker also changes from day to day, in respect to such details as the consistency of the mucal lining (House, 1957). As Van Riper and Irwin point out (1958: 388-389), when factors such as these are coupled with the complicated anatomy of the velopharyngeal system, and its relative inaccessibility to direct observation, broad generalizations about the functioning of the system are difficult to make, and it is not surprising that a certain amount of controversy should still prevail.

Some agreement does exist, however, about the anatomy and physiology of the velopharyngeal system and the mechanisms which serve to raise the velum to give velic closure.

There are certainly four muscles involved in raising the velum, and possibly six. The four chief muscles are the palatal tensor, the palatal levator, the superior pharyngeal constrictor, and some fibres of the upper part of the palatopharyngeus. The two other muscles possibly involved are the uvular muscle (the azygos uvulae), and the salpingopharyngeus. (See Figure 11).
The paired palatal tensor has a number of points of attachment on the skull above and to the sides of the soft palate. It is connected vertically downwards to a tendon, which then bends at right angles round the hamular process of the pterygoid like a rope around a capstan, and continues inwards to insert horizontally into the sides of the soft palate. Contraction of the tensor, as the name suggests, tautens the tendon, which serves to spread and tense the soft palate laterally. It also provides a fixed band within the body of the soft palate to which other palatal muscles are attached.

The paired levator also has its points of origin on the skull, above and behind the soft palate. It passes downwards, forwards and inwards to insert laterally in the back upper surface of the soft palate, forming most of the velic mass. Contraction of the levator lifts the body of the velum, which has been tensed by the tensor. The bending of the soft palate to form a palatal 'knee' may be helped by the contraction of the small uvular muscle, the azygos uvulae, which forms the body of the uvula and serves to shorten it.

One view of the lifting sling action of the levator, together with the other components just mentioned, is that it should not be seen as a totally independent mechanism, but rather as the forward component of a velopharyngeal sphincter slightly tilted from the horizontal. Van Riper and Irwin (1956: 391) cite measurements made by Calnan (1953) of this sphincter, giving the transverse diameter (at rest) of about three centimeters and the sagittal diameter of one centimeter or less, so that the opening is oval, flattened from front to back. The back half of the sphincter, in the form of a
sling with its concavity directed forwards, is composed of some fibres of the palatopharyngeus, which are said to have fused with the upper part of the superior pharyngeal constrictor. In contraction, these two muscles pull the back wall of the nasopharynx slightly forward. (The physiological complexity of the velopharyngeal system can be seen in the double function of the palatopharyngeus muscle, which acts both as a palatal depressor (primarily), and as a palatal elevator (synergistically)).

Luchsinger and Arnold (1965: 449) emphasize the importance of the sphincteric view of the closure mechanism:

'It is to be stressed that the velum does not move like a hinged trap door, (...) as is so often claimed in various books. In reality, the palate represents the anterior portion of the complex velopharyngeal valve, which functions mainly as a circular sphincter. The actual point of contact between the palate and the protruding pharynx occurs at the level of the palatal knee (...). Besides palatal elevation, the entire closely interwoven and always synergistically active musculature of the pharynx is part of palatopharyngeal occlusion'.

Not all writers agree, however, on the importance of the contribution of the palatopharyngeus - superior pharyngeal constrictor combination. Kaplan (1960: 207) suggests that 'In the normal person the closure of the nasopharyngeal valve is effected chiefly by the highly movable soft palate, so that the importance of Passavant's cushion is questionable'. Passavant (1869) discovered that, in a case of cleft palate, velopharyngeal closure was achieved by the formation of a muscular bar, or cushion, on the back wall of the nasopharynx at the point where it is touched by the lifted velum.
This cushion is formed by the contraction of the superior pharyngeal constrictor and the upper fibres from the palatopharyngeus that are fused with it, and which were stated above as pulling the back wall of the nasopharynx slightly forward in closing the velopharyngeal sphincter. Kaplan's position is supported, against that of Luchsinger and Arnold (1965), by a number of writers on the subject of Passavant's cushion, who all deny the importance of the cushion in normal speakers (Calnan, 1954; Bosma, 1961a; Greene, 1964). Greene (1964: 47) cites Russell's (1931) x-ray photographs which 'do not show the bulge of Passavant's muscle, or significant forward movement of the posterior pharyngeal wall'. It may be that vigorous contraction of the superior pharyngeal constrictor and palatopharyngeus is a compensatory action in speakers with a short soft palate or a cleft palate, providing a large cushion to make up for velic inadequacies, while normal speakers manage efficient velopharyngeal closure largely by the movement of the velum, as Kaplan suggests, and hence do not need to exploit the potential contribution of the posterior part of the sphincter mechanism.

Van Riper and Irwin (1958: 389-391) argue for a compromise position, in the sphincter-sling controversy, where they suggest that the sphincteric action is secondary to the primary sling-type action of the levator and tensor muscles. They point out that the amount of lateral contraction of the nasopharyngeal orifice is greater than can be accounted for by sphincteric action, and is probably due to the contraction of the salpingopharyngeus, a paired muscle independent of the sphincter system. They also cite an observation by Podvinec (1952) that 'the closure of the nasopharyngeal passageway is too rapid in speech to be accomplished by sphincteric
action of large muscles, particularly in view of the relatively large size of the opening', and that 'anatomically, there is no recognizable dilator that would open the sphincteric passageway quickly or surely and with control'.

A description of the salpingopharyngeus was given in the section on pharyngeal settings, where Greene, (1964: 46) was quoted as writing that 'Contraction of the salpingopharyngeus muscle which runs from the eustachian tube to the thyroid cartilage plays an important part in closure of the pharynx by causing bunching of the lateral pharyngeal mucosa, and diminishing the transverse diameter of the pharynx by one third'. The importance of the contribution to velopharyngeal closure of the salpingopharyngeus is supported by Luchsinger and Arnold (1965), and Harrington (1944).

In thinking about the physiological correlates of nasality, it is probably sensible to adopt the compromise position of Van Riper and Irwin (1958: 391), that the action of the velopharyngeal system is 'a combination of valvular movement on the part of the soft palate and sphincter movement by the superior constrictor and its related fibres'.

It may be useful at this point to note some observations on the speed of velic movement, and characteristic areas of the velopharyngeal opening in different degrees of nasality, made by Björk (1961) using cineradiography and tomography, synchronized with sound spectrography. He found that the velum moves from closure to an open state in 130 msec, and from an open to a closed state in 160 msec. He also established that the rate of velic movement does not vary directly with speech rate, but lags behind; with a speech rate change of
100 to 200 to 300, the rate of velic movement in the same nominal terms was 100 to 130 to 160. The areas of the velopharyngeal opening that he found were 60 mm$^2$ for slight nasality, and 250 mm$^2$ for heavy nasality. So we are discussing a structure that moves fast, over rather small distances.

The outline of the anatomy and physiology of the velopharyngeal system given above is one which, within certain limits, most workers in the study of speech would accept, as a simplified account of a very complex area. Unfortunately, while an understanding of the working of the velopharyngeal system is necessary to an understanding of the phenomena of nasality, it is not enough.

This inadequacy arises from three aspects of the concept of nasality. Firstly, nasality is above all else an auditory concept, and not primarily an articulatory one able to be specified in terms of the position of the velum during speech. This means that the interim articulatory definition of the neutral velopharyngeal setting will have to be modified. Secondly, 'nasality is a cover term for a number of auditorily similar but not identical phenomena, as noted earlier, and the 'apparent homogeneousness of nasality is caused .... by empathic reactions of the listener in referring what he hears to the speaker's "nose"' (West, Ansberry and Carr, 1957: 156). West et al. also say, as an extension of this, that some sounds heard as 'nasal' are made

'in such a manner as to exclude the possibility that resonance through or in the nasal chambers plays any part in the production of the "nasal" quality. Indeed, it is unfortunate that the term nasal has been applied to this quality of tone' (West, Ansberry and Carr, 1957: 196-197).
Thirdly, nasality is a condition of resonance of a special kind. West et al. suggest that

'The timbre, or overtone structure, usually given the name nasality is the result of resonance in a cul-de-sac resonator, a chamber opening off from the passageway through which a sound is resonated and delivered to the outer air' (West, Ansberry and Carr, 1957: 196-197).

They refer to Russell (1931) for the original discussion of cul-de-sac resonance in nasality. Russell (1931), West (1936), and West, Ansberry and Carr (1957) suggest various possible locations for this cul-de-sac resonance, apart from the most usual location, the nasal cavity. We shall return to this cul-de-sac resonance theory in a moment.

As we shall see below, nasality is a complex area, and one should be cautious about extending explanations of the mechanisms underlying segmental nasality to cover the production of nasality as a feature of voice quality as well. Heffner (1950) writes that

'The contraction of the pillars of the fauces is a feature of the production of nasal vowels. However, a generally accepted genetic explanation of true nasality, or nasal twang, is not yet available. In any event it is evident that the mere passage of all or part of the breath stream through the nasal cavities does not of necessity produce what we call nasality' (Heffner, 1950: 31-32).

Segmental nasality is treated in most textbooks on general phonetics in a somewhat simplified fashion, and the simplifications are not always explicitly acknowledged. Of course, considerations of pedagogic expedience can quite reasonably inhibit writers of introductory texts
from embarking on the often complicated explanations necessary to do justice to the detailed realities of nasality. So it may be helpful here to try to give a brief, explicit account of these simplifications, in order to see more clearly how the mechanisms normally put forward to explain the production of segmental nasality may relate to mechanisms responsible for nasality as an extrinsic setting in voice quality.

The first simplification is the view that when the velum is closed speech is free from nasality, and conversely, when speech is free from nasality, the velum is closed. For example, Sweet (1877: 7-8) wrote that

'In forming all the non-nasal sounds the uvula is pressed up so as to cover the passage into the nose. If the passage is open the sound becomes nasal'.

This view, not unreasonable in Sweet nearly a hundred years ago, is still not uncommon in current writing on articulatory phonetics. Chomsky and Halle (1968: 316), for instance, write that

'Nasal sounds are produced with a lowered velum which allows the air to escape through the nose; nonnasal sounds are produced with a raised velum so that the air from the lungs can escape only through the mouth'.

This may be largely true, especially of the articulation of particular intermittently-occurring nasal segments in the stream of connected speech where segmental nasality is spasmodic. But it is not the only available means for producing nasality. It is not for the lack of available research data that this partial, simplistic account of the mechanism of nasality is perpetuated. Even in the late nineteenth century, Rousselot (1901) noticed that the velum is held slightly open throughout most of the course of normal speech,
without audible nasality. Van Riper and Irwin (1958), in one of the best reviews of the literature of the difficult area of nasality, say that experimental evidence has been offered 'for the belief that the velum does not completely close off the pharyngeal passage to the nose in normal speech' (Van Riper and Irwin, 1958: 239-251), by Hixon (1949), Kaltenborn (1948), Nusbaum, Foley and Wells (1935), and Wolfe (1945). Later in their book (1958: 292), Van Riper and Irwin also mention Kantner and West (1941), McDonald and Baker (1951) and Heffner (1949, ( - 1950 is the date given in Heffner's book) as resisting the simplification. Berry and Eisenson (1956) also support this view, and further point out that the degree of velic opening is variable. Warren (1964: 161) has shown that velic opening up to a maximum of 10 mm² is 'adequate for the required oropharyngeal pressure of occlusive (....) consonant production'. Kaltenborn (1948) quantified the characteristic sizes of the openings into the mouth and the nose from the pharynx in non-nasal and nasal speakers: the typical size of the opening to the nose (presumably on the front to back diameter) for non-nasal speakers was 1 mm, and for the opening to the mouth 11 mm; the measurements for speakers judged as nasal were 8.8 mm for the opening to the nasal cavity, and 3.1 mm for the opening to the mouth. He is quoted by Van Riper and Irwin (1958: 241) as concluding that

'nasality is caused by having too wide an opening into the nasopharynx in comparison with the opening into the oral cavity'.

The second simplification commented on by Heffner (1950: 31-32), is that nasal airflow always gives rise to nasality, and conversely, nasality always requires nasal airflow. It is certainly true that
airflow can give rise to nasality, but only under certain conditions. If, as we have seen above, normal non-nasal speech is in fact characterized by a lack of complete velic closure, then logically it follows that it should be possible to have airflow through the nasal cavity without giving rise to audible nasality. Benson (1951) found 'no relationship between the degree of judged nasality and the amount of nasal airflow in his subjects' (Van Riper and Irwin, 1958: 242). Nusbaum, Foley and Wells (1953) assert that 'It is quite possible to utter vowel sounds free from nasality even when air is flowing out of the nose' (Van Riper and Irwin, 1958: 243). Thus, clearly, airflow through the nasal cavity is not itself a necessary or a sufficient condition for the production of audible nasality. (Which, incidentally, makes the nasal airflow trace on an electrokymogram a less than reliable indicator of nasality). Nasality is essentially a condition of resonance, as asserted above, and the nasal cavity can resonate without the passage of air through it; one has only to think of the possibility of very marked nasality where the nostrils are held tightly closed, for example.

Considering for a moment exclusively the participation of the nasal cavity in the production of nasality, it is possible to assert, following Kaltenborn (1948) and Van Riper and Irwin (1958), that a vital factor in inducing resonance in the nasal cavity is the ratio of the latitudinal cross-sections of two openings, - the relatively horizontal opening from the pharynx into the nasal cavity, and the relatively vertical opening from the pharynx into the mouth. We have seen that Kaltenborn's cineradiographic study of velopharyngeal action showed that in non-nasal and nasal voices there is a major difference in the ratio in the two cases: - in non-nasal voice, the ratio of the nasal
port to the oral port was 1:11, and in the nasal voice was 3.8:3.1 (Kaltchborn, 1948). We also noted that Björk (1961) found that the cross-sectional area of the nasal port increased from 60 mm$^2$ in slight nasalization to 250 mm$^2$ in heavy nasalization, presumably with a concomitant decrease in the oral port areas. That the oral port is rather smaller in nasal voices than in neutral voices is a conclusion reasonably drawn from the comment reported earlier by Hixon (1949), that nasal speakers retract and raise their tongues more than normal speakers. This was attributed above to the effect of the palatoglossus, which, in pulling the velum downwards also tended to pull the tongue body upwards and backwards. Van Riper and Irwin (1958: 241), commenting on Hixon's finding, suggest that nasality sets in whenever the nasal port is relatively larger than the oral port.

One of the difficulties in achieving any precise quantification of the actual cross-sectional areas involved is that complete confidence in cineradiographic results is not usually possible, (Trenschel, 1969). This is especially the case with the small distances concerned in velopharyngeal activity, particularly given that one is looking at the outline of tissues lacking the sharper definition of any mobile bony structure. It is also difficult to be sure of the representative nature of radiographic evidence acquired with the help of radiopaque paint, which makes for an unusual speaking situation. Another technique has recently been introduced into the range of experimental phonetic methods of analysis, however, which can give fairly exact results, if a number of aerodynamic factors are known. This is the use of a hydrokinetic equation to predict the velopharyngeal orifice area from a knowledge of the pressure differential across the orifice, and the rate of airflow through the orifice. Warren (1964) and
Warren and Dubois (1964) were the first to apply this, in relation to work on cleft palate speech. They proposed a slight modification to the hydrokinetic equation for use in the investigation of speech: they introduced a correction factor $K$, ($= 0.65$), for the unsteady, non-uniform and rotational characteristics of airflow in the vocal tract, in the equation:

$$A = \frac{V}{K \sqrt{2 \left( \frac{P_1 - P_3}{D} \right)}}$$

where $A$ is the cross-sectional area of the orifice in cm$^2$, $V$ is the rate of airflow through the orifice in cc/sec, $P_1 - P_3$ is the pressure differential across the orifice in dynes/cm$^2$, and $D$ is the density of air (0.001 gms/cm$^3$).

This approach was tested by Lübker (1969) in an experiment with an actual model, and he reported that 'the area of the orifice can be predicted with considerable accuracy, thus strongly supporting the use of the Hydrokinetic Equation for predicting velopharyngeal orifice areas'. Given that airflow and pressure sensing devices can be used which do not interfere seriously with the naturalness of articulation, a more precise and confident quantification is possible of the interaction and aerodynamic factors in speech, (not only for measurements of velopharyngeal orifice areas, but equally for measures of cross-sectional areas in fricative constrictions, for example). One important additional factor in the aerodynamic-cum-articulatory study of nasality has to be added to this approach, however. Nasal airflow is 'dependent not only upon the amount of velopharyngeal opening, but also upon the amount of oral constriction' (Lübker and Moll,
1965: 271). "This implies that one measure, such as the velopharyngeal orifice distance, is not sufficient to describe the articulatory factors related to nasal airflow" (Lübker, 1965: 270). In this connection, Van Riper and Irwin (1958: 245) cite the findings of Kelly (1934) who showed "that his subjects who had excessive nasality had much narrower mouth openings" (i.e. the maximum forward oral constriction) 'in speaking than did his normal subjects'; they also cite Williamson (1944) as corroborating this, but say that Smith (1950) could find no such consistent tendency.

The most important single factor in the production of nasality, then, according to Van Riper and Irwin (1958) and Kaltenborn (1948), is the ratio of the sizes of the posterior oral and nasal openings. The actual dimensions of the openings will naturally vary from speaker to speaker, as part of the intrinsic differences of voice quality, and the degree of possible nasal airflow without nasality will depend on the anatomical differences between the speakers. Berry and Eisenson (1956) say that in some bass resonant voices, the nasopharynx is open most of the time, without producing audible nasality. It seems statistically reasonable to assume that most men with bass voices are men of larger stature and hence tending to have larger vocal organs. The absolute size of the posterior nasal opening can thus be quite large, and nasal airflow consequently quite copious, before audible nasality begins to be produced.

The third and last simplification often found in introductory phonetic textbooks is that resonance of the nasal cavity is the only resonance responsible for the production of nasality, and conversely, nasality always requires resonance of the nasal cavity.
Some experimental work has been done on the measurement of acoustic energy in the nasal cavity, but the findings are contradictory. Weiss (1954) found a correlation of .74 between the judged hypernasality of voices in his sample and sound pressure level in the nasal tract. But Shelton, Arndt, Knox, Elbert, Chisum and Youngstrom (1969) claimed that 'measurements of oral and nasal sound pressure level do not correlate highly enough with nasality judgments to serve as indices to nasality'.

We can concede that the resonance of the nasal cavity itself is the most common, and the most important factor to be considered in discussions of the acoustic and articulatory correlates of nasality. There are alternative possibilities, however, and to elucidate these, we return now to the notion of 'cul-de-sac resonance' mentioned above, first put forward by Russell (1931: 18). West, Ansberry and Carr (1957: 196-197) give a concise summary and illustration of the notion:

'The timbre, or overtone structure, usually given the name nasality is the result of resonance in a cul-de-sac resonator, a chamber opening off from the passageway through which a sound is resonated and delivered to the outer air. If a bottle were attached to the side of a saxophone, with the bottle opening tightly fitted to one of the lateral vents of the instrument, the bottle would serve as a cul-de-sac resonator. The saxophone would have a nasal quality not unlike that of a bagpipe, in which there is a system of cul-de-sac resonance. If a small hole were pierced in the bottom of the bottle, the nasal quality, though still present, would be somewhat reduced. If this hole were enlarged until it was greater than the lateral vent to which the bottle was attached, the nasality would disappear and the bottle cease to act as a cul-de-sac
resonator.

Wherever, along the tube from the larynx to the outer air, there is a side chamber whose only opening is into the main tube, there is a chamber capable of acting as a cul-de-sac resonator and of producing a quality of tone usually referred to as nasal; and wherever this side chamber has an accessory opening through it to the outer air, it may still function as a cul-de-sac resonator if the necessary opening is smaller than the aperture connecting the side chamber with the main tube.

Amalgamating these comments on resonance produced in a side chamber (to adopt Pike’s term (1943: 87) in preference to ‘cul-de-sac’) with the earlier discussion about areal ratios of the entries into the back of the mouth and the back of the nasal cavity, we begin to approach a reasonable summary specification for the configuration of the vocal tract in the production of nasality, as far as the nasal cavity is involved. Four cross-sectional areas are concerned: that of the entry to the oral cavity, and that of its exit (that is, the narrowest oral constriction); that of the entry to the nasal cavity, at the velopharyngeal orifice, and that of its exit, at the nostrils. In producing resonance auditorily acceptable as nasality, the following conditions will apply. Firstly, either the nasal cavity or the mouth cavity has to constitute a side chamber relative to the other. Whichever cavity has the smaller exit becomes the side chamber, provided that the exit is itself smaller than the entry to that cavity. Secondly, if Van Riper and Irwin (1958) and Kaltenborn (1948) are right in their comments reported above about the ratios of the areas of the entries to the two cavities, the side chamber will generate auditorily acceptable nasality only when the entry to the side chamber has an area approximately equal to or
greater than that of the entry to the other cavity. The nasal cavity is the usual side chamber of the two, in nasal voice, and anatomy facilitates the provision of a large entry area and a small exit area for the cavity. It will be recalled that the value for a typical velopharyngeal opening in heavy nasalization was given by Björk (1961) as 250 mm². The entry to the oral cavity is made correspondingly smaller, in nasalization, other articulatory factors being equal, by the intrusion of the lowered velum into the back of the cavity, which constrains the oral airflow to pass through the relatively narrow openings of the faucal arch on each side of the pendent uvula. And the fixed area of the exit from the nasal cavity into the nostrils is relatively very small compared with the large areas involved in the contribution of the velopharyngeal orifice to the production of nasality.

We can now consider briefly the relation between the production of nasality for purposes of segmental performance and its production in voice quality.

In the production of most nasal stop contoids (those with a place of articulation forward of velar), the effective side chamber is the oral cavity inwards of the oral closure. In velar nasal stops, there may be a small side chamber formed by the approximation of the back of the tongue behind the velar contact to the undersurface of the uvula (Hahn, Lomas, Hargis and Vandraegen, 1952), but as in

3. See Fant (1960: 60-61) for a discussion of the nasalizing effect of a 'shunting side branch' on the acoustic output of the vocal tract, and the influence of the area of the coupling of the side branch to the main tract. The larger the coupling area, the greater the nasalizing effect.
uvular nasals, there is normally a lateral space between the two sets of faucal pillars. Also, in velar and in uvular nasals, the surface of the tongue forward of the closure with the soft palate, if touched with a finger tip, can be felt vibrating with any but the very weakest degree of voicing. The surface of the tongue can therefore excite the resonances of the front of the mouth, and the oral cavity forms a resonant chamber in the formation of these two nasal stops. The oral cavity is made to resonate here in very much the same way that the nasal cavity can be in the voice quality popularly labelled as 'cold in the head voice', when entry to the nasal cavity is blocked by catarrhal mucus in a heavy cold, with the acoustic excitation being transmitted through the tissue of the soft palate itself, or perhaps through the mucal plug.

That velar and uvular nasals have oral as well as pharyngeal and nasal cavity resonance can very easily be demonstrated by producing either of them with a strong whisper, and changing the position of the lips repeatedly from a spread posture to a rounded one and back again. The pitch of the resonances of the front of the mouth can be quite clearly heard, falling markedly with increasing lip-rounding and rising again with the progressive lip-spreading.

The auditory effect that we perceive as 'nasality', in the case of velar and uvular nasals, is likely also to have part of its acoustic basis in the tuning that the nasal system applies as a resonator neck to the resonant frequency of the pharynx (Fant, 1960: 112).

In the production of nasalized segments, as opposed to nasal segments, conventional phonetic description specifies the nasal cavity as the necessary side chamber.
So the nasal cavity is implicated as a normal part of the production of all nasal and nasalized segments. Given that nasality, as a segmental quality, has to be able to be switched on and off rather rapidly, in the spasmodic articulatory fluctuations of continuous speech, it seems almost certain that conventional phonetic theory is quite right in attributing the control of segmental nasality to the action of the velopharyngeal system.

In nasality as a feature of voice quality, however, where the auditory effect has to be nearly permanently present, the involvement of side chambers other than the nasal cavity has to be considered a distinct possibility, in at least a minority of cases. This is all the more plausible in view of the variety of auditory qualities that we are willing to accept as 'nasal', in terms of voice quality.

The possible location of side chambers other than the nasal cavity has been discussed by a number of writers. Van den Berg (1962) says that

"Nasality is immediately recognized by the human ear, but the acoustical correlate is difficult to describe exactly (....). This might seem unimportant for the phonetician and not for the phoniatrist, but this would be a mistake. (....) Nasal qualities, at least qualities which are interpreted as being nasal, may arise without participation of the nose, by too large damping factors at other places of the vocal tract, primarily in the vicinity of the larynx (....) the clinician needs to be aware of this" (quoted by Greene, (1964: 183)).

As in this comment, discussion of side chamber resonator locations other than the nasal cavity is very often in work addressed chiefly to workers in speech therapy and pathology. This is to be expected
given that the use of other side chambers than the nasal cavity is more likely in speakers who suffer from some disability in speech, such as velic inadequacies of various sorts, and that speech therapists are more likely to meet speakers whose speech is idiosyncratic enough, whatever the cause, to distinguish them from most of the rest of their community, than are workers in other disciplines such as phonetics.

West, Ansberry and Carr (1957), in their book on the rehabilitation of defective speech, mention various possibilities for the locations of the nasalizing side chamber:

"Frequently there comes to the speech clinic a person whose voice is distinctly "nasal" in quality but whose vowel sounds are made with the nasal port unmistakably shut tight. Where is the cul-de-sac responsible for his nasality? Many guesses have been made in answer to that question: a pouch formed back of the larynx at the mouth of the esophagus; one formed between the epiglottis and the root of the tongue; two side pouches formed between the alveolar ridges (sc. the external alveoli J.L.) and the cheeks, But wherever the cul-de-sac is in such a case - and it may be in any or all of these places - this "nasal" quality usually disappears when the patient learns to phonate with a relaxed throat and tongue. It may well be that the contraction of the muscles pulls apart surfaces of the larynx and pharynx, of the epiglottis and tongue, or of the cheek and alveolar ridges, that would otherwise be in contact, thus creating cavities in which "nasal" quality can be produced" (West, Ansberry and Carr, 1957: 199-200).

The most frequently posited location for the side chamber, other than the nasal cavity, is the pharynx (Russell, 1931; West, 1936; Tarneaud, 1941; Wise, 1948); Tarneaud (1941), for example, talks about nasality caused by tension in the pharyngo-buccal cavity, as
'timbre mi-guttural, mi-nasalisé' with complete velic closure. Another interesting example is Paget (1930), who published his work on experiments with vowel-models a year before Russell's discussion of cul-de-sac resonance in nasality. Van Riper and Irwin (1958: 244) say Paget stated that suddenly pressing upward and backward on a speaker's pharynx just above the hyoid bone produces involuntary nasality, and they say that this tends to support the cul-de-sac theory of Russell and West. They go on, 'However, the act may also serve to tense the glossopalatinus and pharyngopalatinus muscles, thereby lowering the velum. Thus we doubt its validity as indicating a laryngeal pouch resonator responsible for nasalization'. Against this last comment by Van Riper and Irwin, Paget did demonstrate that it was possible, using an artificial vowel resonator made from 'a rubber tube (about one inch in diameter), attached to an organ reed and fitted with a cork tongue' to add 'an appreciable twang (....) to the vowel sound' by suitably pinching the tube near the opening at the reed end. Paget concluded that his experiment 'indicates that a part, at least (....) of nasal quality (....) is probably due to a constriction of some part of the pharynx' (Paget, 1930: 96). This experiment would lose some of its relevance, of course, if the sort of constriction Paget introduced in his flexible tube were more narrowly localized than the human vocal organs could produce in that region of the pharynx by muscular adjustment.

Paget shows that the notion of side chamber resonance, as an 'additional resonating cavity' was established as a possible cause of nasality before Russell's book appeared in 1931, by quoting Eijkman (1926: 278) who specifies a pharyngeal contribution to
nasality in the same terms that were applied here to the source of nasality in a uvular nasal stop. Eijkman, after referring to the theory that nasality derives from a constriction in the pharynx, says

'Now the additional resonating cavity of small size (....) is actually formed through the narrowing of the uvular aperture; for the smaller this aperture becomes, the nearer the palatopharyngeal folds will approach one another, thus giving rise to a small resonance chamber, enclosed by these folds and the back wall of the pharynx. It would seem, therefore, that it is not the narrowing of the uvular aperture that actually causes nasality, but that it is brought about by the formation of a small resonance cavity in consequence of it' (Paget, 1930: 96).

Greene (1964: 184) supports Paget (1930), Russell (1931), Tarneaud (1941), West (1936), West, Ansberry and Carr (1957), and Wise (1948), and localizes the possible side chamber in the lower pharynx and upper larynx, saying that

'it should not be forgotten that nasality may also be imparted to the voice by muscular constriction in the laryngeal cavity and the relative positions assumed by the ventricular folds, aryepiglottic folds and the epiglottis, also elevation of the larynx by the suprahyoid muscles'.

The sources of nasality in voice quality suggested by the writers above are rather varied, including as they do hypothesized resonatory contributions from the nasal cavity, the pharynx and the larynx. However, it is not too unreasonable to assume that the comments of all the above writers contain elements of truth, if one takes the position emphasized in this thesis that the muscular systems of the vocal tract form a unified, complex, interlinked and interacting unit, mechanically,
physiologically and acoustically. From the details of the physiology of the velopharyngeal system outlined earlier, it seems safe to assume that any adjustment of the system will inevitably affect, in varying degrees, many settings elsewhere in the vocal tract. We saw for example that adjustment of the velum entailed changes in laryngeal, pharyngeal, faucal and lingual settings.

The assumption that velic action is an essential component in the production of every quality that we are ready to accept as nasal has been shown to be unsatisfactory, and has to be abandoned. This means that it is unnecessary to try to construct an elaborate description of velopharyngeal activity hedged about by cautious qualifications to cover the possibility of nasality made by other components of the vocal apparatus. Recognizing that 'nasality' is an auditory concept should allow us, ideally, to distinguish between velopharyngeal nasality, as it were, and other types such as pharyngeal 'nasality', faucal 'nasality', laryngeal 'nasality', and so forth. Nasality which is not produced by the action of the velopharyngeal system can be separately attributed to settings of the relevant mechanism, when enough becomes known to do this with confidence, and nasality due to velopharyngeal action can be considered independently.

The discussion of side chamber resonance other than that involved in resonance of the nasal cavity is speculative at present, and it gives rise to a problem of descriptive terminology concerning the term 'nasality' and related forms. Should 'nasality' be applied to the whole field of side chamber resonance, or only to resonance of the nasal cavity? West, Ansberry and Carr (1957: 200) opt for the second position, and conclude that 'it would be better to refer
to this quality as cul-de-sac resonance rather than nasal resonance, and that 'we are then justified in employing the term nasal to describe the speech of a person who, because of imitation of those in his environment or because of indifference to standards of good speech, utters all his vowels and semivowels with the nasal type of cul-de-sac resonance'. Leaving aside their prescriptive attitude about 'standards of good speech', they are right about the desirability of having a separate cover term for the variety of phenomena currently included under the umbrella of 'nasality'; and 'cul-de-sac resonance', or 'side chamber resonance' might be suitable candidates. The only sub-category of side chamber resonance that we currently really know enough about to use with some degree of confidence is resonance of the nasal cavity, - nasal resonance in the strict sense of 'nasal'. Not enough is yet known, in phonetics at least, about the sources of side chamber resonance in the pharynx, larynx, faucal pillars, to be able to differentiate accurately between the different types on an auditory basis, and until this is possible, it seems premature to develop an over-delicate descriptive terminology.

It is sufficient for the phonetic study of voice quality to note the generality of the concept of a side chamber in producing the effect we refer to as nasality, and to retain the well-established labels of conventional phonetics in their technical meanings: 'nasality', 'nasal', 'denasal' and 'nasalization' will now be taken to refer to control of resonance involving the nasal cavity by means of velic action, unless there is indication to the contrary. The point is taken that the criticism levelled here at expedient simplifications in some phonetic textbooks about the mechanisms
producing nasality might now perhaps be legitimately directed at this position also. However, the usage adopted here is at least taken after an explicit discussion of the general principle of side chamber resonance, rather than tacitly.

Analytic attention will now be confined to velopharyngeal settings in nasality and denasality, unless otherwise stated, in their strict interpretation.

In view of the auditory nature of the concept of nasality, the interim description of the neutral velopharyngeal setting, which was couched in articulatory terms of velic opening and closure, will have to be modified.

We can now say that the neutral velopharyngeal setting is one where the setting of the velum produces audible nasality only on those segments whose phonological identity requires them to be nasal, and on the minimum stretch of segmental performance immediately preceding them where anticipatory nasalization is physiologically inevitable. The actual positions of the velum are statable in terms discussed earlier of the critical ratios of the areas of the entry ports to the nasal and oral cavities, and of the exit of each cavity to the entry.

The nasal settings of the velopharyngeal system is then any setting of the velum which produces more audible nasality than can properly be attributed to the neutral setting. The different degrees of audible nasality in nasal voice can be treated in a scalar fashion, with labels such as 'slight', 'moderate', and 'severe'. The general principle of scalar labelling will be discussed later in this chapter,
and it is necessary here merely to say that any judgment of scalar degrees has to be made on absolute grounds, not grounds relative to the accent of the speaker's speech community, nor any other relative measure which is not general to all anatomically and physiologically normal human beings.

The denasal setting of the velopharyngeal system is one which prevents the occurrence of nasality. One problem here is that the term 'denasal' has been used sometimes to refer to the voice quality of a speaker who has a cold in the head. Luchsinger and Arnold (1965: 684) list as a synonym of 'denasal speech, hyperrhinolalia, hyponasality' the label 'head-cold speech'. A speaker who has a cold in the head may well have a blockage of the nasal port above the velum, and the consequence is that no nasal airflow is possible. We saw earlier, however, that the presence of nasal airflow is by no means an obligatory factor in the production of an auditory quality listeners are ready to accept as nasality. A posterior nasal blockage of this sort does not necessarily prevent the resonance of the nasal cavity; the cavity may well be acoustically excited by sound waves travelling either through the nasal plug or the tissue of the velum itself, just as earlier it was suggested that the oral cavity could be made to resonate quite audibly by whispering while maintaining a uvular closure and velic opening. If the tongue can demonstrably allow the transmission of the low amplitude sound waves produced in whispering, then it doesn't seem unreasonable to suggest that the velum may allow the higher amplitude, voiced waves to be transmitted into the nasal cavity. It has often been my impression that the quality of some (not all) voices of speakers with a head cold was not 'denasality'
in the strict sense of a complete absence of nasality, but rather a special, very highly damped kind of nasality.

The problem of describing the voice quality of speakers with head colds is not, however, the responsibility of this descriptive system, which is explicitly confined to the description of extrinsic settings able to be controlled by any speaker of normal anatomy and in reasonable health. It is legitimate to speculate on how healthy speakers go about the business of reproducing a quality of a speaker with a cold in the head, or other obstructions of the nasal port, such as adenoidal swellings, by means of extrinsic adjustments of the normal vocal apparatus. This is the situation Abercrombie describes in the Liverpool accent. He suggests that

'people can be found with adenoidal voice quality who do not have adenoids - they have learnt the quality from the large number of people who do have them, so that they conform to what, for that community, has become the norm. (Continuing velic closure, together with velarization, are the principal components needed for counterfeiting adenoidal voice quality)' (Abercrombie, 1967: 94-95).

By restricting the denasal setting to refer to the prevention of nasality (assuming as before that nasality is produced by the action of the velopharyngeal system), the position is taken that speakers using this setting treat segments which are nasal in the speech of other speakers as susceptible to the influence of the setting, making nasal stops oral. The problematic question of on what basis segmental susceptibility to the effect of voice quality settings can be explained is noted but left aside.

The physiology of the velopharyngeal system which is relevant to the velum always being raised enough to prevent audible nasality has
already been specified in the earlier discussion of the palatal elevator complex.

We come now to a brief discussion of the acoustic characteristics of nasality.

There has been a fair amount of research devoted to this area, mostly as it relates to segmental nasality in nasal contoids and nasalised vocoids. The conclusions of the research tend to be rather varied, although there is a measure of agreement about the principal features. The variability can be partly explained by the multiplicity of physiological components which has just been discussed. Another important source of the variability derives from the anatomical variations between different speakers (Bjuggren and Fant, 1964), and to some extent to the variability within the speech of a given speaker on different occasions, noted earlier (House, 1957), in details of the consistency of the mucal lining of the nasal cavity, and the underlying tissues. The effect of all these variabilities is that a fairly wide range of acoustic phenomena are perceptually acceptable as indicators of nasality, as will be seen in the summary below of the acoustic characteristics.

Most accounts of the acoustics of nasality assume a simplified anatomy of the nasal cavity, treating it as a single tube. Fujimura (1962), however, writes that

'In reality, of course, the nasal system at a central point branches into two separate tubes, each of which opens at one of the nostrils. If the system is symmetric with respect to the point at which the sound is measured, this geometrical branching is acoustically immaterial' (Fujimura, 1962: 240).
In the case of substantial asymmetry of the tubes, Fujimura (loc. cit.) and Fant and Bjuggren (1964: 6) state that additional high frequency nasal formants and anti-formants are introduced.

Fant (1960: 140-141) gives an account of the anatomy of the nasal cavity which is relevant to a discussion of its acoustic properties. He estimates the 'overall length of the nasal pathways measured as the shortest distance from the uvula to the outlet at the nostrils' is about 12.5 cm. Fant continues:

'The left and right nasal channels run approximately parallel for a distance of about 8 cm. from the nose opening and combine in the nasopharynx. Each of the frontal halves contains a bottom, middle, and upper branch in full communication at any cross-section. These appear to be too closely coupled to function as independent resonators. Provided the left and right parts show complete symmetry, they will function as a single cavity system. This is the ideal configuration (....), but it can be expected that asymmetry will cause an additional diffusion of spectral energy owing to the occurrence of formants from the left and the right pathways, and to the particular mixing in the nasal radiation. A greater damping of resonance in the nasal part than in the oral part of the vocal tract can be expected owing to the greater surface outline in area ratio of any cross-section except in the nasopharynx. (....) Nostril hairs will also contribute to the damping'.

The acoustic results of adding a nasal tract to an oral tract can be discussed under two headings: the resonatory characteristics of the nasal tract, and their effect on those of the rest of the vocal tract.
The most commonly reported nasal formant has a centre-frequency between 200 and 300 c/s (Curtis, 1942; Delattre, 1951; Fujimura, 1962; Hattori, Yamamoto and Fujimura, 1956; House and Stevens, 1956; Potter, Kopp and Green, 1947). Another nasal resonance is reported at about 1000 c/s by Smith (1951), House and Stevens (1956) and Joos (1948); and one at about 2000 c/s by Smith (1951), Delattre (1951), and Tarnóczy (1948). Tarnóczy specifies a slightly higher figure of 2500 c/s.

The bandwidths of nasal formants are said by House (1957) to be 300 c. for the lowest formant increasing to 1000 c. for those near 2500 c/s.

Anti-resonances, or anti-formants, which derive from any shunting side-branch present in the vocal tract (House, 1957: 199), are each paired with a nasal resonance (Fant 1960: 60). Values for anti-resonances have been specifically mentioned by Hattori, Yamamoto and Fujimura (1956), at 500 c/s, by Fujimura (1962) at 700 c/s as the lowest he found, and between 900 and 1800 c/s, varying with segmental articulation, by House and Stevens (1956).

The effects on the rest of the vocal tract of coupling the nasal tract to it are reasonably well agreed. The most general effect is an overall loss of power (Dickson, 1962; House and Stevens, 1956; Kelly, 1932). This attenuation is directly due to the introduction of anti-resonances by the side chamber resonator, which tend to absorb acoustic energy, especially in the higher frequencies (Curtis, 1942; House and Stevens, 1956; House, 1957). A reflection of the general attenuation is a broadening of all formant bandwidths (Kelly, 1932), which has another consequence, that of flattening
spectral peaks in the region between 800 c/s and 2300 c/s, giving an intensity plateau (Fujimura, 1962). It is almost certainly this general attenuation which is responsible for the drop in intelligibility of nasal voices reported by Diehl and McDonald (1956) and Moser, Drehor and Adler (1955) though Glasgow (1944) attributes the diminished intelligibility to the notion that nasality can call attention to itself, distracting the listener.

Considering more specific indicators of nasality, the most widely reported finding, Fant says (1960: 149), is a marked drop in intensity of the first formant (Björk, 1961; Delattre, 1954; House and Stevens, 1956; Kelly, 1932; Smith, 1951). Björk (1961) also notes a drop in the frequency of the first formant, in heavy nasalization. Always comparing the characteristics of nasality with the non-nasal equivalent, in the case of the second formant Björk (1961) found a drop in intensity, although Delattre (1954) disagrees. The third formant shows a drop in both intensity and frequency (Björk, 1961; Delattre, 1954; Kelly, 1932; Smith, 1951). On the fourth formant there is both a rise in intensity (Smith, 1951) and a drop in frequency (Delattre, 1954). There is a general drop in intensity on all formants above the fourth.

The characteristic drop in intensity of the first formant is brought about by a number of factors: the low frequency nasal formant boosts the intensity of the low harmonics, below the frequency of the first formant, and another nasal formant at about 1000 c/s reinforces the harmonics just above most values of the first formant. Antiresonances and an increase in all formant bandwidths then combine to lower the intensity of the first formant (Fant, 1960: 149).
The acoustic characteristics of denasality, as strictly defined to mean the absence of all audible resonance of the nasal cavity, are self-evident, - they are the absence on susceptible segments of the acoustic characteristics of nasality just discussed.

2. **Settings of the phonatory mechanism of the larynx**

(a) **Phonation types**

The domain of phonatory settings is limited by the same criterion that was applied to supralaryngeal settings: - only those settings which can potentially be controlled by any speaker with a normal vocal system will be admitted into the descriptive phonetic scheme.

Only a small number of basic types of phonation will be distinguished; although we shall see that this small number can combine with each other in various ways to make up a larger number of distinguishable, composite, phonatory qualities.

As in the previous section, physiological and acoustic specifications will be added wherever possible to the auditorily-based descriptions, and in some cases aerodynamic characteristics will also be given.

Phoneticians on the whole have been much more interested in supralaryngeal activities than in phonatory ones, and consequently more is known about the former than the latter. This is not without justification, when one thinks of the central importance phoneticians properly place on the capacity of speech to embody the distinctive patterns of language. The number of linguistic distinctions embodied by supralaryngeal activities is always much greater than those
signalled by different types of phonation, in every known language of the world. Ladefoged (1971: 16-18) suggests nine different modes of vibration of the vocal folds (which he says does not exhaust the total number found in different languages); and he points out that, taking the feature that he calls 'glottal stricture', no single language makes use of more than three out of the nine modes of vibration for manifesting linguistic oppositions. (See also Catford, 1964: 29).

The established theoretical structure for describing supra-laryngeal events is therefore considerably more delicately sub-divided than that for phonation types.

The question now arises of defining the neutral mode of phonation against which the other modes can be contrastively described. Near the beginning of this chapter, the neutral mode of phonation was said to be one where the vibration of the true vocal folds is periodic, efficient, and without audible friction. This description is worded to facilitate the discussion of other types of phonatory settings where one (or more) of the specified characteristics doesn't apply. So, we have settings where the phonation is not achieved by the true vocal folds alone (as in ventricular voice), or where the mode of vibration is aperiodic (as in harsh voice), inefficient (as in breathy voice), or with audible friction (as in whispey voice and whisper). This still leaves some types of phonation without contrastive identification against the neutral mode of vibration, - for instance, falsetto, creak, and creaky voice. This is partially put right by noting that the term adopted here for the neutral mode of vibration, modal voice, is Hollien's (1971), and that he says he chose the term
'modal' because 'it includes the range of fundamental frequencies that are normally used in speaking and singing' (Hollien, 1971: 320). We can incorporate this comment into the description of the neutral type of phonation, but we have to bear in mind that matters of pitch as such are not strictly relevant to quality. Falsetto is then characterized as having a range of fundamental frequency potentially higher than that of modal voice, and creak as having a range potentially lower. (The reason for the cautious use here of 'potentially' is an awareness that the pitch range of falsetto can overlap the range of modal voice very considerably, in terms of what is physically possible, even though male speakers using falsetto tend to utilise only the middle to high part of the possible range, in Western culture at least; and the inverse applies to creak, which can potentially be made on a high enough pitch to overlap the bottom end of the modal voice range). The characterization of creaky voice is a problem which will be discussed later in this chapter.

None of these descriptions of phonation types is satisfactory, however. The differences between modal voice and each of the other settings mentioned are of course more numerous and more complex than this brief sketch suggests. Even in the case of nodal voice, it is only a description, and remains far from a definition: the issue is merely hedged by saying that modal voice is the type of vocal fold vibration which phonetic theory assumes takes place in ordinary voicing, when no specific feature is explicitly changed or added. It is very hard to construct a satisfactory, short definition of modal voice, and instead I shall try to elaborate its characteristics in a summary outline of the aerodynamic and physiological factors
thought to be involved in producing this neutral mode of phonation. We will then use this base for the further discussion of the different phonation types.

The very widely accepted theory of vocal fold vibration is the aero-
dynamic-xyoelastic theory. The xyoelastic component was first stated by Müller (1837). He experimented with an excised human larynx set in a frame, exerting different tensions on the various muscles while air was blown through the glottis. He established that the vocal folds vibrate when they are adducted (i.e. brought together) to interrupt such an airstream, and that increasing longitudinal tension in the vocal folds is correlated with rising pitch.

The aerodynamic component is a relatively recent addition. Van den Berg (1958b) has shown that tension in the vocal folds is not the only factor to be considered in glottal vibration, in that the characteristic behaviour of airflow through narrow constrictions plays a very important part, particularly in closing the glottis. If we consider the train of events in one cycle of laryngeal vibration, the situation is basically as follows: with the glottis closed by muscular tensions acting on it in a number of dimensions (discussed later), sub-glottal air pressure builds up with pulmonic effort to expel the air from the lungs. It very rapidly reaches the pressure necessary to blow the edges of the glottis apart, and a jet of air shoots into the pharynx, momentarily relieving some of the sub-glottal overpressure. The glottal margins of the vocal folds which a moment ago were in contact, now form a narrow constriction whose aerodynamic effect on the upward flow of air is to act as a venturi tube. This makes the jet of air from the lungs accelerate through the narrow gap.
The Bernoulli effect on the air molecules in this accelerating jet causes a marked, very local drop in air pressure at the level of the glottis, the narrowest constriction in the larynx, and the glottal edges of the vocal folds are 'sucked' inwards towards each other by the negative pressure. The sub-glottal air pressure, momentarily released into the pharynx, has now dropped sufficiently to be overcome by the two forces combining to close the glottis, - the myoelastic tensions acting on and in the vocal folds, and the aerodynamic Bernoulli effect. The moment of glottal closure is usually the point in the phonational cycle at which the acoustic excitation of the supralaryngeal tract is most powerful. With the glottis closed, sub-glottal pressure builds up again, and the cycle of events repeats itself, between 50 and 200 times a second in voiced segments uttered by average male speakers in ordinary conversation.

This picture of the combined aerodynamic and muscular forces at work in laryngeal vibration was built up by the work of a number of researchers in the 1950's, including Smith (1954a, 1954b, 1956a, 1956b, 1957), and Faaborg-Anderson (1957). The major contribution came from Van den Berg himself (1954c, 1954d, 1956b, 1957a, 1957b, 1958a, 1958b) and with colleagues (Van den Berg, Zantema and Doornenbal, 1959). Van den Berg gives his own summary of the forces which interact to produce laryngeal vibration:

'(the aerodynamic-myoelastic) theory postulates that the function of the larynx is based on the interplay of three factors: (1) the aerodynamic properties of the air which actuates the larynx, (2) the adjustment of the larynx, brought about by the proper nervous activation of the various muscles, and the myoelastic properties of the laryngeal components, (3) the aerodynamic coupling
between (a) the subglottal system and the larynx, (b) the left and right vocal fold, and (c) the larynx and the supraglottal system' (Van den Berg, 1968: 291-292).

This comment on the different coupling factors is important in thinking about the fine detail of the auditory quality of the voice. It implies that the detailed mode of vibration of the vocal folds will depend partly on the degree of effort exerted sub-glottally by the pulmonic system, so that fine details of phonatory quality will necessarily co-vary with laryngeal intensity. Also, coupling factors between the left and right vocal folds become auditorily important when intrinsic asymmetries of the larynx are considered, in such cases as nodules and polyps on one or other vocal fold. And most relevantly, the articulatory state of the supralaryngeal vocal tract will influence the detail of the vibratory pattern of the vocal folds. This applies both in the case of the momentary changes of segmental articulation, and in configurational settings of the vocal tract as long-term articulatory tendencies. We noted earlier, for example, in the section on velopharyngeal settings, that nasal voice has been observed to be associated with a particular mode of vibration of the vocal folds. The theoretical implications of coupling between the larynx and the supralaryngeal tract make it reasonable to expect that all supralaryngeal settings, of any marked degree of deviation from the neutral configuration, should have their own characteristic laryngeal vibratory patterns.

Although the aerodynamic-myoeastic theory of laryngeal vibration is very well established now, it is not the only candidate in the area. There are two others, the neurochronaxic theory of Raoul Husson and his fellow-workers, and the mucoidulatory theory of Smith and Perrellö.
Husson's theory is diametrically opposed to the assumptions of the aerodynamic-veloelastic theory. Essentially, it dismisses muscle tension and the Bernoulli effect as the prime causes of phonation, and maintains that each cycle of vibration is the direct muscular response of the vocalis muscles (forming the vocal folds) to an individual neural command (Husson, 1950a, 1950b, 1951, 1952, 1954, 1955a, 1955b, 1956, 1957, 1962, 1964; Laget, 1953; Moulonguet, 1954; Piquet and Decroix, 1956; Piquet, Decroix, Libersa and Dujardin, 1957; Portmann and Robin, 1956). This theory has been widely rejected (Rubin, 1960a, 1960b; von Leden, 1961; Weiss, 1959), and the experimental findings of the vocal folds being made to vibrate by electrical stimulation of their nerve-supply in the complete absence of any airflow have not been replicated elsewhere. The effective importance of the neurochronaxic theory is not then the theory itself: its value is that it stimulated a decade of unprecedented amount of research into the details of phonation, including work towards the aerodynamic-veloelastic theory.

The muco-undulatory theory put forward by Smith (1961) and Perelló (1962a, 1962b) is of much smaller scale, and is complementary to the aerodynamic-veloelastic approach, not antagonistic. It concerns the effect on phonation of longitudinal waves travelling upwards to the glottis and into the laryngeal ventricles, and from front to back on the surfaces of the vocal folds (Greene, 1964: 40), in the loosely attached, wet, mucosal layer which covers the laryngeal surfaces. These waves have been observed in many high-speed cinefilms of the vocal folds in their phonatory state.

Smith (1961: 100) gives this account of the relevance of these
muco-undulatory waves:

'We must look upon the sliding mucosa between the vocal cords as an initial vibrator and we know that in man there are easy means of transmitting the surface movements of the mucosa to the body (sc. of the main mass of the folds, J.L.) because there is found below the edge an attachment to the underlying connective tissue and again to the muscle fibres. We thus have theoretically a double system, which in normal conditions would always act together in one vibratory pattern. In inspiratory voice (the mucosa) may act independently of the underlying mass'.

Smith then adds:

'Imagine (...) two skipping ropes circling against each other at ground level (i.e. upwards, J.L.). A subglottal adductory movement of cushion (i.e. vocal folds, J.L.) plus membrane (i.e. mucosal lining, J.L.) will result from this and the loose membranes will eventually independently be brought into contact by means of the sucking effect of the airstream' (loc. cit.).

He concludes that

'The important factor for initiating vibrations (seems) to be the existence of two thin surfaces in a vertical orientation' (Smith, 1961: 108).

Smith considers that the contribution of these mucosal phenomena is to the fine detail of the quality of phonation, and not to the fundamental frequency. Perelló (1962a, 1962b) believes the reverse. He argues that changes of fundamental frequency in the absence of muscular adjustments can occur when the consistency of the mucosal layer changes, as in laryngitis, or in premenstrual mucosal changes in women. Changes in the mucosal lining in this way constitute intrinsic conditions of the vocal apparatus, and lie outside the
scope of extrinsic control. They will be discussed further in the chapter on indexical aspects of voice quality, and the muco-undulatory theory is mentioned here only to show that the aerodynamic- myoelastic theory does not necessarily define all possible details of phonatory quality.

It will be assumed from this point onwards that the vibrations of the vocal folds in phonation are brought about in the way described by the aerodynamic- myoelastic theory. In the light of this theory, we shall need to distinguish between three general facets of laryngeal vibration: aerodynamic coupling not only within the larynx but between the larynx and the sub-glottal system, and between the larynx and the supralaryngeal system, as discussed above; the elastic resistance to glottal opening, - little in breathy voice, more in whispery voice, still more in modal voice, with severe resistance in harsh voice and ventricular voice; and mechanical constraints which result from the muscular linkage of the laryngeal muscle systems with those of any other settings of the vocal apparatus which happen to co-occur with phonatory settings in any particular voices.

We come now to a summary outline of the physiology of laryngeal vibrations in modal voice, with which that of the other phonation types may then be compared. It is important to emphasize the simplified nature of this outline, which is presented in this way only for the purpose of giving physiological relevance to the discussion of the broad tendencies towards maintaining particular laryngeal muscular settings as features of voice quality.

Three laryngeal cartilages form the basic frame within which
the muscular control of phonation is exercised. These are the thyroid, the cricoid and the paired arytenoid cartilages. Figure 12 is a schematic diagram of the relative position of these cartilages.

The thyroid is the big, shielding cartilage protecting the front and sides of the larynx from injury. It forms the 'Adam's apple' in male speakers, with its characteristic protruding, slightly pointed shape, and a V-shaped notch in the top edge at the centre. It is made up of two quadrilateral cartilaginous wings, or plates, vertically fused at the front under the central notch. The muscles which make up the true vocal folds and the ventricular folds are attached to the front, internal surface of the thyroid, at the point where the lateral plates fuse together.

The cricoid lies immediately below the thyroid, and is the uppermost of the tracheal cartilages. It is different from the other tracheal rings, in that while the others are incomplete rings with a gap at the back, where the trachea shares a common wall with the esophagus, the cricoid is a complete ring. It is often compared to a signet ring as the etymology of the name suggests, because the part at the back of the cricoid ring is considerably larger than that at the front. The vertical dimension of the 'signet' at the back is usually about 25 mm in the adult male, and about 8 mm at the front. Together with the thyroid, it forms an effective external protective structure for the rest of the larynx; it has also been said to be 'the foundation of the structures which, as a functional group, are called the larynx' (Heffner, 1950: 15).

The arytenoid cartilages, much smaller than the thyroid and the
Figure 12. A schematic diagram of the position of the principal laryngeal cartilages.

1. Thyroid cartilage
2. Cricoid cartilage
3. Arytenoid cartilages
cricoid, sit on the upper rim of the 'signet' of the cricoid at the back. They can rotate vertically and horizontally to a certain extent, as well as slide from side to side on the cricoid (Saunders, 1964: 72). They are shaped somewhat like small pyramids on a triangular base (Cates and Basmajian, 1955). The posterior ends of the true vocal folds are attached to the lower, forward angles of the arytenoids, called the vocal processes. The posterior ends of the ventricular folds are attached to the apex of each arytenoid.

The thyroid is connected to the hyoid bone above it by the thyrohyoid muscle and ligament. When the larynx rises, pulled up by the action of the thyrohyoid muscle, or of the stylopharyngeus muscle (discussed earlier in this chapter, and which runs from the skull nearly vertically downwards at each side of the pharynx to insert in the back edge of the thyroid), the thyroid 'slips up under cover of the hyoid' (Kaplan, 1960: 115).

The thyroid is also connected not only to the arytenoid cartilages, as stated immediately above, but also to the cricoid, by the paired cricothyroid muscle. This runs upwards, backwards and laterally from the inner surfaces of the forward part of the cricoid to insert in the lower edge of the thyroid. The effect of its contraction is to pull the front of the cricoid ring upwards towards the thyroid, which has the mechanical consequence of rotating the back of the cricoid, with its attached arytenoids, downwards and backwards from its neutral position. This lengthens and tenses the vocal folds, thus contributing to pitch control in phonation, and to the small changes in phonatory quality arising from
changes in the fine detail of the cross-section of the folds. The retraction of the cricoid also tends to bring the vocal folds slightly closer to each other (Saunders, 1964: 73).

The laryngeal muscles of interest here fall into two groups: firstly, those which, like the cricothyroid, can change the positions of the cricoid relative to the thyroid; and secondly, those which affect chiefly the positions of the arytenoids relative to the cricoid.

Figure 13 is a schematic diagram of the location and actions of the muscles of the first category. The muscles which can change the position of the cricoid relative to the thyroid are the cricothyroid muscle, and the paired thyroarytenoid muscles, which make up the true vocal folds and the ventricular folds. The thyroarytenoids, running from a fixed attachment at the fused angle of the thyroid to the mobile arytenoids, as noted above, are described by Heffner as follows:

"Each (side) is divided into two parts - an upper and a lower - by a recess, or ventricle, which undercuts the upper portion throughout most of its length. The lower portion of each thyroarytenoid muscle is attached to the vocal process of the arytenoid cartilage. The upper portion is attached to the body and the upper tip of the arytenoid. Indeed, some of the upper fibres of the upper portion of the muscle run upward into the folds which join the arytenoids with the edges of the epiglottis (i.e. the aryepiglottic folds). When contracted, the thyroarytenoid muscles tend to draw the arytenoids forward, at the same time tilting them towards the thyroid cartilage. (....) The upper portions of this pair of muscles, with their covering mucous tissue, are known as the ventricular folds (....)
Figure 13. A schematic diagram of the location and action of the laryngeal muscles connecting the cricoid cartilage to the thyroid cartilage, and related organs.

1. Cricothyroid m.  2b. Ventricular fold
2. Thyroarytenoid muscles  3. Ventricle of Morgagni
2a. True vocal fold
The lower portions (....) have a name of their own, the vocalis muscles (and they) constitute the vocal bands (....). In cross-section the vocal bands are triangular, being shaped much like the cushions of a billiard table, and only their median edges are free' (Heffner, 1950: 17-18).

The ventricle mentioned is the ventricle of Morgagni, and it should be pointed out that the ventricular folds have a rather different composition of tissue than the true folds; Kaplan describes the ventricular folds as

'thick rounded folds of mucous membrane developed around the ventricular ligaments. They are soft and somewhat flaccid. Each contains (....) a few muscle fibers and numerous mucous glands' (Kaplan, 1960: 124-125).

Saunders (1964: 73) confirms that they contain only a few muscle fibres. So their vibration will tend to be inefficient, with hypertension needed to adduct them sufficiently to phonate.

Longitudinal tension of the true and ventricular vocal folds can thus be achieved by two different actions: retraction and slight vertical rotation of the cricoid by means of the cricothyroid muscle; and by contraction of the muscles which make up the vocal folds, the thyroarytenoids.

The second category of muscles, those which control the positions of the arytenoid cartilages relative to the cricoid, have the function of opening and closing the glottis. Biologically vital in helping to control the airway to and from the lungs, they are small but powerful muscles, and are capable in combination of setting the glottis in a wide variety of adjustments. Figure 14 is a schematic diagram
Figure 11. A schematic diagram of the location and action of muscles connecting the arytenoid cartilages to each other and to the cricoid cartilage, and related organs.

1. Posterior cricoarytenoid muscles
2. Lateral cricoarytenoid muscles
3. Arytenoid m. or Interarytenoid m.
3a. Transverse arytenoid m.
3b. Oblique arytenoid m.
4. Vicular process of arytenoid cartilage
5. Vocal process of arytenoid cartilage
of these muscles and their action.

Only one muscle is normally used to open the glottis, the paired posterior cricoarytenoids (Kaplan, 1960: 150). These arise from the back, outer surface of the cricoid and run upwards and to the side to join the arytenoids on their rearmost angles, called the muscular processes. Their contraction pulls the muscular processes in an arc towards the back, and the effect is to rotate the arytenoids, pivoting the other ends of the arytenoids, the vocal processes to which the vocal folds are attached, outwards. Heffner (1950: 20) says that this happens 'at every normal inhalation'. In speech, their action is the major contributor to opening the glottis for voicelessness. Heffner continues:

'The action of these muscles is opposed to and can thus be controlled by the direct pull of the lateral cricoarytenoids and also by the direct pull of the thyroarytenoid muscles. The unopposed pull of the posterior cricoarytenoids widens the opening between the vocal bands to its maximum' (Heffner, ibid.).

The muscular action which closes the glottis is more complex. The lateral cricoarytenoids run backwards from the outer and upper surface of the cricoid on both sides, and like the posterior cricoarytenoids, are attached to the muscular processes of the arytenoids. In contraction, as indicated above, the lateral cricoarytenoids directly oppose the action of their posterior counterparts, and swivel the arytenoid cartilages forward and inward (Kaplan, 1960: 152). This action brings the vocal folds together, and closes the glottis.

There are two other muscular actions which help to close the
glottis. The first is that of the arytenoid muscle complex. The arytenoid muscle (sometimes referred to as the interarytenoid muscle) is made up of two sets of fibres: one of these sets is the transverse arytenoid muscle, which is an (unpaired)

'thick, rectangular mass covering the entire deep posterior surface of both arytenoids. It may be considered to originate along the muscular process and lateral border of one arytenoid and to cross over to reach the lateral edge of the other arytenoid. It draws the arytenoids medially by a gliding action which adducts the vocal folds' (Kaplan, 1960: 151).

It opposes the action of the lateral cricoarytenoids. The other part of the arytenoid muscle complex is the oblique arytenoid muscle. It is a paired muscle in the form of the letter 'X', and it lies behind the transverse muscle, on its outer surface. Each branch of the oblique muscle starts low down on the backmost surface of the arytenoid and rises crossing to the highest angle of the other arytenoid. Its contraction tilts the tops of the arytenoids towards each other, and in conjunction with the transverse part of the arytenoid muscle, helps to adduct the vocal folds (Heffner, 1950: 21; Kaplan, 1960: 151; Van den Berg, 1968: 294).

The second muscular action which can help to close the glottis is the contraction of the muscle forming the vocal folds themselves, the thyroarytenoids. The thyroarytenoids make a double contribution. The vocalis muscles (the part of the thyroarytenoids which make up the body of the folds) contract to exert longitudinal tension in the vocal folds, which reduces the length of the glottis (Van den Berg, 1968: 294). Contraction of the outer, lateral parts of the
thyroarytenoid muscles helps the lateral cricoarytenoids to bring the vocal processes of the arytenoids together, in adduction.

Kaplan notes that

'Any section of the glottal edge can be adducted or abducted without necessarily relating to any other part (Pressman and Kelemen, 1955). This is because the paired thyroarytenoid and vocalis muscles seem to be capable of differential segmental contraction. This activity can occur independently of the arytenoid movements' (Kaplan, 1960: 128).

To summarize the muscular actions which lead to glottal closure, then, Van den Berg writes:

'A contraction of the (powerful) interarytenoid muscles primarily adducts the apexes of the arytenoids and closes the back part of them so that no wild air can escape (....) A contraction of the lateral cricoarytenoid muscles adducts the vocal processes of the arytenoids and therefore the body of the vocal folds. This adduction is augmented by a contraction of the lateral parts of the thyroarytenoid muscles (this contraction goes along with an adduction of the vocal folds). These adductional forces provide a medial compression of the vocal folds and reduce the length of the glottis which is effectively free to vibrate' (Van den Berg, 1968: 294).

The notion of medial compression will be helpful when we come shortly to discuss the differences between the various phonation types.

Just before we come to that, it is necessary to set up a simple map for discussing the different locations in the glottis which are relevant to the characteristics of the different types of phonation. A useful point of reference for this is the vocal ligament, which runs
along the glottal edge of each vocal fold at the point where it normally makes contact with the other. We can then follow Catford (1964: 32) in using the term 'glottal', without any further qualification, to mean the whole length of the opening between the true vocal folds, from the front angle of the thyroid cartilage to the back of the arytenoids; and we can distinguish the ligamental glottis, which is the part of the full glottis formed by the vocalis muscles, with the length of the vocal ligaments along each edge, as opposed to the cartilaginous glottis, in the stretch where the arytenoid cartilages are located. The dimensions of these sections are of interest: Morris (1953) says that the 'intermembranous' (ligamental) part of the glottis is normally about 15.5 mm in the male, and 11.5 mm in the female. The length of the cartilaginous glottis is about 7.5 mm in the male and 5.5 mm in the female. This makes the full glottal length in males about 23 mm, and about 17 mm in females. Kaplan (1960: 128) adds that 'The widest part of the glottis is 6 to 8 mm in the male, and this can increase to about 12 mm, according to conditions'.

The nature of the glottis tempts one to regard it as a two-dimensional space; and for many phonetic purposes, that is sufficient. But in the case of contributions to the fine detail of phonatory quality, the need to take account of the changing three-dimensional configuration of the space between the vocal folds is very similar to the situation at the other end of the vocal tract, at the lips. The changing vertical thickness of the vocal folds from the outer wall inwards to the vocal ligaments at the edge of the glottal space reflects the interplay of the different tensions that are exerted in and on the folds by the laryngeal musculature, and this third,
vertical dimension is one factor among others which differentiate the major settings of the phonatory mechanism.

We are now in a position to state the major parameters of laryngeal control which are relevant to our discussion of different phonation types.

There are three parameters of laryngeal muscular tension which have to interact with aerodynamic factors of pulmonic airflow and pressure. They are: firstly, the adductive tension of the interarytenoid and lateral cricoarytenoid muscles; secondly, the longitudinal tension of the vocal folds achieved by the vocalis and cricothyroid muscles; and thirdly, the medial compression achieved by the interaction of tension in the interarytenoid, lateral cricoarytenoid, and lateral parts of the thyroarytenoid muscles.

i) modal voice

The laryngeal characteristics of the neutral mode of phonation, modal voice, are reasonably well agreed. Catford (1964) says that the full glottis is involved 'both ligamental and cartilaginous, functioning as a single unit' (Catford, 1964: 32). He specifies the following details:

'Periodic vibration of the vocal folds under pressure from below (....). For normal voice the liminal pressure-drop across the glottis is of the order of 3 cm of water. Rates of flow vary according to types of the voice ('registers'): for chest voice at about 100 cps the liminal rate of flow is about 5 cl/sec, maximal about 23 cl/sec. These are mean flow-rates: during the open phase of vocal fold vibration flow-rates much in excess of these must occur, and since
the glottal area is small the general aerodynamic picture is of a series of high-velocity jets shot into the pharynx (Catford, 1964: 31).

I don't propose in this outline of phonation types used in voice quality to go into the details of the different registers claimed to exist, but a thoroughly comprehensive account would have to include them. The difficulty of giving such an account is that there are almost as many different terms and specifications as there are writers on the subject: Mörner, Fransson and Fant (1963: 18), interpreting 'register' to refer to 'voice pitch levels'), say that the area 'suffers from an abundance of terms and an ambiguity of their use', and they list 107 different labels.

I shall make the assumption in what follows that modal voice corresponds to the chest voice register, to the extent that different workers seem to agree on aspects of its production.

Van den Berg specifies the physiological aspects of modal voice (i.e. chest voice register) as follows:

'This register is characterized by large amplitudes of the vocal folds at low pitches. This requires small passive longitudinal tensions in the vocal ligaments. The minimal values of the interarytenoid contraction and radial compression are small. The vocal folds are short and thick. An increase of the active longitudinal tension in the vocalis muscle increases the pitch. Contraction of the cricothyroid muscles increases the pitch, but, when the passive longitudinal tension in the vocal ligaments increases beyond a rather small value, the vibrations either cease, at a small medial compression, or transit suddenly to the falsetto type, at a sufficiently

By 'passive longitudinal tension' Van den Berg means the tension exercised on the vocal ligament by the action of the strong cricothyroid muscle; such passive tension in the vocal ligament 'can be increased far beyond the maximal active tension in a contracting muscle, on account of the fact that the forces supplied by the comparatively thick cricothyroid muscles are exerted upon the comparatively thin vocal ligaments' (Van den Berg, 1908: 295-296).

The production of modal voice is thus carried out with only moderate adductive tension and medial compression, and with moderate longitudinal tension when the fundamental frequency is in the lower part of the range used in ordinary conversation. The vibration of the larynx in this condition is regularly periodic, efficient in producing vibration, and without audible friction brought on by incomplete closure of the glottis.

Having outlined the characteristics of modal voice, we can now move on to the description of other phonatory settings, and their various possibilities of combination.

Each of the major phonation types will be discussed individually, but there are two criteria by which they can be grouped into categories, and the classification of the different phonation types into these categories reveals some useful generalities which might be lost sight of in a simple, sequential listing. The description of the different phonation types will be prefaced, therefore, by an outline of the classification.

The two criteria of classification are firstly, 'Can the
phonation type occur alone, as a simple type?, and secondly, 'can the phonation type occur in combination with other phonation types, as a compound type, and if so, with which?'.

On this basis, there are three different categories involved.

The first category is made up of modal voice and falsetto. The qualification for membership of this category is that they can each occur alone, as simple types, and can individually combine with members of other groups, as compound types, but not with each other.

The second category consists of whisper and creak. These can occur alone, as simple types, and together as a compound type, to give whispery creak. They can also occur as compound types with either member of the first group, giving whispery voice and whispery falsetto, and creaky voice and creaky falsetto; and they can occur as compound types with members of the first group and with each other, giving whispery creaky voice and whispery creaky falsetto.

The third category is formed by modificatory settings which can only occur in compound types of phonation, and never by themselves as simple types. These are the local laryngeal contributions to tense voice and lax voice, which are the settings of overall muscular tension throughout the vocal system. Tense voice and lax voice, discussed in detail in the next section of this chapter, can be seen analytically either as constellations of co-occurring local settings in different parts of the vocal system, or as single overall states of muscle tension. It is convenient to isolate the local laryngeal components of these settings, as members of the third category of phonation types, leaving their contribution to overall tension settings
until the next section. To avoid confusion, the terms tense voice and lax voice will be reserved for the overall tension settings and the mode of phonation associated with them will be called harshness and breathiness respectively.

Harshness can combine with modal voice and with falsetto, to produce harsh voice and harsh falsetto. Breathiness can only combine with modal voice, to give breathy voice; the reasons for the incompatibility of falsetto and breathiness will be explored later. Harshness and breathiness cannot combine with the phonation types from the second category, whisper and creak and whispery creak, unless there is a member from the first category also present (this constraint will also be discussed later). Because the incompatibility of breathiness and falsetto still applies, the only multiple compound products of this third category of settings are thus harsh whispery voice, harsh whispery falsetto, harsh creaky voice, harsh creaky falsetto, harsh whispery creaky voice, and harsh whispery creaky falsetto.

There is one sub-category that should be mentioned, and it is a sub-category of the harshness setting. This is the setting where the ventricular folds become involved in the phonation of the true vocal folds by squeezing the ventricle of Morgagni closed and pressing down on the true vocal folds, with the effect that the true and the ventricular folds combine to vibrate as more massive, composite elements. In order to bring the ventricular folds to this position, a high degree of muscular tension is needed, and the effect is normally to make phonation auditorily very harsh. It would be possible to take this ventricular involvement into the descriptive system by means of
a scalar adjective added to the 'harsh' label, such as 'severe harshness'. It is preferable, however, to incorporate the physiologically explicit label 'ventricular' into the descriptive system. Ventricular voice is thus a phonation type which has the same combinatorial restrictions as harsh voice.

ii) falsetto

Hollien (1972: 329) suggests that modal voice and falsetto are 'completely different laryngeal operations'. Earlier, modal voice was described as having moderate adductive tension, moderate medial compression, and moderate longitudinal tension. Falsetto is different in all three respects. Van den Berg (1968: 298) states that adductive tension of the interarytenoid muscles is high, medial compression of the glottis is large, and longitudinal passive tension of the vocal ligaments is also high (though there is little active longitudinal tension in the vocalis muscles).

There is a reasonably wide measure of agreement in recent accounts of the laryngeal mechanisms responsible for the production of falsetto. The summary account below of the physiology of falsetto is based on a number of sources: Chiba and Kajiyama, 1958; Hollien, 1972; Hollien and Colton, 1969; Judson and Weaver, 1942; Kaplan, 1960; Luchsinger and Arnold, 1965; Rubin and Hirt, 1960; Van den Berg, 1968; Van Riper and Irwin, 1958; and Zemlin, 1964.

The consensus of these sources describes the production of falsetto as follows. The arytenoid cartilages adduct the vocal folds, by contraction of the interarytenoid and the posterior and lateral cricoarytenoid muscles. The vocalis muscles along the glottal edge of each vocal fold remain relaxed, but the mass of each vocal fold
is made stiff and immobile by contraction of the thyroarytenoid muscles. The vocal ligaments along the glottal edge of the vocal folds are put under strong tension by the contraction of the cricothyroid muscle. This results in the vertical cross-section of the edges of the vocal folds becoming thin. The glottis often remains slightly apart, and the characteristic sub-glottal air pressure is lower than for modal voice (Van Riper and Irwin, 1958, 228; Kunze, 1964). Van Riper and Irwin suggest here that with the vocalis muscles relaxed and only the thin margins of the vocal folds participating in phonatory vibration, the expenditure of air is bound to be reduced. They cite Trojan (1952) as showing that oxygen consumption is decreased in falsetto voice, compared with modal voice.

The finding that the glottis often remains slightly open has prompted a number of writers to suggest that falsetto voice is usually accompanied by 'friction noises' (Judson and Weaver, 1942: 74) or 'breathiness' (Zemlin, 1964: 155). Given that the width of opening is small, this fricative component is much more likely to be of the whispery rather than the breathy sort. This position is reinforced by Van den Berg's statement that the forces giving medial compression of the folds (i.e. a compressive tendency at right angles to the front-to-back axis of the glottis) are strong (Van den Berg, 1963: 298). On the other hand, the whispery effect may be only slight, if transglottal airflow is small. Chiba and Kajiyan-a (1953: 28) report a finding that supports this:

'It is found that the edges of the vocal chords remain covered here and there with small lumps of mucus, which means that the air is not exhaled abruptly. (In
the chest register, especially in 'sharp voice',
the small lumps of mucus on the edges of the vocal chords
are blown away as soon as the voice starts'.

(Chiba and Kajiyama's 'sharp voice' is discussed in the next major
section, on overall tension settings).

Falsetto is characterized acoustically by two factors: the first
is that the fundamental frequency tends to be considerably higher
than in modal voice. The pitch-control mechanism is different from
that in modal voice; Van den Berg (1968: 298) writes that

'In chest voice the passive tension in the vocal
ligaments needs to remain small when the active
tension in the vocalis muscles is increased to attain the
highest pitches. In falsetto voice, however, the active
tension in the vocalis muscles needs to remain small when
the passive tension in the vocal ligaments is increased
(by the cricothyroid muscle J.L.) to attain the highest
pitches. The registers overlap in the region of medium
pitches'.

Hollien and Michel (1968: 602) found that the average pitch-range
for male falsetto was 275-634 c/s, as against the average range for
modal voice, which was 94-287 c/s.

The second acoustic characteristic derives from the interaction
of high fundamental frequency and the mode of vibration of the vocal
folds. Zenlin (1964: 155) writes that

'High speed motion pictures of the larynx during
falsetto production reveal that the folds vibrate and
come into contact only at the free borders, and that
the remainder of the folds remain relatively fixed.
Further, the folds appear long, stiff and very thin
along the edges. (....). The quality of tone produced
by falsetto is almost flute-like in nature. This is partly due to the rather simple form of vibration executed by the vocal folds, and partly due to the high rate of vibration. When the fundamental frequency is very high, the harmonically related overtones are widely separated in frequency, and consequently in any given frequency range there will be fewer components in the sound produced than there is in a voice with a lower fundamental frequency. This partly accounts for the rich quality of the bass voice when compared with the 'thin' quality of the tenor voice.

iii) whisper

The physiology of the whisper setting of the phonatory mechanism is not controversial. Nearly all writers agree that the chief physiological characteristic of whisper is a triangular opening of the cartilaginous glottis, comprising about a third of the full length of the glottis (Pressman, 1942). The shape of the glottis in whispering is often referred to as an inverted letter 'I' (Luchsinger and Arnold, 1965: 119). In weak whisper, the triangular opening can be fairly long, including part of the ligamental as well as the cartilaginous glottis. With increasing intensity, the glottis is increasingly constricted until only the cartilaginous section remains just open. Taken together, these factors suggest low adductive tension, and high medial compression.

The triangular opening of the glottis is achieved by the following factors: the lateral cricoarytenoid muscles contract, 'toe-ing in' the vocal processes of the arytenoid cartilages (Zemlin, 1964: 169). The muscles which normally approximate the bodies of the arytenoid, the interarytenoid and posterior cricoarytenoid muscles, remain relaxed (Heffner, 1950: 20). This allows the toe-ing in of the
arytenoids as they pivot on the cricoid. As the air flows past the edges of the open cartilaginous glottis, the characteristic 'whisper' sound quality is produced by 'eddies generated by friction of the air in and above the larynx' (Van den Berg, 1968: 297).

The whisper setting is a very uneconomical use of airflow (Luchsinger and Arnold, 1965: 119; Zemlin, 1964: 169). Catford describes the aerodynamic and acoustic aspects of whisper in the following terms:

'Glottis constricted (estimated area, from the smallest possible chink up to about 25% of maximal glottal area). Critical rate of flow about 2.5 cl/sec, estimated critical velocity about 1900 cm/sec. Maximum rate of flow about 500 cl/sec. Turbulent flow, with production of high-velocity jet into pharynx. Acoustic spectrum similar to breath but with considerably more concentration of acoustic energy into formant-like bands. Auditory effect: a relatively 'rich' hushing sound' (Catford, 1964: 31).

iv) creak

Creak is also called vocal fry or glottal fry in the phonetic literature, particularly by American researchers.

Catford gives the following details:

'Low frequency (down to about 40 cps) periodic vibration of a small section of the vocal folds. Mean rates of flow very low - of the order of 1.25 to 2 cl/sec. The precise physiological mechanism of creak is unknown, but only a very small section of the ligamental glottis, near the thyroid end, is involved. The auditory effect is of a rapid series of taps, like a stick being run along a railing' (Catford, 1964: 32).
The low fundamental frequency of the creak type of phonation is one factor that distinguishes it from harsh voice, which is otherwise somewhat similar. While the mean fundamental frequency for creak has been found to be 34.6 cs, in an average range for male speakers of 24-52 cs, the mean fundamental frequency for harsh voice is said to be 122.1 cs, with a range similar to that of modal voice, whose average range is 94-287 cs, according to Michel and Hollien (1968), and Michel (1964). Michel (1968) also reports that harsh voices seem to have fundamental frequencies consistently above 100 cs, and vocal fry (i.e. creak) consistently below 100 cs. It is fundamental frequency characteristics of this sort that led Hollien, Moore, Wendahl and Michel (1966: 246) to suggest that vocal fry 'is best described as a phonational register occurring at frequencies below those of the modal register'.

The specification of fundamental frequency characteristics is not, of course, enough. Factors contributing to auditory quality are also needed. These are suggested by Hollien, Moore, Wendahl and Michel (1966), as follows:

1. The vocal folds when adducted are relatively thick and apparently compressed, 2. the ventricular folds are somewhat adducted also, and 3. the inferior surfaces of the false folds actually come in contact with the superior surfaces of the true vocal folds. Thus, an unusually thick, compact (but not necessarily tense) structure is created prior to the initiation of phonation. Under these conditions it might be expected that the false vocal folds would vibrate in synchrony with the true folds. However, since there is no evidence to support this conjecture, it is possible that their position is either the incidental result of basic laryngeal adjustments or
serves to produce a damping of the vocal fold movement. It would be predicted also that vibration is initiated and maintained by relatively low subglottal pressures; that airflow, if measured, would be considerably less than for most other phonational events' (op.cit. p.247).

A number of experimental studies support their hypotheses. Von Fónagy (1962) investigated the influence of affective states on the mode of phonation, by means of laryngoscopy, tomography and asymmetrical radiography, and described what he called the 'creaky' voice of 'suppressed rage' as having ventricular folds pressed hard against each other, the ventricle of Morgagni wrinkled, the vocal folds held tightly together and the air column vertically through the larynx narrowed to a line, - all matching the picture suggested by Hollien and his co-workers of phonation with strong adductive tension and medial compression, but little longitudinal tension, and with vigorous ventricular involvement.

More recent work by Hollien, Damsté and Hurry (1969), has provided some support also, in their finding that the control of fundamental frequency in vocal fry is not achieved by the same mechanism as in modal voice; while vocal fold length in modal voice increases with fundamental frequency, and vocal fold thickness is inversely related to the frequency (Hollien and Michel,1968), in vocal fry neither the length nor the thickness of the vocal folds seems to vary with changes in pitch. This suggests that control of fundamental frequency is managed by the aerodynamic component of the aerodynamic-myoelastic phonatory action, rather than the myoelastic component, and the sub-glottal air pressure characteristics should reflect this.
This has not yet been established experimentally, although Hurry and Brown (1971) have shown that, consistent with the hypothesis of Hollien et al. mentioned above, the overall sub-glottal air pressure in vocal fry is 'always less than that for the modal phonation' (Hurry and Brown, 1971: 446); McGlone (1967) and Murray (1969) have found lower airflow values for vocal fry than for modal voice.

A different aspect of creak, or vocal fry, is the 'auditory effect (.....) of a rapid series of taps, like a stick being run along a railing' that Catford (1964) mentions in the passage quoted at the beginning of this section on creak. The effect of continual, separate taps in rapid sequence is an essential part of the characteristic auditory quality of creak. Hollien and Wendahl (1968: 506) have described the acoustic correlate of this effect as 'a train of discrete excitations or pulses produced by the larynx', using 'pulse' to mean 'any of a variety of glottal waveforms of brief duration separated by varying periods of no excitation'. In this connection, Wendahl, Moore and Hollien (1963: 254) have suggested that 'the primary criterion which must be met in order for the signal to be perceived as vocal fry is that the vocal tract be highly damped between glottal excitations'. They also note that the fundamental frequency of vocal fry can vary between 20 and 90 cs, and still be heard as vocal fry provided that the vocal tract is nearly completely damped in between the occurrence of successive wave-fronts. Coleman (1963) specifies that vocal fry is perceived whenever the vocal tract wave is allowed to decay by 42 to 44 db of its maximum amplitude for a single pulse, and when the wave is allowed to decay for only 30 db between the excitation pulses, vocal fry is not perceived. This criterion also
allows for the possibility not only of trains of single, discrete pulses, but also for a sub-category of vocal fry where there is 'a vibratory pattern in which the vocal cords separate twice in quick succession and then approximate firmly in a relatively long closed phase'; this was discovered in an investigation using very high-speed cinefilm of the vibrating glottis, and which was given the name 'dicrotic dysphonia' by the investigators (Moore and von Leden 1958: 235). The possibility of triple-pulse trains in vocal fry has also been suggested (Hollien and Wendahl 1968: 509).

This perceptual necessity of damping the laryngeal pulses gives quite strong plausibility to the suggestion by Hollien, Moore, Wendahl and Michel (1966: 247) noted earlier, that the function of the ventricular folds coming into contact with the surfaces of the vocal folds may be to damp the vocal folds' movements, with a corresponding acoustic damping effect on successive glottal pulses.

This damped aspect of vocal fry also lends credibility to the comment reported in the section on velopharyngeal settings, made by Van Riper and Irwin (1958: 244) to the effect that they 'suspect that some of what has been termed "nasal twang" is merely the presence of glottal fry'. The validity of their remark resides in the fact that nasality shares what is arguably one of its most important characteristics with vocal fry, - that of a damped vocal system.

v) 

It will be recalled that 'harshness' is the quality taken on by a number of other phonation types in the laryngeal contribution to tense voice, one of the two settings of overall muscular tension through-
out the vocal system. It will be discussed here as a modification of modal voice, for convenience of exposition. Applied to modal voice, harshness should be thought of not as contributing substantially new parameters to the mode of phonation, but rather as boosting the values of some of the parameters already operating in modal phonation. We shall return to this in a moment.

A number of writers have given auditory descriptions of the quality associated with harsh voice. Sherman and Linke (1952) call it 'an unpleasant, rough, rasping sound'; Holmes (1932) says it is 'a raucous voice quality'; Hillesen (1957) writes that harsh voice is a 'rasping sound associated with excessive approximation of the vocal folds'; Van Riper (1951) calls it 'strident'. The widely used label for this quality, 'harsh', seems well-chosen.

The acoustic characteristics of harsh voice are concerned chiefly with irregularity of the glottal waveform and spectral noise. In the earlier chapter on the history of voice quality analysis it was noted that Fairbanks (1960) said that 'Irregular, aperiodic noise in the vocal fold spectrum is the distinguishing feature of harshness'. Michel (1964) also says that 'harsh voices are characterized by aperiodicity or noise in the spectrum, a normal fundamental frequency level and larger than normal perturbations about the mean fundamental frequency'. The view that small cycle to cycle variations in fundamental frequency are associated with voices judged to be harsh is supported by Wendahl (1963), Moore (1962), Thompson (1962) and Coleman (1960). Wendahl (1964) carried out an acoustic analysis of the role of amplitude variations of the laryngeal waveform in voices judged to be harsh, and found that successive wavefronts tended to be
of unequal amplitude. Using LADIC, an electric laryngeal analogue for producing synthetic waveforms, he established that these characteristic amplitude irregularities made a significant contribution to the perception of harsh 'roughness' (Wendahl, 1964).

The predominant characteristic is the aperiodicity of the fundamental frequency, which is heard as a component of auditory quality rather than of auditory pitch. This aperiodicity has been referred to as a pitch 'jitter' (Cooper, Peterson and Fahringer, 1957: 183). Listeners are very sensitive to even very small amounts of such jitter. Wendahl (1963) used LADIC in his investigations of laryngeal wave-form irregularity to establish the contribution of pitch jitter to harshness. He presented listeners with synthetic stimuli which varied in the magnitude of frequency differences between successive cycles. For each of two median fundamental frequencies, a 100 and a 200 cps condition, stimuli were generated to have frequency variations on successive cycles of +/- 10 cps, 8 cps, 6 cps, 4 cps, 2 cps, and 1 cps. 535 listeners judged the stimuli on the basis of which sounded the most rough. The results show that even very slight frequency variations, as little as +/- 1 cps around a median fundamental frequency of 100 cps, sounded rough' (Wendahl, 1963: 2118).

Wendahl also showed that greater auditory roughness was related to greater deviations from the fundamental frequency, and that the same absolute amount of deviation sounded less rough when superimposed on a fundamental of higher frequency; so if the 100 cps median frequency were taken to represent a male voice and the 200 cps median a female one, then the same deviation, say +/- 5 cps, would make the male voice
sound rougher (op. cit. pp. 248-249). The same principle underlies Hess's finding that harshness in higher-pitched voices is judged as less severe than in lower-pitched ones (Hess, 1959).

Coleman and Wendahl (1967) also used LADIC to make synthetic stimuli in an experiment investigating the relationship between the proportion of pitch-jitter present in a stimulus and the degree of perceived roughness. They found that:

'As the relative duration of jitter elements within a signal is increased, listeners will evaluate the signal as increasing in roughness. It makes little, if any, difference to listeners whether jitter segments occur at the beginning or end of stimuli. A large jitter signal of short duration may be judged to be less rough than a small jitter signal of less jitter excursion (i.e. less degree of frequency deviation from the median J.L.) but having longer duration within a stimulus' (Coleman and Wendahl, 1967: 92).

The relevance of this last study, which will be commented on in more detail in a moment, lies in the relationship between segmental pronunciation and the everyday perception of voice quality. A number of researchers have found that the judged severity of harshness is correlated with some variables of segmental articulation. Rees (1958), for example, found that harshness on vowels was judged to increase with the openness of the vowel; to be greater when the vowel is in a voiced environment; and more marked on vowels in isolation when initiated with a glottal stop than with a 'soft', 'aspirated' beginning. In connection with this last comment, Craig and Sokolowsky (1945) said that excessive and continuous use of a 'glottal attack' on vowels gives a person's entire speech a characteristic harsh
quality. Van Riper and Irwin (1958: 232) agree with this, when discussing harshness as a functional disorder of the voice: 'Very characteristic of this disorder, (....) is the manner of vocal attack. Glottal catches and stops are common. Vocalization is sudden'.

Sherman and Linke (1952) suggest that harshness is judged to be more severe with greater duration of the harsh utterance. Linke (1953) came to the same conclusion as Rees (1958) when she showed that high vowels are judged as less harsh than open vowels; she also showed that, as one would predict, lax vowels are less harsh than their tense counterparts.

We come back now to consider the study by Coleman and Wendahl (1967) in the light of the findings by Rees (1958), Sherman and Linke (1952) and Linke (1953). Although Coleman and Wendahl used synthetic stimuli (very necessary for precise control of the stimulus variables), their results are nevertheless of direct relevance to the perception of voice quality in the normal situation of spoken interaction. Combining the results of the study by Coleman and Wendahl with those of Rees, Sherman and Linke, and Linke, we can conclude that harshness in ordinary voices will be of intermittent occurrence, and of variable relative severity depending on the nature, context and duration of the segments involved. The example of the perception of harshness here can be extrapolated to the general perception of voice quality. This is to say that we perceive voice quality by attending to signals of differing duration and auditory prominence, distributed intermittently and irregularly through the stream of speech, (and this has to be done while we are simultaneously scanning the same phonic material for manifestations of coded linguistic and
paralinguistic signals).

One physiological correlate of harshness is widely agreed. It is laryngeal tension, which underlies what Milisen (1957), quoted above, describes as 'excessive approximation of the vocal folds'.

Gray and Wise (1959: 52), for example, say that harshness 'results from overtensions in the throat and neck; it is often if not usually accompanied by hypertensions of the whole body'. Van Riper and Irwin (1958: 232) write that one of the essential characteristics of the speaker with a harsh voice is tension,

"Most of these individuals show marked hypertension both of the (larynx) and of (....) the pharynx. Both the suprahyloids and the infrahyoids tend to be strongly contracted, as palpation will demonstrate'.

They quote Russell (1936), to the effect that

'As the voice begins to become strident and blatant, one sees the red-surfaced muscles which lie above the vocal cords begin to form a tense channel'.

They add that

'Most harsh voices are relatively low in pitch, with the average pitch level close to the bottom of the range. The intensity appears louder than in the normal voice, though some of the apparent loudness may come from resonation effects due to the tenseness of the oral and pharyngeal cavities' (Van Riper and Irwin, 1958: 232).

Zemlin (1964) agrees with Van Riper and Irwin, when he writes that

'the distinguishing feature which differentiates the normal from the harsh voice is aperiodic noise of aperiodic vocal
fold vibration. Such vocal fold vibration may well be due to excessive tension in the folds. Support for this line of reasoning comes from the fact that persons with harsh voices tend to initiate phonation with glottal attacks. There is some evidence to suggest that persons with harsh voice quality are phonating at an inappropriate pitch level, usually slightly low for their vocal mechanisms. (Zezulin, 1964: 165).

Kaplan (1960) is quite clear about the responsibility of laryngeal tension for harshness. He says that

'Where the folds are drawn too tightly together during phonation rather than being lax, a shrill, harsh, creaking noise, which is called stridency, or stridor, enters the tone. An obstruction of some type is present. Some causes include general tension, spastic paralysis, or often a throat strain or "pinched throat". There is excessive constriction of muscles all through the vocal tract, and the tension is great in the external laryngeal muscles. The vibrations of the vocal folds are hindered, and supraglottal friction noises are introduced' (Kaplan, 1960: 167-168).

Accepting laryngeal tension as established, the question arises of which type of tension, in the categories set up earlier of adductive tension, medial compression and longitudinal tension, is involved.

Kichel (1968) was reported, in the earlier section on creak, as stating that while vocal fry (creak) was characterized by fundamental frequencies consistently below 100 cs, harshness showed fundamental (but not markedly) frequencies consistently/above 100 cs. This strongly suggests that the laryngeal tension in harshness is not chiefly longitudinal tension, the main mechanism in modal voice for controlling the frequency of vibration of the vocal folds. From the comments noted above, such
as 'excessive approximation of the vocal folds', and 'the folds are drawn too tightly together', it seems reasonable to conclude that the exaggerated laryngeal tension in harsh voice is a combination of extreme adductive tension and medial compression, brought about by over-contraction of the muscle systems responsible for these two parameters in modal voice. This is supported by Brackett (1940), cited by Van Riper and Irwin (1958: 232), who describes the inflammation of the vocal folds which results from their traumatic abuse by the deliberate, experimental production of harsh voice.

Earlier in this chapter, it was suggested that when harshness becomes very severe, the ventricular folds become involved in phonation, pressing down on the upper surface of the true vocal folds. Ventricular voice was offered as a physiologically more explicit synonym for severely harsh voice. It may be helpful to end this section on harshness with some brief comments about this mode of phonation.

Van den Berg (1955) says that 'harsh, metallic voice is made (. . .) when the ventricular folds withdraw into the adjacent tissue, leaving almost no space in the ventricles'. He then goes on to discuss the way that this setting of the larynx serves to boost the relative amplitude of the higher harmonics: spectral features of this sort are summarized in the next section, on overall tension settings of the vocal system, and will not be considered further at this point.

Frederickson and Ward (1962), in an article about the possibilities of damaging the larynx by strenuous muscular exertion, say that in pronounced physical effort, 'the true cords are no longer completely approximated, while the false cords remain competent. Under these
circumstances the full force of the intralaryngeal pressure is exerted at the ventricular level'. Ventricular voice can be visualized, then, as phonation at extreme effort, with a fine degree of control over the audible quality made impossible by the comparatively large muscular forces exerted. Plotkin (1964) says that ventricular voice, 'once heard is never forgotten', and that its 'characteristic deep, hoarse voice, alike in male and female, causes an almost sympathetic tightening of the listener's throat'. Freud (1962), quoted by Aronson (1964), gives a rather similar picture, of 'ventricular dysphonia', where he

'depicts it as a total, tight, spastic apposition of the constrictors of the larynx and hypopharynx, giving the voice a groaning, animal quality and suggesting to the listener the exertion of extreme effort. The words sound as if they are being chopped off' (Aronson, 1964: 369).

These descriptions by Frederickson and Ward, and by Freud, are of course descriptions of voices which have been classed as 'dysphonic', and it doesn't follow that all voices that a phonetic description would classify as examples of ventricular voice would use ventricular phonatory effort of quite such extreme degrees. One can hear voices which make use of some phonatory contribution by the ventricular folds fairly commonly in everyday life. I believe that Sweet was referring to what, in the terms of this thesis, would be called ventricular voice, or possibly whispery ventricular voice, when he wrote that

'Narrowing of the bronchial passages gives a wheezy character to the voice, sometimes approaching to strangulation. This effect is sometimes known as "the pig's whistle". It may be heard from Scotchmen (sic), and
combined with high key gives the pronunciation of the Saxon Germans its peculiarly harsh character' (Sweet, 1906: 73).

His comment about 'narrowing of the bronchial passages' is a revision (which he should have rejected) of a wording in the section on voice quality in his earlier *Handbook of Phonetics* (1877: 97-99) which read 'Narrowing of the upper glottis (....) gives an effect of strangulation'.

vi) **breathiness**

'Breathiness' is the quality which is often taken on by modal voice, giving breathy voice, as the laryngeal contribution to lax voice, the other setting (with tense voice) of a relative degree of muscular tension throughout the vocal system.

Catford (1964: 32) describes the characteristics of breathy voice as a

'Combination of breath + voice: glottis relatively wide open: turbulent airflow as for 'breath' plus vibration of the vocal folds. The vocal folds do not meet at the centre line: they simply "flap in the breeze". Auditory effect, "sigh-like" mixture of breath and voice: one form of voiced [h] (sic)'.

The muscle tension adjustments necessary for breathy voice can thus be seen as minimal adductive tension and weak medial compression, just enough to allow aerodynamic forces in the large volume of transglottal airflow to superimpose a very inefficient vibration of the vocal folds, 'not meeting at the centre line', on the outflowing air. The one laryngeal tension factor that is controlled more finely
is longitudinal tension, in the production of appropriate variations of fundamental frequency for the purposes of intonation. We can assume that the degree of longitudinal tension is rather low, generally. High pitched breathy voices seem rare.

Breathiness can combine with only one other type of phonation, in the system of describing voice quality offered here: that is, modal voice. This is because, while modal voice requires only moderate medial compression, all the others, falsetto, whisper, creak, harshness and ventricular voice, need a greater amount than is compatible with breathiness.

For the 'breath' component of breathy voice, Catford (1964: 30) specifies that the glottis should be

- widely open (estimated area of glottis about 60% to 95% of maximal glottal area). Critical rate of airflow about 25 cl/sec, maximum about 890 cl/sec, estimated critical velocity about 240 cm/sec.
- Diffuse low-velocity turbulence. Acoustic spectrum: a "hushing" noise of rather "thin" quality (as compared with "whisper")(...).
- With the same configuration of the glottis, but with rates of flow below the critical value of about 25 cl/sec, airflow through the glottis is non-turbulent and consequently silent.

Many writers have used the label 'breathy' to describe components in given voice qualities that should rather have been called 'whispeiry'. In the descriptive scheme used here it would not be possible, for example, to accept a label which combined 'breathiness' and 'harshness' for the voice quality often described as 'husky' or 'hoarse', because of the mutually-exclusive pre-requisites of breathiness as here
defined and harshness. Such a quality would instead be labelled 'harsh whispery voice'. However, it doesn't seem unreasonable to suppose that the transition from breathiness to whisperyness is part of an auditory continuum. The placing of the border-line between the two categories is then merely an operational decision. I choose to set the boundary well over towards the 'breathy' end of the scale, because I think that in whispery voice the fricative component can be sub-divided into a larger number of auditory increments than can the breathy contribution to breathy voice. This, however, is only a subjective impression, and the question is amenable to experimental investigation, if Catford's values for the aerodynamic factors are taken as a point of departure. His figures show 'breath' and 'whisper' as two very different phenomena, if they are taken to apply to the end-points of the breathy-whispery continuum: comparing maximum airflow rates, breath has almost twice that of whisper (890:500 cl/sec); for critical velocity, breath has a far lower value than whisper (240:1900 cm/sec); and the maximum glottal area is much greater for breath than for whisper (95:25%). All these are relatively easily accessible parameters for measurement, and a definition of breathiness and whisperyness, derived from an investigation of this sort, and finally given on arbitrary but explicitly quantified grounds, would be helpful.

Some writers simply collapse the two phenomena. Kaplan (1960:167), for example, describes 'breathiness' as a voice quality which 'is said to have an aspirate quality, and the effect is as though a whisper were added to the normal tone'. In this and in other cases, although 'whisper' is a term in the writers' descriptive repertoire
for voice quality, it is not used for the description of a compound
type of phonation with (modal) voice or any other type, and 'breathy
voice' is the label used for any quality where there is a fricative
escape of air during phonation. Zemlin (1964) writes, for example,
that

'The most common correlate of Breathiness is a
persistent glottal chink in the posterior portion
of the vocal folds. Critical examination of a large
number of larynxes reveals that a good many persons with
apparently healthy, normal sounding voices display a
glottal chink in the area of the arytenoid cartilages.
We can suppose that there is a point at which the mag-
nitude of glottal chink will result in a breathy voice
quality. The exact relationship between magnitude of
glottal chink and vocal quality is not well understood'
(Zemlin, 1964: 165).

In the situation that Zemlin describes, it would seem likely that as
the 'glottal chink' grew in size, whispery voice would set in first,
and that it would have to enlarge to a much greater proportion of the
total glottal area before breathy voice as described by Catford was
heard.

With the considerable amount of air that is wasted in breathy
voice, there is an inverse relationship between intensity of the voice
and breathiness (Pronovost, 1942). Some of the acoustic energy
would also be lost by the damping effect of the general relaxation
of the muscles of the whole vocal system in lax voice, of which
breathy voice is almost always a component. This damping effect is
discussed in the next section on lax settings of the vocal system.
Breathy voice, however, contributes its own damping effect to the
general energy loss. Fant (1972: 50) points out that the broadening of the bandwidth of the first formant in lax voice can be partly attributed to the high damping effect of 'weak, breathy voice' on the rest of the system. Bandwidth characteristics in tense and lax voice are discussed in more detail in the later section on tension settings of the vocal system.

b) Compound phonation types

A number of combinatorial constraints on the co-occurrence of individual phonation types to form a compound type have already been mentioned. In this section, we shall look briefly at some of the factors underlying the compatibility, or the lack of it, between different individual phonation types, with respect to their potential co-occurrence.

There are two general conditions under which compatibility between phonation types is possible.

The first condition is where the individual settings apply to different parts of the laryngeal structure, so that competition for the same vocal apparatus is avoided.

The second condition is where the same part of the laryngeal apparatus is concerned in the production of two different phonation types, but where the vibratory patterns of the two settings modify each other without either changing substantially enough to lose its auditory identifiability. No implication is asserted here for primary status for one setting and secondary for the other.

Examples of the first condition for compatibility, where different
laryngeal locations are involved, would be whispery creak, whispery voice, and whispery falsetto. The whisper component is assumed here to be produced in a triangular gap between the arytenoid cartilages, in all three compound phonations, and creak made separately at the thyroid end of the glottis, with both modal voice and falsetto being limited to the ligamental section of the glottis. The triple compound types whispery creaky voice and whispery creaky falsetto would be further instances.

Examples of the second condition for compatibility, where two (or more) vibratory patterns modify each other, would be all the instances of harshness in compound phonations. The modification that harshness imposes on all compounds in which it participates is a boost in the parametric values of adductive tension and medial compression to an extreme degree. In this way, the compound type harsh voice is characterized by greater adductive tension and medial compression than the moderate values normally found in modal voice alone. In order to achieve phonation, the sub-glottal pressure has to be given a compensatory boost also, in order to re-assert aerodynamic participation in an aerodynamic-nyoelastic phonatory equation in which the nyoelastic component has had the elastic resistance of the glottis to airflow substantially increased. This doesn't mean that phonation types with low values on these parameters can't combine with harshness. Whisper is a case in point, where low adductive tension is normally one of its characteristics: it does mean though that the whisper component in harsh whispery voice, say, is maintained by a much greater effort on the part of the lateral crico-thyroid muscles to keep the arytenoid triangle open against the vigorous attempt by the interarytenoid
muscles to close it. Alternatively, when the whisper component is achieved not by keeping the cartilaginous glottis apart but by preventing the vocal folds from completely closing, the posterior crico-arytenoid muscles have to exert a considerable effort to prevent the interarytenoid muscles from closing the glottis. The auditory nature of the whisper component is likely to be rather different in such compound phonation, compared with its occurrence as a simple type.

There are two general conditions under which compatibility is not possible between individual phonation types, preventing the occurrence of particular compound phonations.

The first of these conditions is where the pre-requisite actions of the larynx for each of the two types of phonation involved are mutually exclusive.

The second condition is where perceptual factors make it impossible to hear the differences introduced by the addition of one phonation type to the other.

There are a number of examples of the first, physiologically pre-emptive condition. Modal voice and falsetto are one instance. Their membership of the first grouping of phonation types is based partly on this impossibility of co-occurrence. These two types of phonation need quite different types of vibration of the vocal folds, as described earlier in this section, and they therefore cannot combine. Similarly, harshness and breathiness are mutually incompatible, because of their parametric pre-requisites. Where harshness has extremely high adductive tension and medial compression,
breathiness must have very low values of these parameters. Breathiness is a very unusual phonation type in this respect, and is incompatible with almost every other phonation type. With very low adductive tension, and most importantly, very low medial compression, breathiness is compatible with only one other phonation type, — modal voice, giving *breathy voice*. Modal voice has a moderate degree of medial compression, while every other type, falsetto, whisper, creak and harshness, has a high or very high degree. Also, modal voice has only moderate adductive tension, where falsetto, creak and harshness have high or very high degrees. The one exception to this is whisper, which has low adductive tension, — but whisper is the polar value at the other end of an auditory scale from breathiness, so breathiness can't combine with whisper by definition.

Another example, which is covered by both incompatibility conditions, physiological and auditory, is the combination of harshness and whisper, preventing the occurrence of harsh whisper. The physiological incompatibility is on a different basis than that of modal voice and falsetto, or of harshness and breathiness. It is not a matter of directly opposite physiological requirements, but of redundancy. The effect of adding the actions which produce harshness to those producing whisper is merely to boost the tensions pressing the vocal folds and the arytenoid cartilages together. Regardless of whether the type of whisper is one where a gap is left between the arytenoids, or where the ligamental glottis is also kept slightly open, the effect of adding harshness will be to narrow the glottal aperture. This will only result in the audible whisper rising in amplitude until the gap is completely closed. While the whisper lasts, it will be heard as being louder, without a major change in
quality.

The second, auditory, incompatibility condition applies not only to harshness and whisper, but also to modal voice and falsetto, and to harshness and creak.

Harshness and whisper are both characterized acoustically by a factor of aperiodicity. To add the aperiodicity of harshness to that of whisper would be redundant.

The auditory incompatibility of modal voice and falsetto is straightforward: using the same vocal apparatus, they constitute different (though not dissimilar) qualities.

Harshness and creak present a less clear-cut case of auditory incompatibility. It will be recalled that creak is characterized by a certain amount of moment-to-moment variability of its (normally very low) fundamental frequency. To modulate this variability by superimposing the essential variability (aperiodicity) or harshness would not produce a large change in quality. However, the fundamental frequency of creak can rise quite high, up to and above 100 Hz. In the upper part of this range, it might happen that the inherent variability of the fundamental frequency of creak was smaller than the superimposed variability of a harsh component, and to that extent harshness and creak could combine in a compound phonation of harsh creak.

These principles of compatibility and incompatibility are able to give structure to most of the combinations of phonation types that can and cannot occur. But they can only be applied where the physiological mechanisms and the auditory effects are reasonably well
understood or at least plausibly hypothesized. There are some compound types of phonation where the necessary degree of analytic clarity does not yet obtain. Comments about these can therefore only be rather speculative.

For example, the physiological mechanism for the production of creak remains unclear. It has been treated in this description as if it were the product of an independent vibratory system at the forward, thyroid end of the glottis, following suggestions by Catford (1964: 32). The auditory impression of the creak component in compound phonations is sufficiently different in the various compound types to suggest that the auditorily identified phenomena we are willing to call 'creak' may possibly be produced by different mechanisms in different compounds. The creak component in high-pitched creaky falsetto sounds different from that in whispery creak, for example.

There are also physiologically possible phonation types quite outside the descriptive system presented here, omitted because they seem never to be used as habitual voice qualities, whose possibilities of occurrence in compound phonations are not yet well analysed. To give one example, it is possible to produce phonation which sounds like ventricular falsetto, (sometimes referred to as 'seal voice'), by very severe compressive effort of the whole larynx, and the respiratory system. There are also a number of auditorily different kinds of whisper, touched on by Catford (1964: 32-33).

3. Settings of overall muscular tension throughout the vocal system

Discussion so far has been concerned with settings, both of the supralaryngeal tract and of the phonatory mechanism of the larynx, which can suitably be described as specific to some localized part
of the vocal system. We come now to settings of a different order from such local considerations. These are settings of overall degrees of muscular tension which exercise their effect throughout the vocal system. Two categories of tension settings will be discussed, as distinct from the neutral tension specification. These are tense voice and lax voice, which stand for a high degree of tension generally through the system, and a low degree, respectively. It is very difficult to find any absolute measure that can be used to specify the degree of overall muscular tension in the neutral setting, and which could figure legitimately in a general phonetic theory. So a nominal, relative measure is adopted, where the degree of tension in the neutral setting is the one lying midway between the two extremes of maximally tense and maximally lax settings.

There has been a certain amount of discussion in the literature on voice quality about the factors correlated with overall tension settings. The usual contrast is one which is drawn by the use of descriptive impressionistic labels such as 'metallic voice' and 'muffled voice', relating them to high and low degrees of tension. Some other labels for 'metallic' have been 'brassy, bright, clear, keen, piercing, ringing, sharp, strident'. Labels used for 'muffled voice' include 'dull, guttural, mellow, obscure, soft, thick'. Bloomfield (1933: 94-95) uses the terms 'metallic sound' and 'muffled sound', to refer to 'quality of resonance', and says that they have 'not been physiologically analyzed'. In fact, although the impressionistic labels used have been somewhat varied, and most probably cover a range of different degrees of muscular tension, more recent writers seem to agree reasonably well on the physiological, auditory and acoustic factors posited as characterizing such voices.
For Milisen (1957) 'metallic voice' is a 'sharp piercing voice ordinarily associated with tension of the side walls of the oral and pharyngeal cavities', and 'muffled voice' is a 'diffused voice not Projected from the mouth and frequently associated with excessive relaxation of the oral and pharyngeal cavities'.

Greene (1964: 53) says that 'The relaxed muscular walls of the vocal resonators tend to "damp", stop or absorb high frequencies and produce mellow tone, whereas hard or taut muscular walls act as reflectors and produce harsh tone'. Later, considering voice quality in singing, and the relation between singing and nasality, Greene writes that 'Clear, bright ringing tones (voix blanche) are dependent on lifting and tending of the velum. Low mellow tones are achieved with the faucal arches well relaxed, and the soft palate almost pendant' (Greene, 1964: 56).

Chiba and Kajiyama, in their book on the nature and structure of the vowel, published in Japanese in 1941, but which first appeared in English translation in 1958, write that 'soft voice'

'is usually described as a dull, guttural or obscure sound. A voice which contains some noise and has a comparatively strong fundamental and weak higher harmonic partials. It can be most easily produced either when the head is bent upwards or when the larynx is drawn downwards with the head kept in a normal position' (Chiba and Kajiyama, 1958: 17).

'Sharp voice', they say,

'is often described as being keen, ringing, energetic or powerfully penetrating. We can easily produce this voice with our head drawn back while keeping it bent forward' (op.cit., p.18).
It may already be becoming clear that tense voice and lax voice are each manifested by a constellation of more local settings at various points in the vocal system. It is because there is a common factor of general muscle tension underlying the multiple local settings which contribute to each of these two types of voice that it is analytically helpful to consider the effects of different degrees of overall muscular tension in this separate section.

The outline that follows will consider the local effects of tension settings, firstly at the larynx, and then at a number of points along the supralaryngeal tract.

One acoustic finding needs to be mentioned before starting on the discussion of local manifestations of the tension settings. It is that tense and lax voices seem to be acoustically differentiated chiefly by the comparative amounts of energy in the upper harmonics, with tense voice having relatively stronger upper harmonics than lax voice, as noted in the quotations above from Greene (1964: 53) and Chiba and Kajiyama (1958: 17). Van Dusen (1941), in an acoustic study of 'metallic voice', reported that male metallic voice had greater energy in the harmonics above the first six; although she noted that in the female metallic voice the relative concentration of energy was in the low harmonics, with little in the higher ones. This might be explained, tentatively, as being due to a tendency in both male and female tense voices to favour approximately the same part of the spectrum (perhaps about 500 to 1000 c.), taking account of the relatively higher female fundamental frequency and hence more widely separated harmonics, together with the smaller overall acoustic power of female voices. Her finding that male
tense voices have more energy above the first six harmonics implies that lax voice in males has weaker formant amplitudes in this region then tense voice.

Kaplan (1960) relates the acoustic differences between tense and lax voices to characteristics of the pharynx. Referring to Anderson (1942), he says that:

"the role of the pharynx in resonation is more difficult to evaluate than that of the mouth or nose, but (....) it is especially important to provide resonance for the fundamental and the lower overtones. This is said to give the voice a mellow, rich and full quality.

The texture, as well as the size and shape, of the pharynx and its apertures affects speech quality. A hard-surfaced resonator emphasizes the higher partials, or overtones, so that a pharynx tightly constricted by its muscles, takes on a metallic, strident, and tense tone. On the contrary, a soft surface, provided by relaxed throat muscles, increases the responsive range while damping the resonator. This in effect gives relative prominence to the fundamental and lower partials' (Kaplan, 1960: 199).

So one major acoustic variable dependent on the overall tension setting is the system's damping characteristics, or factors of absorption of sound energy, in different parts of the larynx and vocal tract. Formant amplitudes and bandwidths are thus involved, but we shall also see that dynamic configurational properties of the tract play a part, and that therefore formant frequencies are sometimes affected.

Considering now the involvement of the larynx in the production
of tense and lax voice, a number of writers have suggested that some of the control of damping lies with the upper larynx, notably in the ventricles of Morgagni and the ventricular folds. Pepinsky (1942) has claimed that the ventricles may act as a low-pass filter, in normal voices (that is, in the terms of this thesis, voices without a marked degree of tenseness or laxness). Van den Berg (1955) agrees that in normal voice the ventricles behave as a low-pass filter suppressing the higher-frequency components of the glottal waveform, and also says that, in what he calls 'harsh, metallic voice',

'the ventricular bands decline and lay themselves against the wall of the ventricle until they firmly press upon the vocal folds. The ventricle disappears and the higher harmonics are not filtered. This accords with observations on certain singers with a particularly strong voice, who exhibited small ventricles' (Van den Berg, 1955: 63).

Greene (1964: 55) supports this point of view:

'The ventricle of the larynx is also modified by the action of the false folds which follow the movements of the true folds. When relaxed they act as soft surface filters but when constricted they press down upon the folds and obliterate the ventricle. The tension favours high partials and produces the "compressed tone" described by Aikin' (i.e. Aikin 1951).

Chiba and Kajiyama (1958) don't agree completely with Pepinsky, Van den Berg and Greene about the acoustic function of ventricular action in tense and lax voices. They concede that in lax voice ('soft voice'), the 'edges of the false vocal cords are curved upwards and the rimae ventriculorum are wide, while in "sharp voice"
(tense voice) the latter are found to be narrower than in any other type of voice' (Chiba and Kajiyama, 1958: 35). But they believe that the relative adjustments of the false vocal cords are 'caused by the adjustment of the vocal cords', and are 'not designed to change the resonance of the larynx tone in the ventricles of Morgagni; or if there be any such change, the effect is not great' (op. cit., p.36). They prefer to attribute the spectral prominence of the lower harmonics in lax voice ('soft voice') to the following fact:

'In "soft voice" the vocal processes of the arytenoid cartilages are separated and the cartilaginous glottis is kept slightly open. During pronunciation the ligamentous glottis opens and shuts periodically. It opens into the shape of a spindle and shuts completely. But in this case it opens for a considerable time and shuts for a short time. This makes it impossible for air to be emitted abruptly. The result is that the lower harmonic partials are somewhat stronger, and the higher harmonic partials are somewhat weaker, than in any other kind of voice' (op. cit., pp.20-21).

Whichever explanation one accepts, the ventricular hypothesis of Pepinsky, Van den Berg and Greene, or the 'opening quotient' hypothesis of Chiba and Kajiyama, the acoustic results are not in question.

Another involvement of ventricular action can often be found in the production of tense voice. Because of the hypertension of such voices, the phonation type is likely to be either ventricular voice or harsh voice. Notice, for instance, that both Van den Berg (1955: 63) and Greene (1964: 53) include harshness in their comments on what is here called tense voice.

Other laryngeal factors concern the pitch and loudness characteristics of tense and lax voices. These, although not directly aspects
of quality as such, derive from the same underlying cause of greater or less muscular tension than in neutral voices, and their perception is to some extent influenced by the concomitant voice quality, as we shall see.

There is a strong probability that in tense voice the pitch-range will be higher than in lax voice, and the loudness-range louder. It is certainly possible, though, to compensate for these tendencies (just as we noted the possibility of pitch compensation in the earlier discussion of raised and lowered larynx voices). I have the impression that American adult male voices often tend to be tense, harsh, loud, but low-pitched rather than high-pitched. This may be because of the influence of the belief that this type of voice acts as an index to an American cultural stereotype of the virile, dominant male, which in that society is presumably thought to be a socially and psychologically desirable image to try to project.

Leaving these possibilities of compensations of pitch-range aside, it is interesting that even when the pitch-range of tense voice is not in fact high, in physical terms, it may be perceived as being higher than the fundamental frequency by itself justifies, because of spectral characteristics of tense voices: Landes (1953) found that voices with relatively greater energy in the upper part of the spectrum (as is typical of tense voice), tend to be perceived as higher pitched than those with less high-frequency intensity. They may also be perceived as louder: Van Riper and Irwin (1958: 452-453) point out that

'the human ear is much more sensitive to tones in the
frequency range 1,000 to 4,000 than it is to tones below or above the range. Thus (....) voice qualities that have more of their energy in the upper than in the lower frequencies (....) may sound louder than the actual energy reaching the ear would normally justify'.

West, Ansberry and Carr (1957: 72) put this point very appropriately for our purposes:

'Two voices that, physically measured, have equal intensities or energy values may vary greatly in loudness, particularly if one has its energy concentrated in the fundamental tone and the other in a harmonic tone near 1000 vibrations per second, for it is the octaves between C$^1$ and C$^3$ to which the human ear is most sensitive. "Piercing" voices are those whose energy values are in this middle region'.

This lends some support to the tentative suggestion made above (concerning Van Dusen's finding (1941) that male tense voices tended to have most energy in the lower harmonics), that the perceptually greater loudness of these voices arises from their having greatest intensity in the same frequency region, near 1000 c.

In lax voice, the laryngeal hypotension tends to prevent the occurrence of ventricular phonation, and facilitates the occurrence of breathy voice, whose main phonatory characteristic was noted earlier to be incomplete closure of the vocal folds. Chiba and Kajiyama's comment reported above that 'in "soft voice" the cartilaginous glottis is kept slightly open' (1958: 20-21) is relevant here. Lax voice tends to be accompanied by a low pitch-range and a soft loudness-range. The comments and findings cited above in the discussion of tense voice, from Landes (1953), Van Riper and Irwin (1958) and West, Ansberry and Carr (1957), show equally that lax
voices will tend to be perceived as lower-pitched and softer, just as tense voices are heard as higher-pitched and louder, than strictly acoustic measures of fundamental frequency and overall intensity would account for.

We turn now to consider supralaryngeal factors in tension settings. In tense voice, the velopharyngeal musculature tends to be tensed in such a way that the velum closes the velopharyngeal port, with the result that tense voice is less often accompanied by nasality. The remark of Greene (1964: 56) that 'Clear, bright ringing tones (voix blanche) are dependent on lifting and tensing of the velum' supports this. This is not to say, however, that tense voice never has side chamber resonance associated with it. So-called 'nasal twang' may be the popular label for side chamber resonance generated by tense adjustments of the upper larynx, pharynx or faucal pillars, as discussed in the section on velopharyngeal settings. Speculatively, this may well be why a voice quality not infrequently heard in adult male Americans is thought to be nasal, even though resonance of the nasal cavity doesn't seem to be involved; this is a voice which would be classified in the terms used here as tense, harsh, and loud, with side chamber resonance of some other cavity than the nasal cavity. The voice of James Cagney, the American film actor, is an example of this sort. (Voices of film actors are often very clear exemplars of qualities thought to be national stereotypes, for obvious reasons).

There is also the possibility that in tense voice the faucal pillars will be approximated, without any lowering of the braced velum, and with the pharynx constricted and the larynx subjected to
a vertical pull both from below by the infrahyoid complex and from above by the laryngeal elevator complex. The comments by Bell (1908: 19-21), in the section on faucal settings, about faucalization in deaf speakers leading to a quality he described as 'a peculiar metallic ring, somewhat like the tone of a brass musical instrument'; and by Pike (1943: 123-124), associating faucalization with 'lower pharyngeal constriction, glottal tension, and usually a raising of the larynx', are relevant here.

Supralaryngeally, in lax voice, because of the relative relaxation of the musculature of the vocal tract, the local velo-pharyngeal system tends also to be relaxed, and lax voice therefore is frequently accompanied by moderate nasality.

Tense voice and lax voice are correlated with two aspects of lingual settings. The major aspect is to do with the amount of radial movement of the centre of gravity of the tongue from its neutral position, in the midsagittal plane, that characterizes the articulatory gestures of segmental pronunciation. The other, more minor, aspect concerns the shape of the tongue from side to side, in the coronal plane.

Sweet (1877) was one of the earliest writers to relate degrees of overall tension to activities of the tongue, in his use of the two terms 'narrow' and 'wide'. He said that

'The distinction depends on the shape of the tongue. In forming narrow sounds there is a feeling of tenseness in that part of the tongue where the sound is formed, the surface of the tongue being made more convex than in its natural 'wide' shape, in which it is relaxed and
flattened. This convexity of the tongue naturally narrows the passage - whence the name. This narrowing is produced by raising, not the whole body of the tongue, but only that part of it which forms, or helps to form, the sound' (Sweet, 1877: 9).

He made it more clear that he was thinking of the convexity of the surface of the tongue as related to the sagittal plane, and not to the coronal plane as such, in an article in 1911, when he wrote

"In forming narrow sounds there is a feeling of tension in that part of the tongue where the sound is formed, the tongue being clenched or bunched up lengthwise, so as to be more convex than in its relaxed or 'wide' condition. The distinction between narrow and wide (...) generally depends on quantity; length and narrowness, shortness and wideness going together' (Sweet, 1911: 463).

We will return to the relation between tension and segmental duration below, but we can note here that in bunching the tongue along its sagittal axis, there will normally also be an accompanying tendency for the coronal axis of the surface of the tongue to be made more sharply convex. So tense voice, to the extent that susceptible segments reflect increased tension, is likely to have the surface of the tongue more steeply convex in both planes, as well as involving a greater departure of the centre of gravity from its neutral position, for the articulation of susceptible segments, which are raised more than the corresponding 'wide', 'flattened' segments would be in lax voice.

Sweet gives Alexander Melville Bell the credit for being the first to notice the distinction between the two categories that Sweet was to call 'narrow' and 'wide' (Sweet, 1911: 463). Bell's own terms were 'primary' and 'wide'. He wrote that
Primary vowels are those which are most allied to consonants, the voice-channel being expanded only so far as to remove all "fricative" quality. The same organic adjustments form "wide" vowels when the resonance-cavity is enlarged behind the configurative aperture; - the physical cause of "wide" quality being the retraction of the soft palate, and expansion of the pharynx' (Bell, 1867: 71).

But Sweet rejected this explanation of the articulatory actions contributing to the two qualities, and said that because he himself had showed that the distinction depended on the shape of the tongue, he "accordingly substituted "narrow" for Bell's "primary"" (Sweet, 1911: 463).

Heffner (1950: 96-97) discusses some of Sweet's comments about this area, and writes that 'Later scholars have substituted the terms tense and lax for narrow and wide, and recognized that this tenseness applied not merely to the tongue muscles but to the entire articulating complex'. Heffner is here referring to Sievers (1876), though why he refers to him as one of the scholars 'later' than Sweet is difficult to understand. Heffner then goes on to adopt the position that tense and lax(vowels) are differentiated by relative closeness of the glottal opening (in vocal fold vibration) and greater sub-glottal pressure for tense vowels, and relative openness of the glottal opening and less sub-glottal pressure for lax vowels. He suggests that terms such as tense and lax can be retained if their reference is shifted 'from tongue elevations and tongue muscle tensions to laryngeal positions (sc. the maximum degrees of glottal opening in vocal fold vibration J.L.) and air pressures' (Heffner, 1950: 97). So there are various different physiological parameters
posed by the different writers on the topic of segmental tension as a momentary action of the vocal apparatus. Who is right about the basis of segmental tension need not be a primary concern in this study of the quasi-permanent features of voice quality; but what is interesting in the different observations made in this area of segmental tension is that they may all be right, from the point of view of the variety of local manifestations of an overall tension state. Comments by phoneticians about tenseness and laxness as segmental features are thus relevant to the study of voice quality, if we choose to see the various alternative segmental suggestions as possible co-occurring local phenomena which as far as voice quality is concerned all individually arise from a general, single tension state throughout the vocal system.

A summary of hypotheses about the basis for segmental tension can be found in an article by Jakobson and Halle (1964), where their own comments are prefaced by a short history of the discussion about tenseness and laxness as segmental concepts. They explore comments on the topic by Sweet, Bell, Winteler, Sievers, Meyer, Stumpf, Jones, and Fant. Among the variety of articulatory and acoustic correlates which have been suggested, one comment by Jakobson, Fant and Halle (1952: 242) is particularly interesting. They maintain that tense vowels show a greater deformation of the vocal tract, while lax vowels show a smaller deformation of the vocal tract from its 'neutral, central position'. They define the 'neutral position' of the tract as the one assumed by the vocal organs in producing a very open \[ \infty \], which is different from the specification of the neutral position in this thesis, but in this regard not importantly so. Chomsky and Halle (1960: 325) write that
'we note that the difference between tense and lax consonants (... involves) a greater versus a lesser articulatory effort and duration. The greater effort is produced by greater muscular tension in the muscles controlling the shape of the vocal tract.'

The tendency for segmental articulations to depart further from the neutral vocal tract position in a tense voice, and less in a lax voice, may explain some of the auditory implications in the impressionistic labels given at the beginning of this section, — particularly about a tense voice being a 'sharp', 'penetrating' voice, and a lax one being 'dull', 'guttural' or 'obscure', and a 'diffused voice not projected from the mouth'. In this connection also Fairbanks (1960: 170) says that 'muffled voice' (lax voice) 'is a general term for the slighted consonants and neutralized vowels that result from limited or inconclusive movement of the articulatory structures (oral inactivity)'.

We return now to consider the comment quoted from Sweet (1911: 463), to the effect that 'The distinction between narrow and wide (....) generally depends on quantity; length and narrowness, shortness and wideness going together'. Jakobson and Halle (1964: 97) summarize the findings of a number of investigators when they say that

'The heightened subglottal air pressure in the production of tense vowels is indissolubly paired with a longer duration (....) The tense vowels are necessarily lengthened in comparison with the corresponding lax phonemes (sic). Tense vowels have the duration needed for the production of the most clear-cut, optimal vowels; in comparison with them the lax vowels appear as quantitatively and qualitatively reduced, obscured and deflected from their tense counterpart toward the neutral formant pattern'.

This last comment neatly characterizes one aspect of the acoustic differences between tense and lax voice - formant ranges are narrower in lax voice than in tense, because of the tendency of the susceptible vocoid segments to be centralized in lax voice, with less extensive radial movements of the centre of gravity of the tongue away from the neutral configuration, and hence less deviation of the formant frequencies from their neutral values.

Jakobson and Halle conclude their article with a brief summary of their survey:

'In producing lax phonemes the vocal tract exhibits the same behaviour as in generating the cognate tense phonemes but with a significant attenuation. This attenuation manifests itself by a lower air pressure in the cavity, a reduction in the size of the cavity (....), by a smaller deformation of the vocal tract from its neutral, central position, and/or by a more rapid release of the constriction. The tense consonants show primarily a longer time interval spent in a position away from neutral, while the tense vowels not only persevere in such a position optimal for the effectuation of a steady, unfolded, unreduced sound, but also display a greater deformation of the vocal tract' (Jakobson and Halle, 1964: 100).

So Sweet's observation about the correlations between length and 'narrowness', and shortness and 'wideness', is upheld. Fant (1960) also supports this position, and adds a comment on the contribution of muscle tension in the walls of the vocal tract to the acoustic output of the system:

'The effect of the muscular strain traditionally claimed to be associated with the tense stops is to prolong the fricative interval of semi-closure. In the case of the
tense vowels, the muscular strain cannot be expected to affect the damping of the cavity walls and thus influence the formant bandwidths to any significant extent. The tenseness and the longer duration condition on articulation further away from the neutral position, and those formant bandwidths variations that do occur, are due to the varying degrees of opening' (Fant, 1960: 225).

The moment-to-moment variations in formant bandwidth that can be observed during continuous speech are thus to be attributed not to the fluctuations of muscle tension in the walls of the tract, but to the effect of momentary configurational changes of segmental articulation. However, the degree of difference of muscular tension between tense voice and lax voice is likely to be considerably greater than the differences of tension in the speech of a given speaker between tense and lax segments. Accordingly, in tense voice and lax voice, it seems reasonable to link some of the narrowing of formant bandwidths with the reflective effect of tense cavity walls, and some of the broadening with the attenuating effect of lax walls. It was also noted earlier that some of the broadening of formant bandwidths in lax voice is due to the damping effect of the inefficient phonation of 'weak, breathy voice' (Fant, 1972: 50). Momentary changes of bandwidths can still be attributed to articulatory factors differentiating tense and lax segments, rather than to factors of muscle tension. We thus see a situation where the momentary values of formant bandwidths oscillate around a long-term average for the particular speaker. The oscillations can be attributed to the effect of segmental performance, and the long-term average largely to the extrinsic tension setting.
4. Structural and scalar conventions in the construction of descriptive phonetic labels for composite voice qualities

As a direct consequence of setting up a phonetic system for describing voice quality as the product, like phonetic quality, of co-occurring but analytically separable components, we reach a situation where a given composite voice quality can potentially be the subject of an awkward multitude of descriptive labels. It is clearly undesirable not to have some explicit structural conventions governing the shape of the necessary composite labels. Such conventions have to specify sequential and combinatory constraints on certain groupings of labels, and on individual labels within these groups. The first part of this section suggests a number of conventions of this sort.

a) structural conventions for descriptive phonetic labels

The first convention is that composite labels should have basically two possible gross structures: if labels for tension settings are represented by $T$, those for configurational settings of the supralaryngeal vocal tract by $C$, and those for phonatory settings by $P$, then the first of the alternative structures for a full label is $TCP$. An example of this would be the label tense velarized falsetto. The second possible structure is the order $TPC$, where the label representing $C$ is the phrase 'with ....', as in tense falsetto with velarization; another example of the $TPC$ structure would be lax breathy voice, with horizontal expansion and vertical constriction of the interlabial space, and labial protrusion. (Alternatives to the cumbersome labial description will be discussed in a moment). The only other structure which
should be used is an amalgam of the ones above, where the labels representing C are split into the two possible positions, in a T C P C structure. An example is tense pharyngalized creaky voice with nasality.

Most often, a composite label will not need to fill all three structural places explicitly. Whenever a structural place is left unfilled, it can be conventionally assumed that the neutral setting relevant to that place in structure is operating, or that the nature of the setting applicable to the unfilled place is irrelevant. So in a label like nasal creak, the assumption holds that the tension state of that voice is neutral or irrelevant. The one structural place that should always be explicitly filled is P. This will most often merely be voice, as in nasal voice, and the assumptions can be made that the use of voice alone means modal voice.

Some conventional internal structure can be given to the different individual settings within each of the major elements of a composite label. The discussion of these constraints also affords a useful opportunity to bring together in one place all the different descriptive phonetic labels that have been suggested in the different parts of this chapter, and to set them out in summary form.

No sequential convention is necessary for the settings of overall tension; no co-occurrence is possible, and the choice is merely of one from the three possible settings tense, neutral and lax.

Conventions for the sequence of labels within the configurational settings are helpful. Arbitrarily, the convention can be that the order of labels from left to right in the C element of the composite
label is as follows: firstly, **longitudinal** settings, then **latitudinal** settings, and lastly, **velopharyngeal** settings. Within these groupings, there can be up to three lower degrees of internal structure governing the sequence of the labels for actual settings. Within each category, the assumption still holds that neutral values apply in the absence of explicit mention of a particular setting.

The sequential order of labels for settings within these categories largely follows the anatomical progression from the lips to the larynx, as follows:

<table>
<thead>
<tr>
<th><strong>Longitudinal</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>labial</strong></td>
<td>- labial protrusion</td>
</tr>
<tr>
<td><strong>laryngeal</strong></td>
<td>- raised larynx</td>
</tr>
<tr>
<td></td>
<td>- lowered larynx</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Latitudinal</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>labial</strong></td>
<td>- horizontal expansion of the interlabial space</td>
</tr>
<tr>
<td></td>
<td>- vertical expansion</td>
</tr>
<tr>
<td></td>
<td>- horizontal constriction</td>
</tr>
<tr>
<td></td>
<td>- vertical constriction</td>
</tr>
<tr>
<td></td>
<td>- horizontal expansion and vertical expansion</td>
</tr>
<tr>
<td></td>
<td>- horizontal constriction and vertical constriction</td>
</tr>
<tr>
<td></td>
<td>- horizontal expansion and vertical constriction</td>
</tr>
<tr>
<td></td>
<td>- horizontal constriction and vertical expansion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>lingual</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>tip/blade</strong></td>
<td>- tip articulation</td>
</tr>
<tr>
<td></td>
<td>- blade articulation</td>
</tr>
<tr>
<td></td>
<td>- retroflex articulation</td>
</tr>
<tr>
<td><strong>tongue-body</strong></td>
<td>- dentalized</td>
</tr>
<tr>
<td></td>
<td>- alveolarized</td>
</tr>
<tr>
<td></td>
<td>- palato-alveolarized</td>
</tr>
<tr>
<td></td>
<td>- palatalized</td>
</tr>
<tr>
<td></td>
<td>- velarized</td>
</tr>
<tr>
<td></td>
<td>- uvularized</td>
</tr>
<tr>
<td></td>
<td>- pharyngalized</td>
</tr>
<tr>
<td></td>
<td>- laryngo-pharyngalized</td>
</tr>
<tr>
<td><strong>tongue-root</strong></td>
<td>- advanced tongue-root</td>
</tr>
<tr>
<td></td>
<td>- retracted tongue-root</td>
</tr>
</tbody>
</table>
The terms describing labial activity are suitable for a precise discussion of the settings, but very often such precision is not necessary. When brief reference is adequate, then the extended labels for labial settings can be exchanged for more familiar descriptive terms, as briefly discussed in the section on labial settings.

The phrase with spread lips or with lip-spreading can be used for every label in which horizontal expansion occurs, and with rounded lips or with lip-rounding for every label in which horizontal constriction occurs. The specification for labial settings can be
given a degree of precision half-way between these two extremes, by
translating horizontal constriction with vertical expansion as with
open lip rounding; the phrase with close lip-rounding can be used
for both horizontal constriction and vertical constriction and
horizontal constriction. This only leaves vertical expansion and
vertical constriction, when the horizontal parameter is neutral, without
a ready translation into more familiar phonetic terms. To illustrate
the difference between the detailed composite label and the corresponding
abbreviated labels, we can consider the voice quality of the Reverend
Ian Paisley, the Northern Irish politician. The labial component
in his voice is similar to that of Eric Heffer, the M.P. mentioned
earlier, - they differ in that Mr. Paisley has a more open jaw
setting and slightly less horizontal constriction of the interlabial
space. The full composite label for his voice would be a loud, tense
harsh voice with labial protrusion, horizontal constriction and
vertical expansion of the interlabial space, with an open jaw setting.
The abbreviated labels for this would be either a loud, tense harsh
voice with labial protrusion, open rounding and an open jaw setting,
or a loud tense harsh voice with protruded rounded lips.

Similar conventions can be adopted for the labels for phonation
types representing the $P$ element in the composite label. The first
constraint, that of permissible combinations of phonatory settings,
has already been discussed in the section on compound phonation
types. An arbitrary convention deals with the sequential order of
the phonatory labels.

It will be recalled that the different phonation types were
classified into three groups, depending on criteria of combinatorial
potentials. The first group was made up of modal voice and falsetto, the second of whisper and creak, and the third of harshness and breathiness. The sequential constraints on these labels take account of this classification. The convention is that if either member of the first group, modal voice or falsetto, is present, then it will take the rightmost, terminal position in the element representing P in the composite label. If a member of the second group, whisper or creak, is present, then it will take the penultimate place in the P element; if the first group is not represented, then whisper or creak takes the terminal position. If both whisper and creak are present, then, arbitrarily, whisper will appear to the left of creak, giving whispery creak. Whenever a member of the third group, harshness or breathiness, is present, then it will always take the leftmost, first position in the element representing P. In this way we get composite labels such as creaky voice, for the combination of creak and modal voice, whispery creaky falsetto, harsh whispery voice, and so forth.

If all incompatible combinations of compound phonation types are omitted, the complete list of all possible phonation types, in the descriptive framework offered here, gives twenty items, as follows:

**Simple phonation types**
1. modal voice
2. falsetto
3. whisper
4. creak

**Compound phonation types**
5. whispery creak
6. whispery voice
7. whispery falsetto
8. creaky voice
9. creaky falsetto
10. whispery creaky voice
11. whispery creaky falsetto
12. breathy voice
13. harsh voice
14. harsh falsetto
15. harsh whispery voice
16. harsh whispery falsetto
17. harsh creaky voice
18. harsh creaky falsetto
19. harsh creaky voice
20. harsh whispery creaky falsetto

The twenty items are not, of course, of equally frequent occurrence. The phonation types involving falsetto are much rarer than the others, for instance; and breathy voice is much more common than the single appearance in the list of the label breathy might imply.

It will be recalled that the mode of phonation mentioned earlier in connection with harshness, namely ventricular (voice or falsetto), was said then to have the same distribution as harshness. While 'ventricular' can replace 'harsh' in items 13 to 20 in the above list, nevertheless, because not enough is yet known about phonation achieved with the participation of the ventricular folds, the reference of ventricular as a mode of phonation is limited here to the one application in which one can be confident of its involvement, — the production of severe harshness. This reduces its status, until more is known, to that of a sub-division of a particular type of phonation rather than allowing it full independent status, as it were.

b) scalar conventions for voice quality labels

The final part of the descriptive apparatus for discussing phonetic
aspects of the phenomena of voice quality leads on from the mention above of a 'severe' degree of harshness: it is the matter of scalar labels for the perceptual prominence of any given setting in a speaker's voice quality. At least three degrees of perceptual prominence can be fairly easily distinguished, - slight, moderate, and severe. We can thus refer, for example, to a slightly velarized voice, a moderately breathy voice, a severely nasalized voice. When the description of a particular voice needs more than a single setting-label, then the application of each scalar label can be clarified by punctuation, as in a slightly tense, moderately velarized, slightly nasalized, slightly whispery voice. The use of the TPC structure can also help with longer composite labels, as in a moderately pharyngized severely creaky voice, with slightly lowered larynx, open rounding, slightly open mandibular setting and severe nasalization - (open rounding being used here as the equivalent of slight rounding).

Some economy can be made in the use of scalar labels by the convention that when no scalar label is explicitly given, the moderate value can be taken to apply. There will also be many instances where it is not relevant to use scalar labels.

The one component of a composite label that can't be given a scalar specification is the last item in the element representing the phonatory settings, P. There are four phonation types that can appear in this position, modal voice, falsetto, whisper and creak. The first two never appear in any other position than the terminal one in the P element, so never have scalar labels attached to them in any circumstances. The last two, however, can appear not only in the terminal position, but also in the penultimate position.
penultimate, whisper and creak can properly have scalar labels attached when needed.

With the descriptive apparatus developed in this chapter, it should be possible to communicate reasonably reliably about phonetic aspects of the quality of any organically normal voice. This should also mean that the translation of impressionistic labels for voices into descriptive phonetic terms can be attempted (Laver, 1968: 48), and some illustrative examples are suggested in the following list:

<table>
<thead>
<tr>
<th>Impressionistic label</th>
<th>Descriptive phonetic label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ginny voice</td>
<td>Whispe~ry creaky voice</td>
</tr>
<tr>
<td>Hoarse voice</td>
<td>Harsh whispery voice</td>
</tr>
<tr>
<td>Sepulchral voice</td>
<td>Whispe<del>ry creaky voice wi</del>th slight laryng~o-pharyngalization and slightly lowered larynx</td>
</tr>
<tr>
<td>Pig's whistle voice</td>
<td>Tense, ventricular, severely wispe~ry voice</td>
</tr>
</tbody>
</table>

Because of the inherent unreliability of reference of the impressionistic labels, not every reader might agree with the suggested translations. But at least the descriptive phonetic system allows explicit, testable statements to be made about hypothesized components.

C. The acoustic synthesis of voice quality

One way to test the validity of the suggestions offered earlier in this chapter about the acoustic specifications of particular voice
qualities is to produce a synthetic version of the voice quality concerned, for comparison with the human quality under investigation. Using speech synthesis to do this, in the way outlined at the beginning of section B of this chapter, is a useful complement to the usual reliance on the sensory skills of the phonetician. The great benefit of a speech synthesiser is that not only is the acoustic output exactly known, but also that variation of the output can be precisely controlled. Crystal (1969: 93) writes of the 'maximal controllability provided by a speech synthesiser' in the study of voice quality. This 'controllability' is simultaneously the strength and the weakness of a synthesiser: it is possible to make changes on one spectral parameter at a time, which allows very precise specification of the acoustic stimuli in experiments with subjects judging fine differences of quality; but no human being can produce single parameter changes of this sort in his own speech; because of the inter-linked nature of the various parts of the speech apparatus. So the interpretation of experimental results involving listener-judgments (whether by the experimental subjects or the experimenter himself, in setting up the stimuli for the experiment) should only be made with this caveat very firmly in mind. Synthetic speech (which is almost always based on the principle of using the minimal acoustic specification needed to achieve an acceptable quality of speech, with acceptability normally judged on criteria of intelligibility) is nearly always an approximate simulation, not an exact copy or duplication of the acoustic correlate of the human behaviour being modelled. Accepting this reservation, speech synthesis remains potentially a very useful way of testing the validity of acoustic suggestions about voice quality.
Success in the attempt to synthesise the acoustic correlates of the extrinsic settings in voice quality would have some important consequences. People professionally interested in voice quality, such as phoneticians and speech therapists, could be trained to analyse it into its different strands, using stimuli which were 'maximally controllable'. Most importantly, in an area rife with failure to communicate accurately, for reasons discussed in the first chapter, performance in judging voice quality could be standardized more reliably than is possible at the moment with recorded samples of human voices. The reliability of training with synthetic samples is further enhanced by the possibility of controlling scalar increments of any component in the composite stimuli concerned.

The work briefly reported in this section is an informal feasibility study, designed to establish the possibility of synthesising voice quality using chiefly those acoustic parameters normally considered adequate for the synthesis of linguistically intelligible utterances. The constraint of using basically the same parameters in the two applications, linguistically intelligible utterances and voice quality, reflects the view explicitly stated earlier that extrinsic features of voice quality exploit largely the same parameters of vocal control as phonetic features of spoken language. If this position is valid, then a synthesiser suitable for the production of the latter features should also be largely suitable for the production of the former.

A previous exploratory study along these lines is reported in Laver (1964). An account of some of the work about to be described is given in Laver (1967). The synthetic work is not complete, and forms part of a project still in progress. Nevertheless the results of the project so far are of relevance to the discussion of voice
quality in this chapter, and an account of the project is therefore
given below.

A number of different extrinsic settings in voice quality were
synthesised, mostly with reasonable success, on PAT (Anthony and
Lawrence, 1962; Lawrence, 1953) as 'backgrounds' to a sentence taken
from E.T. Uldall's synthetic program for a long piece of connected
speech, 'The North Wind and the Sun' (Uldall, 1962). The sentence
was

'Then the North Wind blew as hard as he could, but
the more he blew, the more closely did the traveller
fold his cloak around him, and at last, the North
Wind gave up the attempt'.

A tape recording of the results of this synthesis of different
voice qualities is provided with this thesis.

The procedure in this exploratory study of the feasibility of
synthesising voice quality was as follows:

1. A particular setting was selected for simulation.
2. Using normal phonetic techniques of kinesthetic
   and proprioceptive introspection, I produced the
   setting as part of my own pronunciation of the
   sentence (my normal pronunciation of the sentence
   is not markedly different from that of the speaker
   whose reading of the North Wind and the Sun was
   analyzed for the original synthesis).
3. The likely acoustic correlates of the effect of the
   setting were worked out in terms of constraints on
   the acoustic parameters of PAT.
4. These constraints were reproduced on PAT by changing
the limiting values of the parameters affected, and by other means discussed in a moment.

5. The control program for the sentence was run through PAT and the output recorded.

6. The voice quality reproduced by PAT and the original quality of my own performance were auditorily compared.

7. Any necessary adjustments to PAT were made, until the match between the two qualities allowed me to conclude that the same type of setting was involved. A perfect match was not demanded, and different scalar values of the setting in the two voices were accepted as adequate.

On the principle of permuting a small number of basic components to produce a large number of different composite effects (as discussed in the preceding section), eight different settings were synthesised and combined to give 72 different voices.

The eight settings included three supralaryngeal and five laryngeal settings. The three supralaryngeal settings were raised larynx, velarization, and nasality. The five laryngeal settings were modal voice, falsetto, creak, whisperyness and harshness. In this early, ad hoc work, creak was treated as a phonation type on a par with modal voice and falsetto. Whisperyness was treated as a modification that could be superimposed on modal voice, falsetto and creak, and not as an independent phonation type (though it is quite straightforward to produce whisper as the laryngeal excitation of the resonatory tract in PAT; - it was left aside as a possibility for later incorporation, precisely because it was known already that the synthesis of whisper as an independent phonation type was feasible).
Raised larynx and velarization were mutually incompatible because the ad hoc values for the limits to the formant ranges were in competition. Allowing for these factors, combinations of up to four simultaneous settings were superimposed in turn on modal voice, falsetto and creak phonation, to give the 72 different voices.

There are two different series of recordings, differing only in details of how the simulation of nasality was achieved. The two series will be referred to as the K-series and the L-series. The K-series recording has the 72 different voices using one type of nasal simulation in the examples with nasal components, and the L-series recording has 36 different voices using the other type of nasal simulation.

The account given here of the acoustic details of the components of the voices will be followed by lists of the sequences of the voices on the tape itself; the sequence of items within each series' list was determined by the sequence in which it was convenient to change the control settings on PAT between the recording of the items. Each item is made up of one particular complex of extrinsic settings superimposed on the three independent phonation types, modal voice, falsetto and creak, in that order.

The project was governed by the principle of least effort, as it were. In other words, only the fewest and smallest changes from the neutral condition of the PAT parameters necessary for the audible production of the quality concerned were made. It was in this spirit that creak, which in the earlier study reported in Laver (1964) was given a special larynx-pulse shaping, in this later work was differentiated from modal voice and falsetto only by the range of
fundamental frequencies allowed, since a very low fundamental seemed sufficient by itself to produce a synthetic type of phonation auditorily acceptable as a simulation of creak.

The acoustic specification of the PAT settings will be listed in terms of their deviation from the specification of the normal, neutral settings for modal voice. It will be recalled that the normal fundamental frequency range on PAT is 50 - 250 cs; the range of the first formant (F1) is 100 - 1000 cs; that of the second (F2) is 500 - 2500 cs; that of the third (F3) is 1400 - 3400 cs; and the fourth (F4) has a fixed value of 3800 cs. Formant amplitudes have a standard 12 db fall-off per octave. Formant bandwidths for the first three formants are standard at 100 c. Other factors will be mentioned as relevant. The auditory judgments of adequacy of simulation made by colleagues who have listened informally to the recordings will also occasionally be included where relevant.

The only change for the falsetto setting was in the range of the fundamental frequency, which was raised to 120 - 480 cs. The result sounds like adult male falsetto, and not like a female voice, because the formant frequency ranges remain in their neutral values, which are appropriate to an adult male speaker.

The fundamental frequency range for creak was 32 - 128 cs. Since the segmental program for the sentence only occasionally prescribes a rise in the value of the fundamental frequency parameter to near the top of the range, the auditory impression is quite close to that of the human creak; but it should be noted that one of the characteristics of normal creak, mentioned earlier, is severe damping of the larynx pulses, and this was not included in the
synthetic version of the phonation type. As with many qualities
of the voice, we seem to be prepared here to accept a certain variety
of phenomena as manifestations of the particular category of quality.

Nasality in the X-series in the tape recorded material was
synthesised by changing the bandwidth of the first formant from
100 cs to 250 cs. The auditory effect of this is only a partial
success. Some phoneticians who have heard the recordings say that
they are confident the quality sounds nasal; others say that they
would rather call the quality 'muffled' (or other similar terms).
Nasality of course shares a 'muffled' auditory effect with lax
voice, as noted earlier, but it seems better to conclude that the
synthetic means of achieving this effect here was not really suitable,
since in real terms a bandwidth change of this sort would be associated
with similar changes, perhaps of less degree, in the bandwidths of
the other formants also. The 'minimal change' strategy adopted
here was not very effective, then.

Nasality in the L-series of recordings was simulated in quite a
different way, closer in some aspects to the acoustic phenomenon of
nasality in real voices. An additional parameter was included, which
was a 'nasal' formant. The frequency of the formant was fixed at
300 cs, realistically, in the sense that nasal formants in the real
situation don't change in frequency, because the volume of the
nasal cavities doesn't change. The amplitude of the nasal formant
was controlled by that of the fundamental amplitude parameter, with
the two parameters 'ganged' together, so that the value of the
amplitude of the nasal formant rose and fell exactly as the fundamental
amplitude varied. The amplitude relationship of the fundamental
and the nasal formant was that, under the standard condition (where,
as reference, 1 volt = 0 db, with F1 = 500 cs, F2 = 1500 cs, and F3 = 2500 cs), the nasal tract output is 2 db greater than the oral tract output at the maximum output level. The bandwidth of the nasal formant was 290 cs. This type of simulated nasality was acceptable to more phoneticians than the other type, but some still found the simulation unrealistic, and 'muffled' was once again a frequent comment. So the synthesis of really acceptable nasality was not achieved in either simulation, and this remains a problem for further exploration as the voice quality synthesis project continues.

Whisperyness was reasonably easy to simulate, although some phoneticians thought, I believe rightly, that the proportion of the amplitudes of the whisper component and the voice component was not right; a frequent comment was that the whispery examples had an abnormal, pathological tinge to them. Whisperyness was simulated by linking the amplitude of the hiss-through-formants parameter to the fundamental amplitude, with the fundamental amplitude controlling the parametric values for both. The relationship of the two amplitude values was changed, by raising that of the hiss-through-formants parameter by 13 db over the same range as in the neutral condition, and lowering the fundamental amplitude by 3 db over the same range as in the neutral condition. The cut-off points for the frequency of hiss-through-formants were set at 500 and 5000 cs.

Harshness was simulated by superimposing an irregular pitch-jitter on the fundamental frequency parameter. This was done by linking the fundamental frequency parameter to a random-noise generator, whose output had been low-pass filtered at 80 cs, with a filter slope characteristic of 18 db per octave. The result was that
the fundamental frequency swung irregularly about the intonation curve prescribed by the parametric control program, up to a maximum of 60 times per second, by an average amount of +/- 5 cycles, and a maximum amount of +/- 20 cycles.

The auditory effect of harshness was convincing on most voices, but was not very perceptually prominent on voices with either a creak component or a whispery component. This was presumably because of the very low fundamental frequency of creak, which meant that most of the prescribed intonation curve fell below the maximum frequency of the pitch-jitter at 80 cs, and which also meant that irregularity of the fundamental frequency was more difficult to perceive wherever the fundamental approached the bottom end of its range at 32 cs, where the auditory sensation of pitch begins to lose its character. In the case of whispery voices, the effect of adding the random-noise element of harshness to what is already a random-noise element of whisper (although neither is completely random, nor are they identical in the details of their aperiodicity) did not give striking differences.

Velarization was reasonably successfully simulated. This was done by changing the frequency ranges of the second and third formants, F2 to 850 - 2150 cs and F3 to 1800 - 3000 cs.

The synthesis of the raised larynx setting is a problematic case. This setting in the live situation is often associated with a raised fundamental frequency range. In an earlier stage of the project, a quite acceptable simulation of the quality of raised larynx voice had been achieved, with the formant range alterations listed below together with a moderate rise in the fundamental range. However, as noted above, a rise in fundamental frequency is not an
essential concomitant of raising the larynx, because of compensatory possibilities of re-adjusting the setting of the pitch mechanism of the larynx. So, in accord with the decision to make only the minimal necessary changes from the neutral setting, raised larynx was specified in the present sets of recordings with the fundamental frequency range raised only slightly at the bottom end of the range, giving a scale of 60 - 240 cs. But the result was that the formant range adjustments which had sounded acceptable when the fundamental range was higher now gave an effect less like the stereotyped quality of a longitudinal distortion of the vocal tract, and more like a latitudinal distortion. The effect was that of a tongue setting where the body of the tongue was lowered and retracted, giving the auditory impression of laryngo-pharyngalization, - rather like the voice quality impressionistically called a 'hot-potato voice' or 'plummy voice', mentioned in the section on tongue settings, as if the speaker actually had an object in his mouth that depressed the tongue. When I tried maintaining a raised larynx position with an unraised fundamental frequency range, the auditory effect was very like the synthetic quality. It may be that what we are prepared to accept auditorily as corresponding to constriction of the upper larynx and lower pharynx is the same resonatory distortion of the vocal tract that results from raising the larynx; and that the upward shift of the fundamental frequency in most raised larynx voices is the factor chiefly responsible for the auditory differentiation of the two qualities.

The formant range adjustments that were made for the qualities labelled as raised larynx voice were that F1 was changed to 150 - 1050 cs, F2 to 600 - 2100 cs, and F3 to 1500 - 3000 cs.
The feasibility of synthesising the effects of extrinsic settings in voice quality, using largely the same vocal parameters and their acoustic correlates as in the synthesis of linguistic utterances, seems reasonably demonstrated. More controlled work on the synthesis of voice quality can now be undertaken, both by prescription, as it were, as in the project reported here, and through a prior acoustic analysis of individual live voices. Comment on the area of acoustic synthesis of voice quality has been included in this thesis partly as a pointer to the sort of directions that work on voice quality in the future might profitably take, and partly as an acknowledgement that theoretical discussion about topics like the description of voice quality has eventually to be framed in a way that makes it amenable to empirical testing, both by the sensory skills of the phonetician and by the techniques of the phonetic laboratory. It was also for this reason that it was felt necessary to include physiological and acoustic specifications wherever possible in the discussion of the different types of settings in the descriptive part of this chapter.

List of items on the tape recording of the K-series of synthetic voice qualities

Each item shows one or more voice quality settings superimposed on three phonation types, modal voice, falsetto and creak, in that order.

K1 Unmodified modal voice, falsetto, creak
K2 Nasal voice, falsetto, etc.
K3 Whispers voice, etc.
K4 Nasal whispery voice
K5 Nasal harsh whispery voice
K6 Harsh whispery voice
K7 Nasal harsh voice
K8 Harsh voice
K9 Velarized voice
K10 Velarized nasal voice
K11 Velarized whispery voice
K12 Velarized nasal whispery voice
K13 Velarized nasal harsh whispery voice
K14 Velarized harsh whispery voice
K15 Velarized nasal harsh voice
K16 Velarized harsh voice
K17 Raised larynx voice
K18 Raised larynx nasal voice
K19 Raised larynx whispery voice
K20 Raised larynx nasal whispery voice
K21 Raised larynx nasal harsh whispery voice
K22 Raised larynx harsh whispery voice
K23 Raised larynx nasal harsh voice
K24 Raised larynx harsh voice

List of items on the tape recording of the L-series of synthetic voice qualities

Voice quality settings superimposed on voice, falsetto, and creak, as in the K-series recording

L1 Raised larynx nasal harsh voice, falsetto, creak
L2 Raised larynx nasal harsh whispery voice etc.
L3 Raised larynx nasal whispery voice
L4 Raised larynx nasal voice
L5 Velarized nasal harsh voice
L6 Velarized nasal harsh whispery voice
L7 Velarized nasal whispery voice
L8 Velarized nasal voice
L9 Nasal harsh voice
L10 Nasal harsh whispery voice
L11 Nasal whispery voice
L12 Nasal voice
Part III

Chapter 4  A semiotic view of spoken communication

Introduction

Having offered in the preceding chapter a means of describing phonetic aspects of the phenomena of voice quality, we turn now to consider some semiotic aspects of voice quality as a factor in spoken communication. This is such a wide and diverse field that it would be impracticable to give a comprehensive account within the limited scope of this thesis. The summary suggestions set out below, particularly those on the topic of indexical information conveyed by voice quality, should be regarded as perspectives for future study, rather than as definitive proposals.

It was asserted earlier that voice quality stands in a complementary relation to phonetic quality. It is necessary now to expand this, and to elaborate the view that within voice quality, and within phonetic quality, there are internal aspects of quality which also stand in complementary relation to each other. Within voice quality, we have already seen that intrinsic and extrinsic features of quality have to be distinguished; in a moment we shall explore the implications of the idea that the perception of extrinsic quality can only be accurately achieved when the contribution to the overall voice quality of the intrinsic features is known.

Within phonetic quality, the aspects which constitute the exponents of linguistic units have to be distinguished from those which are the manifestations of paralinguistic units. This relationship between linguistic and paralinguistic units will be considered later.
The chapter begins with a brief discussion of the semiotic terms to be used, and then considers voice quality and the communication of indexical information. The final part of the chapter looks in more detail at semiotic aspects of the relationship between voice quality and phonetic quality.

A. Semiotic terminology

Semiotics was said in the Introduction to the thesis to be 'the study of communicative signs' (Morris, 1938: 80): the semiotic concepts discussed below can help to illuminate not only the communicative function of voice quality as such, but also the relation between voice quality as an extralinguistic sign and phonetic quality as a linguistic (and sometimes paralinguistic) type of sign.

The original application of the term 'semiotic' was in an area close to the indexical interests of this thesis: Morris (1946: 285-287) says that it was first used in Greek medicine, to refer to the theory of medical symptoms used as signs in the diagnosis and prognosis of disease; the Stoic philosophers then used 'semiotic' to mean the general theory of signs, in which was included logic and epistemology, with Stoic philosophy being divided into semiotic, physics and ethics.

Morris then traces the history of semiotic into medieval Europe, through the works of Augustine and Boethius. There, the subject became known as 'scientia sermocinalis', in the work of such figures as Petrus Hispanus, Abelard, Roger Bacon, Thomas of Erfurt, Sigur of Courtrai and William of Occam. During this period, two divergent traditions developed: one led to the work of the British empiricist philosophers like Francis Bacon, Hobbes, Locke, Berkeley, Hume and Bentham; the other, through the work of Leibniz, led to that of
modern symbolic logicians such as Boole, Frege, Peano, Peirce, Russell, Whitehead, Carnap and Tarski. It is to the work of a member of this second group, Charles Peirce, the American pragmaticist philosopher of the late nineteenth century mentioned in the introduction to this thesis, that the semiotic discussion in this chapter owes most.

It was Locke who was responsible for the re-introduction of the term 'semiotic', in the late seventeenth century, as

"the doctrine of signs (....) the business whereof it is to consider the nature of signs (that) the mind makes use of for the understanding of things, or conveying its knowledge to others" (Locke, 1690).

Peirce gave a slightly different definition of 'semiotic' from that of Locke: semiotic, he wrote, is the

'formal doctrine of signs (.... where ....) a sign is something which stands to somebody for something in some respect or capacity' (2.227 - 228).

(Specific reference to Peirce's writings will be to the first six volumes of his collected papers, edited by Hartshorne and Weiss (1931-1935), citing volume and paragraph number, as in the above citation). This neutral, all-embracing definition of a sign by Peirce will be adopted in the discussion below.

Peirce was a strikingly original thinker, who produced an array of seminal ideas, some of which are borrowed here. It is because of his importance in the field of semiotics, which has only recently been generally acknowledged, and also because of the current growth of interest in semiotics on the part of linguistics and related disciplines, that I feel some obligation to indicate
correspondences between semiotic terms and ideas used in this thesis and those of Peirce. It has to be said immediately, however, that Peirce's work is not always the easiest to follow: Macdonald (1935), one of the most interesting commentators on Peirce, writes that 'his theory of signs is very involved, often obscure and incomplete, like everything he wrote' (Macdonald, 1935: 108). This is not to deny the value of some of his ideas: Bertrand Russell, in the foreword to a book on Peirce by Feibleman, wrote that Peirce 'reminds one of a volcano spouting vast masses of rock, of which some, on examination, turn out to be nuggets of (...) gold' (Feibleman, 1946: XVI). Some of Peirce's ideas are seminal then, but because of the obscurities and occasional contradictions implicit in his writings (cf. Burks, 1948-1949: 675) it will often be necessary to modify, and usually to simplify, the concepts borrowed from him.

Peirce divided the different sorts of signs into three mutually-intersecting trichotomies. The second is the most relevant here. Feibleman (1946: 90) gives a condensed quotation of this as follows:

'The second trichotomy of signs consists of the icon, a sign which refers to an object by virtue of characters of its own which it possesses whether the object exists or not (2.2\#7); the index, a sign which refers to the object that it denotes by virtue of being really affected by that object (2.2\#8); and the symbol, a sign which refers to the object that it denotes by virtue of a law, usually an association of general ideas, which operates to cause the symbol to be interpreted as referring to that object (2.2\#9)'

In these definitions, the reference of the sign to its object is non-arbitrary in the cases of the icon and the index, and arbitrary in
the case of the symbol.

Icon, index and symbol are the three basic semiotic terms borrowed from Peirce in this thesis. The notion of an icon was used in the first chapter, on principles of labelling voices, in the Peircean sense which has become part of the general vocabulary of our culture. This is the sense broadly captured by Peirce when he wrote that 'Anything whatever, be it quality, existent individual, or law, is an Icon of anything, in so far as it is like that thing and used as a sign of it' (2.247). As Feibleman (1946: 91) nicely puts it: 'The functioning of an icon as a sign is dependent upon its capability of similarity of structure'. Strictly, it is less the notion of an icon, as such, that was useful in the first chapter, but more the notion of an iconic relationship holding between a sign and the object to which it makes reference.

The concept of a symbol will be of relevance later in this chapter, but that of the index, which has been used in a number of places throughout the thesis, will be the present focus of attention.

The notion of an index is potentially a very fruitful one. Unfortunately, it is also the least well-defined concept of Peirce's trichotomy of icon, index and symbol. He used the term 'index' in many different senses at different times, and it will be necessary to modify and limit the definition of the term.

Feibleman's (1946: 91) condensed quotations illustrate Peirce's varied, sometimes obscure and sometimes contradictory uses of the term:

'The index "is a real thing or fact which is a sign of its object by virtue of being connected with it as a matter of fact" (4.667). A genuine index and its object must
be existent individuals (whether things or facts) .... The index refers to its object "by virtue of being really affected by that object .... In so far as the index is affected by the object, it necessarily has some quality in common with the object, and it is in respect to (this) that it refers to the object" (2.248). Hence "the index is physically connected with its object" (2.299) and "they make an organic pair". An index "is a sign which would, at once, lose the character which makes it a sign if its object were removed, but would not lose that character if there were no interpretant" (2.304).

Indices have three characteristics which distinguish them from other signs. They "have no significant resemblance to their objects .... they refer to individuals .... they direct the attention to their objects by blind compulsion" (2.306). "A rap on the door is an index. Anything which focuses the attention is an index. Anything which startles us is an index, in so far as it marks the junction between two portions of experience" (2.285).

There are more examples of yet other senses of Peirce's use of the term. But following Macdonald (1935: 115) it is possible to reduce the variety to two general senses: the first is the one to which we can refer as the evidential sense. The best illustration from Peirce's writings of this sense is the comment quoted above, where he said that the index refers to its object "by virtue of being really affected by that object" (2.248). The orientation of a weathercock would, in this usage, be evidence for, or an index of, wind-direction, and the height of a column of mercury in a thermometer would be an index of heat.

The second sense is what we might call the demonstrative sense. This is the sense of 'indexical' which has been perpetuated in philosophy
since Peirce by, for example, Macdonald (1935), Burks (1946-49), Bar-Hillel (1954) and Searle (1969), (cf. also Lyons, 1972: 86n.22). The more usual technical linguistic term for this sense of 'indexical' is 'deictic' (cf. Lyons, 1968: 275). As Macdonald puts it

"we must ultimately point out our meanings and this is done by means of an Index. The Index in this sense is a sign which denotes an object without describing it; "The Index asserts nothing; it only says 'there!'. It takes hold of our eyes, so to speak, and forcibly directs them to an object and there stops" (3.362). Hence the demonstrative pronouns "this" and "that" are indices' (Macdonald, 1935: 115).

To quote Peirce again, directly, 'Anything which focuses the attention is an index' (2.285). So such linguistic devices as pronouns, adverbs of time and place, and personal and proper names would all qualify for this demonstrative sense of the term 'index'. Even more generally, it could be said that all referring signs used in language have, to some extent, an indexical element which serves to focus the listener's intellectual attention on the referents of the signs.

Peirce's uses of the term 'index' are exemplified here at some length because it is important to be quite clear about the sense in which 'index' and related terms such as 'indices', 'indexical', 'to indicate', 'indicative' and 'indication' are used in the following discussion. To avoid Peirce's ambiguity and obscurity, a restricted and explicit definition is proposed here, based partly on his evidential sense, and discarding his demonstrative sense. In this chapter, the terms 'index', 'indexical' and related forms, will now be applied only to intrinsic and extrinsic vocal features which serve, by acting as evidence for physical, psychological and social
attributes, to identify or characterize the speaker. One way in which this definition differs from any offered by Peirce in the material quoted above is that, because extrinsic features can be involved, the reference of an index to its object may be in some cases arbitrary.

This general view of the concept of an index derives from that of Abercrombie (1967). It will be recalled from the first chapter of this thesis that he divides indices in speech into three classes: those that indicate membership of a group; those that characterize the individual; and those that reveal changing states of the speaker (Abercrombie, 1967: 7). One aspect of this classification will be particularly relevant to the second part of this chapter, that of indices that 'reveal changing states of the speaker'. Lyons (1972: 72) has given the apt term symptomatic to indices in this class. One such group of symptomatic indices that Abercrombie discusses is made up of what he calls affective indices, - those that 'do not have a direct physical cause, (.....) from which we infer feelings such as amusement, anger, contempt, sympathy, suspicion, and everything else that may be included under "tone of voice"' (Abercrombie, 1967: 9). The point to be noted here is that paraphonological elements (the aspects of vocal behaviour that Abercrombie subsumes under 'tone of voice' in the quotation above) can quite properly be considered on two different levels of analysis. On one level they are regarded as manifestations of the code of arbitrary, conventional signs called paralanguage. On another level, they are seen as events which indicate the momentary affective state of the speaker. This is analogous to the situation involving language: a particular
linguistic utterance can be analysed as a sequence of manifestations of the code of arbitrary, conventional signs called language; on another level a variety of aspects of the utterance can be taken to indicate that the speaker is an American rather than a Scot, for instance.

B. Voice quality and indexical information

Distinguishing between extrinsic indices and intrinsic indices, as extrinsic or intrinsic aspects of vocal behaviour that can indicate a speaker's identity or characteristics, it is of some interest to explore the question of their different degrees of relevance to general phonetic theory.

Extrinsic indices are properly of central relevance if one accepts that the task of general phonetic theory is to give an account of all aspects of extrinsic vocal behaviour which can figure in spoken communication. It is this view that is taken in the account of indexical information below. If one prefers a more traditional definition of the aim of general phonetic theory, where it is an account of all extrinsic aspects which have been found to act as linguistic signs, then extrinsic indices have central relevance only to the extent that they indicate, through factors of accent, sociolinguistic characteristics of the speaker.

Intrinsic indices are of largely peripheral concern to general phonetic theory, under any definition of 'phonetic', because of the 'general' criterion. While general phonetic theory tries to account for vocal features which are general to all normal human beings, intrinsic indices are by definition concerned with speaker-specific factors such as the geometry and action of a given speaker's vocal
apparatus. Intrinsic indices do have one aspect of basic relevance to general phonetic theory, however: that is, that it is against the background of the intrinsic features that the extrinsic features of central concern to phonetic theory gain their perceptual and analytic definition. It is in this interpretation that a discussion of intrinsic features has been included in this phonetic study. Intrinsic features in the voice by definition reflect only the physical attributes of the speaker: a short outline of different physical indices is given immediately below.

1. The indication of physical attributes by voice quality

One complication in the consideration of intrinsic indices is that the auditory phenomena arising from the physical attributes of one speaker can sometimes be extrinsically imitated by another speaker. The example cited at the beginning of the previous chapter may be recalled, where

'a certain music-hall performer, a large, middle-aged man, who had learnt to produce, completely convincingly, the voice-quality of a seven-year-old girl, showing that it is possible to compensate, by muscular adjustments, for extreme anatomical differences' (Abercrombie, 1967: 94).

Another example of the use of extrinsic features to mislead listeners into concluding that the speaker possesses certain intrinsic characteristics, is the use by some popular female singers and film actresses of whispery voice. This simulates the effect on phonatory quality of the change in consistency of the mucal lining of the larynx which takes place during sexual arousal.
These are, however, somewhat unusual examples, and listeners' judgements of physical attributes from intrinsic clues in voice quality are usually amongst the more accurate conclusions drawn. This is presumably precisely because the clues derive from a largely invariant, normally non-manipulable aspect of vocal action.

It is useful here to draw a distinction between those physical parameters which are universal, on which every speaker can be placed, and those which are much more limited in application, specific to only a small section of the population. All speakers can be classified in terms of size and physique, sex and age (Laver, 1968). These parameters will be discussed first.

There seems to be a general correlation between a person's size and physique and the dimensions of his larynx. If we hear a very deep-pitched, loud voice, with a corresponding phoratory quality, over the telephone, we confidently expect the speaker to turn out to be a large, strong male; in general, our expectations are fulfilled, within a reasonable margin of error. Bonaventura (1935) gave subjects pictures and voices to match, and found that fair accuracy was achieved: in terms of Kretschmerian body-types (Kretschmer 1925), judgements of pyknic types were most accurate, less accurate for leptosome types, and least for athletic types. Moses (1940, 1941) gives general support to this, and Fay and Middleton (1940a) report a more detailed finding: they found that in judging body-types from voices transmitted over a public address system, the results were 22% above chance for pyknic types, 20% above for leptosomes, but only 1% above chance for athletic types.
There is one class of voices where the general correlation does not apply, but where listeners nevertheless seem to be able to reach successful indexical conclusions about the physical attributes. That is where the formant ranges of the voice are radically discrepant with the fundamental frequency, as in particular types of dwarfism (Vuorenkoski, Tjernlund and Perheentupa, 1972; Weinberg and Zlatin, 1970).

Exceptions to the general rule of our ability as listeners to attach a particular size and physique to a given voice are sufficiently rare to take us aback when they occur.

One usually forms fairly accurate impressions, from voice quality, of a speaker's sex and age (Mysak, 1959; Ptacek and Sander, 1966; Shipp and Hollien, 1969; Tarneaud, 1941; Zerffi, 1957). Deviations from 'normal' expectations about the correlation between a speaker's voice and his sex and age seem to have a powerful effect on impressions of personality.

A number of writers have commented on the acoustic correlates of male and female voices. Apart from the obvious tendency for females with smaller vocal dimensions to have higher ranges of fundamental frequency, the most frequent comment is to do with the spectral correlates. Tarnóczy and Fant (1964) note that while average spectra for speakers of both sexes vary with the language spoken - echoing the discussion in the previous chapter of this thesis about the contribution to voice quality of the relative frequency of occurrence of the different linguistic segments - there is a tendency, usually irrespective of the language spoken, for a spectral minimum to occur at about 900 cs for males and 1000 cs for
females. They report this as true for Swedish, Hungarian and English, but not for German male speakers, where they say 'this effect could be counteracted by the specific language structure' (Tarnóczy and Fant, 1964: 11).

Fant (1960) gives a list of acoustic values for average male subjects, and compares them with those for females and children. He writes that:

'The natural range of variation of the voice fundamental frequencies for non-nasal voiced sounds uttered by average male subjects is as follows:

- \( F_0 \) - 60 - 240 cs
- \( F_1 \) - 150 - 850 cs
- \( F_2 \) - 500 - 2500 cs
- \( F_3 \) - 1500 - 3500 cs
- \( F_4 \) - 2500 - 4500 cs

Females have on average one octave higher fundamental pitch but only 17% higher formant frequencies; see Peterson and Barney (1952); Fant (1953). Children about 10 years of age have still higher formants, on the average 25% higher than adult males, and their fundamental pitch averages 300 cs. The individual spread is large' (Fant, 1960: 21).

In an article in 1966, Fant amended this position slightly, when he said that:

'The common concept of physiologically induced differences in formant patterns comparing males and females is that the average female F-frequencies are related to those of the male by a simple scale factor inversely proportional to the overall vocal tract length (i.e. female F-pattern about 20% higher than
male). (...) simple scale factor rule has important limitations' (Fant, 1966: 22).

He points out that the deviations from the rule are obscured if an average is taken over all vowels, and says that female-male relations are

'typically different in (1) rounded back vowels, (2) very open unrounded vowels, (3) close front vowels. (...) The main physiological determinants of the specific deviations from the average rule is that the ratio of pharynx length to mouth cavity length is greater for males than for females, and that the laryngeal cavities are more developed in males' (loc. cit.).

'The scale factor relating average female formant frequencies to those of men is a function of the particular class of vowels. (...) The female to male scale factor is of the order of 18% averaged over the whole vowel system. (...) The scaling of children's data from female data comes closer to a simple factor independent of vowel class' (Fant, 1966: 29).

Certain indices in voice quality of the age of the speaker are reasonably familiar to us all, including the quality associated with the 'breaking' voice of puberty, and the quality of extreme old age. Luchsinger and Arnold (1965: 132-137) give an interesting survey of some of the physiological factors involved in both cases.

Vocal indications of puberty, referred to in clinical literature as 'vocal mutation', often include whispy voice. Luchsinger and Arnold (1965: 132) write that

'In addition to the lowering of the average speaking pitch, the voice is frequently husky during mutation, or it may sound weak';
They later expand on this comment, in saying that a 'typical laryngoscopic sign of mutation is the incomplete closure of the cartilaginous glottis. One then sees the so-called mutation triangle, which is similar to the "whispering triangle", or the deficient posterior glottal closure in hypokinetic dysphonias (op. cit., p.133).

The senescent voice of extreme old age derives from a complex of endocrinial, anatomical and physiological changes. The mucal fluid supply often becomes disturbed, either greatly increasing or decreasing, tissues become increasingly less elastic, and cartilages become calcified and ossified (Flrfe and Naylor, 1958; Luchsinger and Arnold, 1965; Meader and Huyskens, 1962; Terracol and Azémar, 1949). Meader and Huyskens (1962: 77) comment that 'Since the rigidity of tissue is one determination of its resonating qualities, the gradual deposition of lime in (...) cartilages (replacing them by bone) helps to explain the shrill voice and thin voice (deficient in harmonics) of age'.

Because muscles atrophy, the glottis of old speakers often has a bowed appearance (Luchsinger and Arnold, 1965: 136; Tarneaud, 1941); this means that to achieve phonation, greater adductive effort has to be exerted, and rather harsh voice is often the result. When this is combined with inefficient phonation because of an excess of mucus, the type of voice that results is harsh whispery voice, as suggested by the following comment from Luchsinger and Arnold (1965: 136):

'Tracheal and laryngeal mucous secretions are increased, sometimes on an allergic basis. Together with a tendency to chronic bronchitis, this over-secretion of mucus produces the hacking, coughing, throat-
Anatomical changes include an accumulation of fatty tissue in the laryngeal ventricles (Ferroni, 1959), and a shrinking of the ventricular folds towards the sides of the larynx, giving a wider entrance to the ventricles (Luchsinger and Arnold, 1965: 136); all these factors can contribute significantly to the fine detail of the acoustic quality of any phonation type used. Luchsinger and Arnold mention work by Braus (1924), who pointed out that the larynx and the entire respiratory and digestive tract are in a lowered position with senility (Luchsinger and Arnold, 1965: 137), because of the loss of elasticity of the muscular and ligamental structures from which these organs are suspended, which presumably results in lowered larynx voice.

Luchsinger and Arnold end their section on senescence of the voice with an interesting indexical comment. They say that

'In the geriatric sector of phoniatrics, counseling regarding the hygienic use of the voice in old age is a matter of psychological advice. As he nears the end of life, the elder citizen may continue to play a productive role in social life despite the progressive loss of his physical powers. At this age the voice no longer serves for physical attraction as it does in the young male who is about to found a family and to strive for their care. Likewise, the senile voice is no longer suitable for the militant leadership of assembled males, be it in army barracks or in executive offices of industrial enterprises. As in Areopagus, the assembly of elder statesmen in ancient Athens, the wise old man is expected to give counsel to those of lesser experience. It is well understood in theatrical art that the figure of the benevolent patriarch is portrayed by calm speech with a low and relaxed voice.'
The discussion immediately above concerned universal parameters on which all speakers can be placed. We now turn briefly to consider the case of some physical attributes which can be indicated by voice quality which normally apply to smaller sections of the population. In these more specific cases, the largest grouping is one which might be called the speaker's medical state and it is this that will be exemplified below. A speaker's 'medical state' covers such aspects as abnormalities of anatomy or physiology, the physical effects of trauma or disease, the noxious effects of alcohol, drugs or smoking, transient effects of endocrinal changes, signs of fatigue, and others.

Voice quality indicates a surprisingly wide range of information about a speaker's medical state (Laver, 1968: 49). It is useful to distinguish, following Abercrombie (1967: 92), between relatively permanent aspects of a speaker's medical state, as life-long or very long-term conditions, and more transient medical states. Information about permanent aspects can involve matters which are primarily anatomical, physiological, or endocrinal, both in 'normal' and 'abnormal' conditions. The discussion below will be confined to 'abnormal' conditions.

Abnormalities of anatomy can be revealed by voice qualities associated with a number of conditions: cleft palates; unusual patterns of dentition (Lawson and Bond, 1968); the use of dentures (Lawson and Bond, 1969); unusual conformations of the jaw. A number of abnormal congenital conditions can also be indicated: vocal fold sulcus (a furrow along the glottal edge of the vocal fold, which gives rise to so-called diplophonic voice, a mode of phonation
with two different simultaneous fundamental frequencies (Kinsl, 1962); some types of cleft palate (Jaffe and de Blanc, 1970; Lovr, 1970); sub-glottal bars (Howie, Ladefoged and Stark, 1961); and intrinsic factors in the voices of monozygotic twins (Alpert and Kurtzbarg, 1963). One bizarre oddity which is interesting is that the result of a particular type of injury to the larynx from excessive vocal abuse can alter the laryngeal anatomy in one rather drastic fashion. The laryngeal ventricle has a small appendix which can be ruptured by extreme intralaryngeal air pressure. In extreme exertion, as noted earlier, the true vocal folds separate slightly, because of their particular cross-sectional shape, while the ventricular folds 'remain competent. Under these conditions the full force of the intralaryngeal pressure is exerted at the ventricular level' (Frederickson and Ward, 1962) and if the pressure is sufficiently great, the appendix dilates and herniates, with air finding its way into the tissue of the neck. A hernia of the ventricular appendix of this sort is called a laryngocele, and Wright and Maguda write that:

'Clinically, the first description of the laryngocele was in 1829 by Larrey, who was Surgeon-in-Chief of Napoleon's army. In Egypt, he found that some of the criers, who were the blind men engaged by the chief priests to cry the verses of the Koran from the top of the minarets, developed sub-maxillary swellings which were distended with air during phonation. In time, these tumafactions became so voluminous as to require a stiff collar for support. Ultimately, these persons known as criers had to change occupation. He also noticed this condition among some drillmasters, and believed it to be due to vocal abuse' (Wright and Maguda, 1964).

Abnormal physiological conditions such as those due to pathology, as in Parkinsonism and laryngeal paralysis, can be indicated by voice

Clues in voice quality to endocrine abnormalities are found in the cases of voice disorders resulting from diseases of the thyroid, adrenal, and pituitary glands (Luchsinger and Arnold, 1965: 188-217). Systematic research into the possible use of voice quality as a diagnostic sign (in the original use of a 'semiotic' sign) of these and similar medical states would be extremely valuable; so far, the area has only sporadically attracted investigation (McCallum, 1951; Palmer, 1956; Punt, 1959; Somnin, 1960; Laver, 1968).

More transient medical states can be indicated by voice quality when the speaker is suffering from conditions of local inflammation of his vocal organs, as in laryngitis, pharyngitis, and tonsillitis, and from nasal catarrh, adenoids or a cold (Abercrombie, 1967: 92; Laver, 1968: 49).

Other transient factors in voice quality derive from changes in the copiousness and consistency of the supply of lubricating mucus in the larynx, and in the characteristics of the mucal lining covering the actual vocal folds, affecting the efficiency of their vibration. One such state has already been mentioned, - the whispery voice of sexual arousal in both men and women. The effect can last for several days, as witnessed by an amusing anecdote by John Morgan, writing in the New Statesman some years ago:

'I think it's a pity that Covent Garden's prices are going up in the new season (....) Perhaps opera singers should be paid less. When this dark thought crosses the mind, I always remember being told by one of the world's great male singers that his life
involved him in great sacrifices. Making love affected his voice. Smoking didn't, drinking didn't - just the act of copulation enjoyed even three days before a performance. He earned £500 every time he trod the boards. Few people can be faced with such a poignant choice so often, would you say? (Morgan, J., 'London Diary', New Statesman, 61: 71, 15 January 1965).

A similar endocrinial effect is found in the voices of women in the pregnant or premenstrual state, according to Tarneaud, (1941) and Perelló (1962). Greene (1964: 80) cites Hildernisse (1956) as reminding us that 'singers often have a clause in their contracts to exempt them from singing during the menstrual period'. See also Anado (1953). It is here that the notion of waves travelling through the mucal fluid covering the vocal folds, (differently in the various different consistencies and copiousness of the mucus), which constitutes the Smith-Perello hypothesis of the 'muco-undulatory' action of vocal fold vibration mentioned in the preceding chapter, can be seen as a necessary complement to the basic aerodynamic-myoelastic theory (as far as considerations of fine differences of voice quality are concerned).

Examples of transient states which can become relatively permanent, and which can be detected in voice quality, are the effects of alcohol and hot tobacco smoke. In excess, these agents tend to damage the vocal folds. 'Whisky voice', 'ginny voice', 'brandy voice' and 'trummy voice' are popular labels for the deep-pitched, harsh whispery voices that tend to signal one result of excessive consumption of alcohol. The phrase 'smoker's larynx' is a fairly frequently used medical label for the pathological effect of excessive hot, toxic tobacco smoke on the vocal folds (Myerson, 1950; Devine, 1960).
Lastly, information about transient states such as fatigue can sometimes be found in voice quality. In extreme fatigue, the mode of phonation can become inefficient, resulting in whispery voice or in lax breathy voice. It should be noted, however, that in less extreme cases of fatigue, Fay and Middleton (1940b) showed that the ability of listeners to judge a speaker’s rested or tired condition from his voice alone appears to be based on stereotypes, in that judges agree with each other in their judgements but are much less often accurate.

It follows from the great variability of intrinsic aspects such as these, that the accurate perception of extrinsic quality in any given voice is a quite remarkable perceptual feat. Yet we seem to have the ability, as listeners hearing a particular speaker for the first time, to reach perceptual decisions quite rapidly about the allocation of particular features in his speech to the intrinsic base of his voice quality; to the extrinsic overlay on this of his voice quality settings; and to the extrinsic complex that the speaker exploits for the purposes of linguistic and paralinguistic communication. No doubt some of these decisions take longer to reach than others, and perhaps some are more provisory, more interim than others: the fact that we can do it at all is only slightly less surprising than that we seem to be able to do it rapidly. Exactly how we go about this task, in psychological terms of perceptual strategies, is an extremely interesting question beyond the scope of this thesis. But the solution to that question lies in the very foundation of our ability to perceive spoken language. The role of intrinsic features as a perceptual background against which the extrinsic features are discerned is thus of some considerable importance to a general phonetic theory, even though intrinsic
features as such lie outside that theory.

2. The indication of psychological attributes by voice quality

We seem to be prepared, as listeners, to draw quite far-reaching conclusions from voice quality about psychological attributes of speakers (Laver, 1968: 49-50). In our Western culture, we are ready to believe, for example, as noted earlier, that a harsh voice is correlated with more aggressive, dominant, authoritative personalities, and a breathy voice with more self-effacing, submissive, weak personalities. This belief occasionally finds expression in some rather eccentric assertions. Addington (1963) suggests that particular voice qualities often indicate different traits in men versus women. A 'breathy' quality showed that men are 'aesthetic' and women 'pretty and callow'; 'flat' that men are 'distant', women 'hard and lethargic'; 'nasal' that men are 'unattractive and self-effacing' and women the same; 'tense' that men are 'cantankerous' and women 'high-strung'; 'throaty' that men are 'stable' and women 'oafish'; 'rotund' that men are 'suave' and women 'aggressive'. and so on. The validity of a connection between voice and character is a central belief of the writers on physiognomics (Foerster, 1893; Delestre, 1866), as was mentioned at the beginning of the chapter on the history of voice quality analysis. Delestre (1866: 370) for example, wrote that 'Un aveugle pourrait déterminer le temperament d'un orateur, par la nuance de sa voix'.

This idea that personality characteristics (both normal and psychopathological) are correlated with voice quality, has been tested more scientifically by many writers, mainly in the medical and psychological fields (Allport and Cantril, 1934; Brody, 1943;
Cohen, 1961; Diehl, White and Burle, 1959; Eisenberg and Zalowitz, 1938; Fay and Middleton, 1939a, 1940b; Froeschels, 1960; Goldfarb, Braunstein and Lorge, 1956; Kramer, 1964; Mallory and Miller, 1958; Moore, 1939a, 1939b; Moses, 1954; Pear, 1957; Sapir, 1927; Starkweather, 1964; Taylor, 1934).

Some controversy remains, but in general writers seem to agree that some such broad correlations do exist. From our general social experience, most of us would agree with them, but one major obstacle in the way of reliable scientific statements has been the lack of any standard system for labelling the voice qualities concerned, and a related inability to specify more than a fairly crude quantification of the voice quality variables which act as the experimental variables. This has mostly meant that only those voice qualities whose labels are fairly generally familiar have been investigated in this connection. They include breathy voice, nasal voice and harsh voice. Moore (1939a, 1939b), for example, suggests that breathy voice indicates a personality which has high neurotic tendencies and a strong tendency to introversion, but which has low dominance; that a 'harsh metallic voice' indicates dominance and emotional stability; and that a 'nasal whine' is indicative of emotional instability and low dominance, though less so than in the case of speakers with a breathy voice. The problem here is that one can't be sure that 'breathy voice', say, means the same for Moore as for other writers. The problem becomes acute when the situation arises of challenging the findings of such research. Diehl, White and Burle (1959), for example, say that they found no relationship between nasality and anxiety (op. cit.: 285), which apparently contradicts Moore's finding. But we can only finally be sure that this is a valid
confrontation when it can be established that 'nasality' in the two research reports has the same referent.

The uncertainty of reference of the labels for voice quality used in this area of research is more insidious when reasonably well-established terms like 'nasal' and 'harsh' are discarded in favour of a writer's own impressionistic terms, which give a misleading air of accurate, specific, and communicable precision. The link, if any, between personality factors and voice quality becomes that much harder to analyse when the descriptive terms used lack currency, and specific conclusions harder to justify. This colours the validity of comments about personality attributes such as made by Greene (1941), when he wrote that

'The psychophonasthenic is a variant of the hysterical personality. He is abnormally suggestible, but usually cultured, intelligent, and well-educated. On the other hand, he is unsocial, retiring, and aloof. The voice is tremulous, pinched, and grating, with pitch irregularities. It cracks frequently (....) breaking to a lower key or choking off completely'.

The other type of research in this area doesn't try to describe the voices used as stimuli, but focuses on the ability of judges to make accurate diagnoses of the personality underlying the voice. Fay and Middleton (1939b) show that speakers can judge Spranger personality types from voice alone, as well as intelligence (1940a), but not sociability (1941b), lying (1941a), introversion (1942), nor leadership (1943).

If it is true that information about personality is indicated
by voice quality, then the information must be chiefly carried by
the effects of extrinsic settings, which are learnable, rather than
by the invariant intrinsic features.

The indication of a speaker's affective state by the extrinsic
vocal features of paraphonology does not directly concern voice
quality, since, like language, paralanguage is a matter of an
independent, arbitrary, culturally-relative code (Abercrombie,
1968; Laver and Hutcheson, 1972; Laver, 1975). But they have
an interesting connection in one indexical respect: voice quality and
paraphonology can quite often involve the same auditory phenomena,
and when this happens, listeners sometimes mis-allocate the auditory
feature concerned. For example, laryngitis or a heavy cold often
result in a phonation type that sounds very similar to whispery voice.
Most of us have probably had the experience of suffering from this
condition, and having listeners reply to one's intrinsically whispery
voice in whispery voice or whisper themselves, mistaking the physical
index for an affective one, joining in the conspiratorially confidential
interaction they thought was being signalled. Thus it may well
also be the case that conclusions drawn by listeners about a speaker's
personality can be influenced to some degree by the auditory similarity
of features in his habitual voice quality to those of paraphonology.

3. The indication of social attributes by voice quality

Social behaviour is largely learned behaviour. Because of this,
indexical clues in voice quality to social information about a
speaker must lie in the features which can be acquired by imitation,
the extrinsic features. Extrinsic settings which are characteristic
of a particular accent can then act as a clue to social factors that
are typical of speakers of that accent (Laver, 1968: 50).

In this way, voice quality serves to indicate factors such as regional origin and social status, where these are associated with a particular accent. The comments noted earlier, at the beginning of the chapter on the descriptive system, on the characterization of speakers of different languages by their articulatory settings, by Wallis, Wilkins, Holder, Cooper, Webster, Sweet and Honikman, are relevant here.

Nasality is one extrinsic setting that is often characteristic of accents of a particular region: it typifies speakers of Received Pronunciation from England, and of many regions in the United States (Bullen, 1942; Johnson, 1951) and Australia. Similarly, velarization acts as a regional marker in the speech of speakers from Liverpool and Birmingham, and from parts of New York.

Voice quality can also act as an index of membership of a social group that is not necessarily an accent-group. Luchsinger and Arnold exemplify what they refer to as 'conscious affectation of a misunderstood speaking style' by commenting that

'The old speech pathologists, notably H. Gutzmann, Sr. (1901), and Nadoleczny (1929), quoted the examples of habitually nasal speech among many Prussian Imperial Guard lieutenants, and the widespread nasality among priests and pastors in the eighteenth century, of whom it was said "Humilitatis gratia nasalitatem affectant". (For the sake of humility, they affect nasality)'

(Luchsinger and Arnold, 1965: 666).

Occupation can be indicated in two ways: by an extrinsic setting
voluntarily acquired as an index of membership of the particular occupation, as in the case of the Prussian Imperial Guard lieutenants, or by the effect of practising the particular occupation on the intrinsic vocal apparatus of the speaker, as in the case of laryngeal damage by vocal abuse. The two categories sometimes include members of the same profession; military drill sergeants, for example, seem characteristically to have harsh voices, - these are either the direct result of habitual vocal abuse, or are perhaps acquired by imitation, in the hope of projecting the typical persona of their profession. Harsh voice, and harsh whispery voice, is the occupational hazard and index of a number of professions with a similar demand on the larynx: singers and actors (Brodnitz, 1954, 1962, 1964), clergymen (Craig and Sokolowsky, 1945; Russell, 1936), executives (Gardner, 1958, cited by Luchsinger and Arnold, 1965: 140), and cheerleaders (Jensen, 1964, 1965). The rather bizarre example of an effect on the voice by the practice of a profession may be recalled from earlier in this chapter, in the laryngocele suffered by the blind criers of the Koran, and by some drillmasters, noted by Napoleon's Surgeon-in-Chief.

According to Fay and Middleton (1939a), Pear (1931) and Herzog (1933), listeners are moderately good at guessing occupation from clues in voice alone, presumably, as Fay and Middleton suggest, because occupational stereotypes exist.

4. Stereotyped judgements in voice quality

We all act, as listeners, as if we were experts in using indexical information in voice quality to reach conclusions about physical, psychological and social attributes of speakers (Laver, 1960: 50).
Long experience of inferring such attributes from voice quality, presumably often successfully confirmed by information from other sources, encourages us to invest our implicit ideas about the correlations between voice quality and indexical information with an imagined infallibility. It is worth questioning the validity of this judgmental process. We draw indexical conclusions, and we act on them, shaping our social interactions by their influence; but is the information we infer accurate, or is there the possibility that it is quite false? There is a good deal of evidence that in these subjective judgements we operate with stereotypes (Cantril and Allport, 1935; Eisenberg and Zalowitz, 1938; Fay and Middleton, 1939b, 1940b; Kjeldergaard, n.d.; Licklider and Miller, 1951; Starkweather, 1964). Listeners, if they are from the same culture, tend to reach the same indexical conclusions from the same evidence, but the conclusions themselves may, on occasion, bear no reliable relation to the real attributes of the speaker (Laver, 1968: 50 - 51).

Of the three types of indexical information in voice quality, physical, psychological and social, it is the physical information which probably tends to lead to the most accurate conclusions, especially as to sex and age. Psychological and social conclusions run a higher risk of error because of their culturally-relative nature, and because they derive from the more variable strand in voice quality, the extrinsic settings, as opposed to the invariant intrinsic aspect (Laver, 1968: 51).

C. Voice quality and phonetic quality

A position that has been maintained throughout this thesis is that voice quality and phonetic quality are complementary concepts,
and partly mutually defining. It was said in the Introduction that 'Since phonetic quality is the prime datum for the study of the spoken mode of language, the crucial part played by voice quality in the definition of phonetic quality is very relevant to any general phonetic theory, and hence to any well-founded general linguistic theory'. The intention of this concluding section is to try to elucidate the relationship claimed to exist between the two sorts of quality, by placing them in an overall semiotic view of spoken communication.

A useful point of departure into further discussion of semiotic aspects of communication is a description of the elements of a communication system. There are at least five elements (themselves made up of sub-components): a producer of a transmission containing signals; the transmission itself, which in a moment will be re-named the medium; the signals themselves, as certain aspects of the transmission; a communication-channel, capable of being manipulated by the producer to produce a transmission which can be projected through space and/or time; and a receiver of the transmission capable of abstracting the signals from the transmission. There is also the information conveyed by the signals, which is in a sense external to the communication system as such.

In the case of spoken language, the producer is the speaker; the transmission is the phonic totality of the utterance; the signals are particular patterned aspects of sound within the utterance; and the receiver is the listener.

In speech, the signals which are transmitted do not themselves exhaust the totality of the transmission, in the sense adopted here of the term signal. A signal is the patterned aspect of a transmitted
utterance which acts as a *linguistic symbol* or a *paralinguistic symbol*: there are other aspects of the transmitted utterance which are not part of the symbolic signal as such, - the aspects to be attributed, for example, to extrinsic and intrinsic features of voice quality.

The most interesting of the five suggested elements in the communication system, as far as this study of voice quality is concerned, is the *transmission*. A transmission of this sort is essentially an artefact created by the producer to achieve communication with the receiver by means of a symbolic code. The artefactual nature of the transmissions necessary to achieve linguistic communication is the essence of Abercrombie's concept of a *medium*. He writes that a piece of spoken English and a piece of written English

'are the same *language* embodied in different *mediums*, one medium consisting of shapes, the other of noises.

It is possible for the same language to be conveyed by different mediums because the language itself lies in the *patterns* which the mediums *form*, and not in the physical objects or events, as such, of which the medium consist. When we distinguish language from medium, what we are doing is to distinguish a pattern from its material embodiment, of which, in a sense, it is independent. Language, we could say, is *form*, while the medium is *substance*' (Abercrombie, 1967: 1).

Abercrombie amplifies the artefactual nature of a medium:

'One thing all mediums have in common is that they *mediate* between the producer and the receiver of language. Thus every medium has associated with it two sorts of human activity: a producing activity from which the medium results, and a receiving activity by which the medium is apprehended. The first involves acts of mobile organs, the second involves acts of a perceiving sense. When one person communicates with another by
language, the activity of the former is not directly perceived by the latter. What is perceived, and what affects the communication, is the product of the activity, the artefact, either noises or shapes - the medium, that is to say. The contact is indirect' (Abercrombie, 1967: 3).

This is exactly the sense of the term 'transmission' used here, and 'medium' will now normally be used in its place. It was helpful to begin this section with 'transmission' rather than 'medium', though, because many writers use the term 'medium' not in the sense borrowed here from Abercrombie, but as the equivalent of the term 'communication-channel', which is here sharply differentiated from the idea of a medium. We would want to be able to say, for example, that the spoken medium exploits both the auditory and the visual communication-channels, in that the productive acts of articulation can be seen, as well as their acoustic results being heard. Earlier, for example, it was suggested that a descriptive label for a voice such as 'orotund' had a visually iconic element in that a part of the articulation of the beginning of the word was visibly similar to the characteristic lip-posture of the type of voice to which reference was being made. The separation of the concepts of a medium and a communication-channel facilitates this sort of discussion.

I would like now briefly to consider the relation between a medium and the symbolic signs used in language.

In his definitions of icon, index and symbol, Peirce saw each of these categories as signs referring to objects. This attitude that reference is a special semiotic relation holding between a sign and an entity which can be empirically observed, and that the relation
cannot apply to entities which are not empirically observable, is one which is held by a number of philosophers and linguists (cf. Lyons, 1963: 1-4; 1968: 425-427). However, as the present objective is only to explore the relation between a symbolic sign and a medium, the notion of reference can be used in a looser sense: this sense is the equivalent of 'standing for' in Peirce's neutral definition of a sign as 'standing to somebody for something in some respect or capacity' (2.228). It will be terminologically convenient here to say that a symbol refers to its referent, without concern for the observability of the referent, (though to qualify as a symbol, a sign must have an arbitrary, conventional relation with its referent).

Language is sometimes said (simplistically) 'to consist of a system of conventional signs', which implies by the definitions used above that language is a symbolic process. It is probably true to say that language is primarily a symbolic process, as 'symbol' is here defined, but there is one kind of linguistic unit that cannot in any circumstances be said to be symbolic, because it has no reference: that unit is the most abstract molecular unit at the phonological level. In a phonemic view of phonology, this unit would be the phoneme. What could a phoneme be said to symbolize, or to what could it be said to refer? It is not particularly useful to think of a phoneme as a symbol referring to itself.

Hjelmslev recognized that a phoneme was neither a symbol nor a sign, when he called this type of entity a 'figura', and maintained that such phonological units were cenemes, or 'empty' units, and distinguished them from pleremes, or 'full' units (Hjelmslev, 1953);
see also Hockett, 1953). This unique status of the most abstract molecular phonological level in the theory of language arises from what many linguists have regarded as a principal defining characteristic of language, - that of having two distinct levels of structure (variously called 'double structure', 'double articulation', or 'duality of structure', (Hockett, 1958: 574-575; Lyons, 1972: 65)).

These two levels of units in all languages are the phonological (cenematic) level, not capable of reference, and the grammatical (plerematic) level, made up of morphological, lexical and syntactic units, all capable of reference 4. The sole function of the always very limited number of phonological units (cenemes) in each language is to act as molecular units capable of being combined into grammatical symbols (pleremes), which can then be further concatenated into a potentially infinite number of unique sequences (cf. Lyons, 1968: 54).

All that is necessary for the cenemes of a language is that they should be able to be distinctively encoded by the producer and distinctively identified by the receiver of the transmission. Their essence is thus that they serve as distinctive building blocks for combination and concatenation into the grammatical pleremes of the language.

To achieve perceptual distinctiveness, cenemes as abstract entities have to have a manifestatory connection with a physical

---

4. I am aware that in borrowing the terms 'ceneme' and 'plereme' from Hjelmslev and Uldall's glossematic theory, I have simplified their position. I am also aware that a loose definition of 'reference' becomes unsuitable here for many cases of plerematic symbolization, when the pleremes are not lexical items referring to observable objects. However, my point in both cases is merely to emphasize the dual level of symbolization.
medium. In the spoken mode of language, this connection is with the patterned aspects of a transmitted utterance referred to above as signals. The signals were said to act as linguistic symbols, and by this was meant that they 'stand for' ('refer to', 'manifest', 'embody', 'expound') or symbolize the cenematic units. The phonemic level of phonology is seen in this view as acting as a bridge between two major categories of linguistic symbols: the patterned aspects of speech here called signals, and the plerematic units of grammar. The signals symbolize the cenematic units internal to language, and the pleremes, made up of cenemes, symbolize semantic aspects of the world external to language.

This is acknowledged to be a simplistic view of the symbolic structure of language; it has been introduced into the discussion here for a very important reason, however, - the relationship between cenematic units and the physical artefacts which serve to symbolize the cenemes is crucial to any theory of linguistics that concerns itself with the pragmatic fact that people use language and that linguists sometimes need to observe them doing so. The relationship between ceneme and medium is vital for this sort of linguistic theory, because it is through the successful analysis of this relationship, and only through it, that we can gain any observational access to the cenematic and thence to the referential plerematic complex of language. A pre-requisite in the linguistic phonetician's success in establishing the link between the relevant signal in the medium and the particular cenematic unit is his ability to perceive the signal in the first place. This sounds self-evident. But the signal is not a simple percept. It is an abstraction of a particular
pattern of phonetic quality (leaving aside for the moment the parallel phonetic factors of pitch, loudness and duration). Phonetic quality is often treated in writings on linguistics as if it were 'raw data', as part of the 'directly observable', 'real', 'concrete' world. One of the fundamental attitudes repeatedly asserted in this thesis is an insistence that this is not so. The phonetic level of analysis is a level of considerable abstraction from the phonic artefacts produced by each act of speaking, as was very clearly pointed out by H.E. Palmer (1930) and Daniel Jones (1938). The same is true of the abstract status of voice quality. Every analytic decision about phonetic quality entails a complementary decision about voice quality, (and of course vice versa). Every decision about either is the product of a number of steps of analytic abstraction from the phonic artefacts of the utterance concerned.

Vehemence of assertion is not a satisfactory substitute for rational persuasion, though, and the following section will be devoted to a consideration of the relationship between voice quality and phonetic quality in more detail, and of their place in the semiotic process of spoken communication.

1. Semiotic analysis of the phonic medium

A fruitful way to try to elucidate the relationship between voice quality and phonetic quality is to trace the network of complementary distinctions that need to be drawn between different sorts of vocal features in an exhaustive analysis of the phonic medium.

A summary diagram of the relations between some of the distinctions about to be drawn is provided in Figure 15.
Figure 15. Summary diagram of the typological distinctions between individual features in the phonic medium.
The phonic medium is the totality of the vocal sound production by any speaker of any language during speech. The basic premise of this discussion is that the phonic totality can be completely accounted for, with no uncategorized residue, by the typological distinctions put forward.

The first distinction in the phonic medium is one already discussed at some length in earlier chapters: the distinction between intrinsic features dependent solely on the normally invariant physical foundation of the speaker's vocal equipment, and extrinsic features, made up of all aspects of vocal activity under the potential volitional control of the speaker.

The sub-division of the intrinsic physical foundation of the vocal apparatus has already been outlined in some detail in the earlier section on the indication of physical attributes by voice quality.

The extrinsic features are made up initially of two different categories: unique features such as a momentary clearing of the throat to remove a build-up of mucus on the vocal folds, which are not of interest to this study, and are to be discarded from analytic attention; and recurrent features, which are the central focus of the whole analytic machinery of linguistic phonetics, and also, in this account, of one aspect of the analysis of a speaker's voice.

The recurrent extrinsic features, without which there could be no concept of phonetic sameness, and whose meaning is surely self-evident, are divided into two major types of vocal features. The first is what we might call the exponent features. These are all
and only the vocal features which act as exponents ('signs', 'realizations', 'manifestations') of phonological and paraphonological units. The second type of recurrent features is the extrinsic component of the voice, the extrinsic settings, to which we might give the name concurrent features, since they exist as a concurrent background, as it were, to the more momentary exponent features of phonological and paraphonological units.

It is convenient to pause here to consider the output of the tree diagram classifying the relationships between the vocal features distinguished so far. There are four classes of features as the terminal elements in the diagram as presently constructed. They are: intrinsic features of the voice; extrinsic features of the voice; extrinsic phonetic features expounding phonological and paraphonological units; and the unique events that may happen in speaking that are relevant neither to the phonetic material of the utterance nor to the speaker's voice, and are therefore excluded from further participation in the analysis. We are left with three classes of features, in an asymmetric classification: two different types of voice features, intrinsic and extrinsic; and one category of phonetic features.

For a moment, let us confine our attention to the linguistically-interesting category, the exponent phonetic features. Although phonological and paraphonological units have been mentioned here, this has only happened in connection with their phonetic exponents. No distinction between 'phonological features' and 'paraphonological features' has been proposed under any of the nodes in the diagram which are dominated by the node labelled extrinsic features.
There are two reasons for this. The first is that this would be a confusion of levels of analysis: the classification is one of distinctions able to be drawn between aspects of substance (aspects of the medium, that is to say), not of form. The second reason is another facet of the first, and arises from the espousal of a particular theoretical position: that is, in a properly general linguistic and paralinguistic theory, phonology and paraphonology are to be distinguished not by the superficial fact of the type of vocal feature by which their units are phonetically expounded, but by their semiotic function, which is independent of the nature of their exponents.

The generality underlying this position is that in a general theory wider than the scope of any given language, the phonetic features available for symbolizing the units of phonology and paraphonology are very widely one and the same set of features. There will perhaps be tendencies for paraphonology to exploit longer-term aspects of the phonetic features than the characteristically shorter-term aspects exploited by phonology, but no necessary difference of principle exists between the phonetic parameters chosen. Perhaps a brief illustration will clarify this: considering the phonetic parameter constituted by the limits within which pitch-movements are constrained to take place in English, phonology and paraphonology both exploit the parameter, in slightly different ways. In Halliday's phonological analysis of tone in English, the terminal tendencies of two of his suggested tones are 'high' for the rising tone 2, and 'mid' for the rising tone 3 (Halliday, 1963). In the paraphonological indication of anger in English, the 'characteristic range of notes, or compass, within which the pitch fluctuation' of a speaker's voice takes place can be raised (Abercrombie, 1967: 99).
So that paraphonology sets slightly longer-term limits to the pitch-movements, and phonology sets more momentary values for the limits of the particular pitch-movement that is the exponent of the particular tonal unit.

If we return now to the analysis of different aspects of the phonic medium, Jones (1938: 2), in his consideration of the concrete or abstract nature of sounds, says that he 'will start from Mr. Perera's hypothesis that a sound is a non-physical "thing" which possesses or can possess certain physical attributes, namely quality (timbre), length, loudness and pitch'. This conception of 'quality' as a physical attribute of a sound is not one sympathetic to the views expressed throughout this thesis. The concepts of 'a sound' and of 'quality' are here seen as notions which are both abstractions of some considerable degree. The same is true of the concept of 'a feature': an exhaustive account of different sorts of features in the phonic medium, 'with no uncategorized residue', is here a matter of a complete typological account of abstract 'features' or 'aspects' of speech, not of any sort of exhaustive arithmetic subdivision. The complementarity of voice quality and phonetic quality is also typological, not arithmetic. But the collocation, in Jones' discussion, in the context of the abstract nature of some aspects of speech, of quality, length, loudness and pitch, prompts another (somewhat less abstract) range of distinctions of phonic aspects that may help us to get closer to the relationship between voice quality and phonetic quality.

We can say that sounds in the phonic medium have two basic types of attributes, - those of quality and those which we might
call dynamic attributes (length, loudness and pitch).

Symmetry in descriptive theoretical structures should be regarded with cautious suspicion: but the distinction just mentioned, between 'quality' features and 'dynamic' features, suggests a tempting analytic symmetry in the further discussion of the phonic medium. We are concerned now with the basic dichotomy between phonetic features and voice features. It seems profitable to suggest that these categories can each be divided into quality features versus dynamic features, to give the concepts of phonetic quality and voice quality, and phonetic dynamics and voice dynamics.

If we consider first the concepts of phonetic quality and voice quality, then they differ partially of course in that voice quality has an intrinsic component that is lacking in phonetic quality as such. But it is then very striking that, at the next lower level of description, phonetic quality and voice quality can both make use of precisely the same recurrent features. Supralaryngeal, laryngeal, and overall tension factors all contribute potentially to both phonetic quality and voice quality. Beyond this level of course, phonetic quality and voice quality diverge, in that voice quality derives from longer-term aspects of the supralaryngeal, laryngeal and overall tension factors, while phonetic quality is served normally by very much shorter-term aspects of their activities, whether phonetic quality is acting here as the exponent of phonology or paraphonology.

The matter of phonetic dynamics and voice dynamics is more problematic. Phonetic dynamics are straightforward enough; they are the extrinsic vocal aspects dealt with by Abercrombie (1967) under the name of 'voice dynamics'. It is a pity that a terminological
suggestion here completely inverts an established usage; it is necessary to make a particular point, and no necessary permanence is urged for the change. Phonetic dynamics in this suggested usage could be seen as made up of temporal dynamics (tempo, continuity, rhythm), pitch dynamics (pitch-range, pitch-movement), and loudness dynamics (loudness-range, loudness-movement). Voice dynamics as outlined here would stand in the same relation to phonetic dynamics as does voice quality to phonetic quality. That is, they are made up of not only an extrinsic component, like their phonetic counterpart, but also of an intrinsic one. This intrinsic component would include the physically-determined limitations on such features as maximum pitch-range and maximum loudness-range. Intrinsic constraints on temporal aspects are more tenuous; they presumably set a limit on maximum tempo, for instance. The extrinsic component of voice dynamic features would consist, as in the extrinsic component of voice quality, of habitual settings of the dynamic features. Values of dynamic features which fulfilled an exponent function for phonological or paraphonological units would then be a matter of phonetic dynamics, and those which were habitual and irrelevant to phonological or paraphonological communication as such would be allocated to voice dynamics.

The point in setting up this four-way relation between voice quality and phonetic quality, and voice dynamics and phonetic dynamics, is to urge the relevance of voice quality and voice dynamics as here defined to general phonetic theory. Two points are salient: excluding the intrinsic components, voice features and phonetic features can be argued to exploit very largely the same vocal parameters. Secondly, just as in the case of phonology and paraphonology
argued earlier, the decision about whether a given vocal feature is a voice feature or a phonetic feature cannot be taken on the grounds of the phenomenal characteristics of the feature: the decision is fundamentally a semiotic one, concerned with the role of the vocal feature in the semiotic process of spoken communication. In this view, the credentials of voice quality and voice dynamics to consideration as having relevance to general phonetic theory, and therefore to general linguistic theory, seem not unacceptable.

2. Voice quality and accent

It is fair to ask, if the relevance of voice quality to general linguistic theory, and in particular to linguistic phonetics, is conceded, how such a concession could change any practical element of a linguistic phonetician's professional activities in a profitable way.

One field where a most effective application has already been elegantly demonstrated is that of sociological linguistics, in Trudgill's sociological studies (1973; 1974) of the urban dialectology of English in Norwich, with regard to its social differentiation. I think it is not mere flattery to say that his work in this area shows the shape of things to come in the sociological development of linguistic theory. He develops a

'diasystem the aim of which is to produce all forms of Norwich English from a common underlying base. The diasystem comprises a single underlying systematic phonemic system, and (a number of different types of rules)' (Trudgill, 1974: 191)
One type of rule incorporated in his diasystem is a type specifying particular articulatory settings of the sort discussed at length in this thesis. Trudgill properly regards them as 'ancillary to the phonetic realization rules' (loc.cit.). He characterizes the value of including extrinsic setting components of voice quality in the phonological specification of accents as follows:

'If we incorporate (....) rules of this type into the Norwich diasystem, (....), the statement of both the phonetic realization and realization level mutation rules can be much simplified. (....) Another advantage of setting rules is that they can relate different types of Norwich English to each other in the diasystem in a much more generalized and significant way than a series of individual rules. Different social types of Norwich English may be characterized by the presence or absence of, say, rule 100 (a setting rule, J.L.), rather than by a whole series of rules. This is an important point, since it is clear that perhaps the single socially most significant feature of linguistic differentiation in Norwich is the type of voice quality produced by the particular type of setting employed by a speaker. It is in any case this feature which most clearly distinguishes WC from MC speakers (working class from middle class, J.L.). This point, of course, did not emerge at all from our atomistic analysis of the co-variation of linguistic and sociological phenomena' (Trudgill, 1974: 190-191).

The productive use that Trudgill has thus been able to make of the extrinsic setting component of voice quality in an application not only of direct sociological relevance, but also of immediate linguistic relevance, buttresses the attitude espoused in this thesis about the phonetic legitimacy of an exploratory excursion into the analysis of voice quality. It lends validity to
the view expressed earlier that perhaps the proper study of phonetics extends not just to the features which have been found to act as linguistic and paralinguistic signs, but can reach beyond these semiotic limits to the definition and description of all extrinsic signs used in spoken communication. Nor does such an expansion of phonetic horizons inevitably threaten a dilution of the fundamental linguistic motivation that is widely accepted as a hallmark of British phonetics. As Trudgill has shown, a readiness to explore the frontiers of one's subject can sometimes provide an enriched view of some of its most central areas. In this particular case, the notion of 'an accent' can now be given a more productive description, through the use of theoretical apparatus which has gained in linguistic and sociological power by drawing on concepts whose relevance to linguistics has previously often been denied, by figures of the linguistic stature even of Sapir, for instance, as noted in the introduction to this thesis.

CONCLUSION

A number of attitudes have become salient in the course of this investigation into phonetics and semiotic aspects of voice quality. One such attitude emphasizes the link between the two aspects: an interest in linguistic phonetics is one facet of a wider interest, that of the uniquely human curiosity about the semiotic process of communication.

Another attitude that has been evident is that an apparently inward-looking interest in the history of our subject can be a source
of strength for the present character of the subject. Sweet (1881), Firth (1946), Abercrombie (1948) and Jones (1943) have all invested considerably more than a passing attention to the work of past phoneticians: insights of very modern relevance can often be seen to have a surprisingly long genealogy. The contributions to some basic notions discussed in this thesis made by phoneticians such as Wallis and Wilkins in the seventeenth century are often of substantial value. In this general connection, Abercrombie makes a comment in his article on 'Forgotten phoneticians' very much in keeping with the historical aspects of this thesis. He writes that

'A recent textbook on phonetics mentions "Sir Henry Sweet" among "other early writers on the subject". My witness to Professor Firth's thesis - that our antecedents are older and better than we think - have been minor ones, but they are more, I think, than curiosities. Sweet deserved a knighthood as much as anyone, but we should remember that a good three centuries of lively interest in phonetics preceded him' (Abercrombie, 1948, reprinted in Abercrombie, 1965: 75).

The attempt to provide descriptive categories for phonetic aspects of voice quality, and to incorporate the descriptive system into accepted general phonetic theory, may one day perhaps lead, after more work, to the possibility of extending the training of the sensory skills of the phonetician to something of the same degree of sensitivity to voice quality as is presently a standard expectation in his professional ability to perceive distinctions of vowel quality, for instance.

The reward in conducting the work reported in this thesis does not lie solely in the fact that it is always interesting to discern
generalities common to all men, but also in that we tend to learn more about some quite fundamental characteristics of our own subject when we explore the more tenuous areas in the boundaries of the subject. This is because, in pushing outwards from the central base of the discipline, it is usually necessary to examine some of the most fundamental assumptions of established general phonetic theory, - assumptions which, precisely because they are so basic, often lie unquestioned, unexamined, and sometimes unsuspected and therefore unarticulated, for long periods. The renewed realization of the strength and validity of these fundamental assumptions, when they are periodically re-scrutinized, is a salutary professional experience.

Lastly, in seeing phonetics set in the broader frame of semiotics, as has been possible in this study of voice quality, it is re-assuring to be reminded again that the concerns of phonetics are not the construction of some elaborate game of taxonomy, but the apprehension of a very central aspect of human communication. In knowing more about phonetics, even in the study of voice quality as one of its less central, less traditional pre-occupations, we learn more about how we construct and sustain our identity as social beings.
REFERENCES


Anthony, J. and Lawrence, H. (1962) A resonance analogue speech synthesiser Proceedings of the 7th International Congress on Acoustics Copenhagen


Ascham, R. (1570) The Scholemaster London

Austin, G. (1806) Chironomia; or a Treatise on Rhetorical Delivery T. Cadell and W. Davies: London


Benson, J. (1951) An experimental study of the relationship between amount of nasal emission of air and judged nasality M.A. Thesis West Virginia University.


Davis, E.B. (1941) *Nasal Twang* *le maître phonétique*, 56: 4-5

Delattre, P. (1951) Les attributs acoustiques de la nasalité vocalique et consonantique *Studia Linguistica*, 8: 103-109
Delestre, J.-B. (1866) *De la Physiognomie* Renouard: Paris


Fant, G. (1962) Descriptive analysis of the acoustic aspects of speech *Loco*, 5: 3-17

Fant, G. (1964) Formants and cavities *Proceedings of the Fifth International Congress of Phonetic Sciences*, Münster


Fay, P. J. and Middleton, W. C. (1930b) Judgment of personality types from the voice as transmitted over a public address system. Character and Personality, 8: 144-155


Fay, P. J. and Middleton, W. C. (1940a) The ability to judge the rested or tired condition of a speaker from his voice as transmitted over a public address system. Journal of Applied Psychology, 24: 645-650

Fay, P. J. and Middleton, W. C. (1940b) Judgment of intelligence from the voice as transmitted over a public address system. Sociometry, 3: 186-191

Fay, P. J. and Middleton, W. C. (1940c) The ability to judge truth-telling, or lying, from the voice as transmitted over a public address system. Journal of General Psychology, 21: 211-215


Foerster, R. (1893) *Scriptores Physico-norici* Leipzig

Franke, F (1899) *Die Umgangssprache der Niederlausitz in ihren Lauten* *Phonetische Studien*, 2: 21-60

Frederickson, J.M. and Ward, P.H. (1962) *Laryngocele Ventricularis* *AMA Arch. of Otolaryng.*, 76: 560-573

Freud, E.D. (1962) *Functions and Dysfunctions of the ventricular folds* *Journal of Speech and Hearing Disorders*, 27: 334-340


Frohkin, V.A. (1965) Some phonetic specifications of linguistic units: an electromyographic investigation *Working Papers in Phonetics*, 3, Department of Linguistics, University of California at Los Angeles


Greene, J.S. (1941) *Psychophorasthenia Syndrome* *Annals of Otolaryngology, Rhinology and Laryngology*, 50: 1177-1184

Gutzmann, H., Sr. (1901) Von den verschiedenen Formen des Halses
Halle/Saale

Haberman, F.W. (1954) English sources of American elocution in
Wallace, K.R. (Ed.) History of Speech Education in America, pp. 105-
126 Appleton-Century-Crofts: New York

Hahn, E.F., Lounas, C.W., Hargis, D.E. and Vandraegen, D (1952)

of Speech Education in America, pp. 219-237
Appleton-Century-Crofts: New York

Halliday, M.A.K. (1963) The tones of English Archivum Linguisticum,
15: 1-28

Hardcastle, W. (1975, forthcoming) Physiology of Speech Production
Mouton: The Hague

Harrington, R. (1944) Study of the mechanism of velopharyngeal
closure Journal of Speech Disorders, 9: 325 -

Hattori, S., Yamamoto, K. and Fujimura, O. (1956) Nasalization of
vowels and nasals Bulletin of the Kobayasi Institute of Phys.
Research, 6: 226-235

Heffner, R.-M.S. (1950) General Phonetics University of Wisconsin
Press: Madison


Hertneg, E. (1933) Stimme und Persönlichkeit (Voice and personality)
Zsch. f. Psychol., 130: 300-379

Hess, D.A. (1959) Pitch, intensity and cleft palate voice quality
Journal of Speech and Hearing Research, 2: 113-125

Neerl. 5: 73 -

Hixon, E.H. (1949) An X-ray study comparing oral and pharyngeal
structures of individuals with nasal voices and individuals with
superior voices M.A. Thesis, State University of Iowa

Hjelmslev, L. (1953) Prolegomena to a theory of Language, Whitfield, F.
(trans.), Indiana University Press: Bloomington

Hockett, C.F. (1958) A Course in Modern Linguistics Macmillan:
New York


Proceedings of the 7th International Congress of Phonetic Sciences
Montreal: pp. 320-331


Husson, R. (1950a) Etudes des phenomenes physiologiques et acoustiques de la voix chantee pendant la phonation Thèse: Paris


Husson, R. (1955a) Données expérimentales nouvelles concernant la physiologie de la phonation Exposés Annuelles d'Oto-rhino-laryngologie
Husson, R. (1955b) Physiologie phonatoire du larynx Encyclopédie Médico-Chirurgicale, Oto-Rhino-Laryngologie, pp. 11- Paris

Husson, R. (1956) Uber einige neue Ergebnisse der Kehlkopfphysiologie und pathologie Arch. Chr.-, Nas., u kehlkheilk, 169: 176-182


Husson, R. (1962) Physiologie de la phonation Paris


Jensen, P.J. (1965) Phonatory Deviation in Cheerleaders NIH Progress Report, Grant No. NB 01,056-01


Joos, K. (1948) Acoustic Phonetics Language 24, Supplement


Kins, J. (1962) Trouble de la voix dans le sillon des cordes vocales Folia Phoniatrica, 14: 272-273

Kjolbergaard, P.I. Changes in perceived personality traits as a function of manipulations of vocal characteristics Harvard University, mimeographed


Kretschmer, E. (1925) Physique and character New York


Ladefoged, P. (1964) Some possibilities in speech synthesis Language and Speech, 7: 205-211


Landes, B.A. (1953) An investigation of the relationships between vocal harmonics and listeners' discrimination of fundamental pitch M.A. Thesis, Purdue University

Laver, J. (1964) The synthesis of voice quality PAT Report, University of Edinburgh


Laver, J. and Hutcheson, S. (Eds.) (1972). Communication in face to face interaction Penguin: Harmondsworth

Lawrence, W. (1953) The synthesis of speech from signals which have a low information rate in Jackson, W. (Ed.) Communication Theory, Butterworth: London


Locke, J. (1690) An Essay concerning Human Understanding London


Michel, J. (1964) Vocal Fry and Harshness. Progress Report of Communication Sciences Laboratory, Department of Speech, University of Florida, for National Institute of Neurological Diseases and Blindness, National Institutes of Health, under Grant NE-043,378, "Pitch characteristics of adult males."


Miller, J.E. (1964) Decapitation and recapitation, a study of voice quality *Paper 38 at 68th Meeting of Acoustical Society of America*

Montanus, P. (1635) *De Spreackonst* Delft


Moore, W.E. (1939b) Voice quality and anxiety *Journal of Speech Disorders*, 4: 33-


Moses, P.J. (1941) Theories regarding the relation of constitution and character through the voice *Psychological Bulletin, 38*: 746 (A)

Moses, P.J. *The Voice of Neurosis* Grune and Stratton: New York


Perello, J. (1962b) La théorie muco-ondulatoire de la phonation Ann. Otol. (St. Louis), 79: 722-


Podvinec, S. (1952) The physiology and pathology of the soft palate Journal of Laryngology and Otology, 66: 452-

Pellux, J. (1706) Onomasticon Amsterdam


Pressman, J.J. (1942) Physiology of the vocal cords in phonation and respiration Arch. Otolaryng., 35: 355-393


Pronovost, U. (1942) Research contributions to voice improvement Journal of Speech Disorders, 7: 313-318


Quintilian Institutiones Oratoriae, Watson, J.S. (trans.), Quintilian's Institutes of Oratory, or, Education of an Orator George Bell and Sons: London (1879)
Journal of Speech and Hearing Research, 1: 155-168


Rousselot, P.J. (1901) Principes de phonétique expérimentale
Didier: Paris

Rubin, H.J. (1960a) The neurochronaxic theory of voice production—
a refutation Arch. Otolaryngol., 71: 913-920

Rubin, H.J. (1960b) Further observations on the neurochronaxic theory of voice production Arch. Otolaryngol., 72: 207-211


Russell, G.O. (1936) Etiology of follicular pharyngitis, catarrhal laryngitis, so-called clergymen's throat; and singer's nodes Journal of Speech Disorders, 1: 113-122

Sandford, H.P. (1938) English theories of public address, 1530-1628 H.L. Hedrick: Columbus, Ohio

Sapir, E. (1927) Speech as a personality trait American Journal of Sociology, 32: 692-705

Saunders CIDL Corporation: New Jersey


Sheridan, T. (1762) A course of lectures on elocution London

Sheridan, T. (1781) A rhetorical grammar of the English language Dublin


Sievers, E. (1876) Grundzüge der Lautphysiologie (later editions entitled Grundzüge der Phonetik) Leipzig


Smith, S. (1951) Vocalization and added nasal resonance Folia Phoniatrica, 3: 165-169

Smith, S. (1951a) Remarks on the physiology of the vibrations of the vocal cords Folia Phoniatrica, 6: 166-178

Smith, S. (1951b) Vibration period of the vocal cords Film, Part I and Part II Government Film Office, Copenhagen


Smith, S. (1956b) Mouvements des cordes vocales Film No. 4, Government Film Office, Copenhagen

Smith, S. (1957) Théorie aérodynamique de la vibration des cordes vocales in Aubin, Larynx et Phonation


Sweet, H. (1885) Elementarbuch des gesprochenen Englisch


Sweet, H. (1900) The History of Language

Sweet, H. (1908) The Sounds of English


Tarnaud, J. (1941) Traité pratique de phonologie et de phoniatric, la voix - la parole - le chant Librairie Maloine: Paris

Tarnóczy, T. (1948) Resonance data concerning nasals, laterals and trills Word, 4: 71-77

Tarnóczy, T. and Fant, G. (1964) Some remarks on the average speech spectrum Royal Institute of Technology, Stockholm, STL-QPSR, 4: 13-14

Taylor, H.C. (1934) Social agreement on personality traits as judged from speech Journal of Social Psychology, 5: 214-218


Thompson, C.L. (1962) Wavelength perturbations in phonation of pathological larynges Progress Report, Grant MD 01 398 National Institutes of Health, Washington


Travis, L.E., Bender, W.R.G. and Buchanan, A.R. (1934) Research contribution to vocal theory Speech Monographs, 1: 65-71


Van den Berg, J. (1954d) Über die Koppelung bei der Stimmbildung Zeitschrift für Phonetik, Bd. 8: 281-292


Van den Berg, J. (1957a) Subglottic pressures and vibrations of the vocal folds Folia Phoniatrica, 9: 65-71


van Dusen, C.R. (1944) A laboratory study of the metallic voice Journal of Speech Disorders, 6: 137-140


Wallis, J. (1653) Grammatica Linquae Anglicanae and Tractatus de Locula. London


Webster, N. (1769) Dissertations on the English Language. Isaiah Thomas: Boston


Wendahl, R.W. (1964) The role of amplitude breaks in the perception of vocal roughness. ASHA, Convention Abstracts, 6: 406


Wilkins, J. (1668) An essay towards a real character and a philosophical language. John Martyn: London

Willis, R., (1829) Vowel Sounds. Philosophical Society


APPENDICES


Illustrative tape recording of the acoustic synthesis of seventy-two voices with different extrinsic voice quality settings.
ABSTRACT

The outline of a componential descriptive model of voice quality is suggested, in the hope of facilitating discussion about voice quality among phoneticians, linguists, speech therapists, psychologists and psychiatrists. Emphasis is laid on the function of voice quality as an index to biological, psychological and social characteristics of the speaker.

INTRODUCTION

As actors in a social world, we interact with other people by virtue of a constant interchange of information on many different levels. Perhaps the most explicit sort of information exchanged in social intercourse is language, and modern linguistic and phonetic theory has developed some elegant and effective concepts for the description of speech, the spoken medium of language. The description of the more narrowly linguistic aspects of speech does not, however, exhaust the possibilities of information carried in utterances. Any given utterance contains not only linguistic information, but also a great deal of information for the listener about the characteristics of the speaker himself. Abercrombie (1967) refers to the features in speech which convey this information as 'indexical' features. It is the purpose of this article to explore one major vehicle of such indexical information in speech, that of voice quality.

While the concepts available for linguistic description are well developed, it is only comparatively recently that phoneticians have begun to show more than a cursory interest in a descriptive theory of voice quality. For a long time other disciplines such as speech pathology, psychology and psychiatry have been more ready to acknowledge the relevance of the study of this topic. With the current expansion of research in the area, this may be an appropriate time to try to suggest the broad outline of an overall descriptive model of voice quality. One motive for attempting to set up such a descriptive model is to facilitate interdisciplinary discussion of the indexical function of voice quality; another is to incorporate the descriptive model into the wider theory of general phonetics. I hope to be able to show that general phonetics constitutes a legitimate framework for the study of this area, in that it offers an appropriate philosophy of analysis, and can make available an established and meaningful body of relevant descriptive concepts.

There has been a variety of usages in labelling voice qualities in the past. With the exception of some phonetically sophisticated systems developed by speech pathology, the majority of previous systems have used single impressionistic labels for given voice qualities. Typical labels have been "husky", "plummy", "thin", "rich", "velvety", "reedy", etcetera. Such labels are often vague to the point of meaninglessness, except in a metaphorical sense, or as arbitrary imitation-labels. The great advantage of a general phonetic approach to the labelling problem is that, in its systems for describing the physiology of articulation, overall labelling is not attempted. The basic philosophy of phonetic analysis is that composite articulatory
events are broken down into their component parts, and each independent physiological component is separately labelled. Thus, for example, the sound at the beginning of the English word 'fat' is described not as a 'sort of "f"-sound', but as a 'pulmonic, egressive, voiceless, labiodental fricative with velic closure', with each important physiological component analytically isolated.

It is the principal thesis of the descriptive model put forward here that voice quality is similarly susceptible of description in terms of components, and that general phonetic theory can supply the concepts necessary for a physiologically meaningful description of each of the components.¹

**OUTLINE OF A DESCRIPTIVE MODEL OF VOICE QUALITY**

Voice quality, the quasi-permanent quality of a speaker's voice (Abercrombie, 1967, p. 91), can be thought of as deriving from two main sources: firstly, the anatomical and physiological foundation of a speaker's vocal equipment; and secondly, the long-term muscular adjustments, or 'settings' (Honikman, 1964), once acquired idiosyncratically, or by social imitation, and now unconscious, of the speaker's larynx and supralaryngeal vocal tract.

The anatomy and physiology of a speaker determine the width of the potential range of operation for any voice quality feature, and the long-term habitual settings of the larynx and the vocal tract restrict this feature to a more limited range of operation. For example, a man's voice may be physically capable of spanning a wide pitch range; in normal speech, however, he habitually selects a more restricted range within the total possibilities. Basic anatomy and physiology thus determine the possible extremes, and voluntary muscular settings determine habitual ranges between those extremes.

While factors of basic anatomy and physiology are beyond the speaker's control, and the habitual settings are to a certain extent within his control, both these sources of voice quality can transmit indexical information, although of different sorts, as we shall see in a later section.

The anatomy and physiology of the vocal organs, and their habitual muscular settings, are all of legitimate professional interest to the phonetician, but for the moment the habitual muscular settings will be taken as the more direct focus of attention.

It is useful to divide the settings into two groups:

A. Settings of the larynx
B. Settings of the supralaryngeal vocal tract

**A. SETTINGS OF THE LARYNX**

Laryngeal settings fall into three sub-categories:

(a) phonation types
(b) pitch ranges
(c) loudness ranges

(a) Phonation Types. Phonation types constitute an area which is still largely open to exploration. Certainly, some labels exist, and are used more or less widely in phonetics, but much research needs to be done before confident statements

¹ Abercrombie (1967), Garvin and Ladefoged (1963) and Fairbanks (1966) are among previous writers who have suggested a phonetic approach to voice quality. This article owes much, in particular, to Abercrombie's suggestions.
about the detailed physiology of a wide range of phonation types can be made. The phonation types relevant to this article about whose physiology something is known, besides 'normal voice', include 'breathy voice', 'whispery voice', 'creaky voice', 'falsetto voice', 'ventricular voice' and 'harsh voice'. One of the most valuable contributions to current phonetic knowledge about phonation types has been that of Catford (1964), and he has done much to give both physiological and aerodynamic definitions to many of the above labels.

Some combinations of these phonation types are possible, as in 'harsh, whispery voice', 'whispery, ventricular voice', 'creaky, falsetto voice', 'harsh, falsetto voice', 'harsh, whispery, falsetto voice', 'whispery, creaky voice', 'breathy, falsetto voice' and so forth. (See Catford, 1964.)

(b) Pitch Ranges. Pitch ranges within the total possible range in any phonation type can usefully be described on a five-point scale: 'very deep', 'deep', 'medium', 'high', and 'very high'.

Although the total pitch-possibilities for each phonation type can be divided into the five suggested ranges, there seems to be a tendency for speakers using a given phonation type to favour a particular pitch range. This happens, for instance, in the case of falsetto voice, where speakers typically select a high pitch range within the possibilities for falsetto voice, or in creaky voice, where the deeper ranges for creaky voice are often used. In acoustic terms, the absolute frequencies involved in deep falsetto voice and high creaky voice (which show less usual choices of pitch range within the possibilities for the particular phonation types), may overlap considerably. The 'normal voice' phonation type does not show this tendency quite as much as the other phonation types, and the pitch ranges used are much more varied.

(c) Loudness Ranges. Loudness ranges can also be described as selections from a five-point scale: 'very soft', 'soft', 'medium', 'loud' and 'very loud'.

The habitual settings of the laryngeal and associated musculature which result in characteristic pitch and loudness ranges are of a different order from the settings for characteristic phonation types, in that while a person might be recognized solely by his phonation type in the utterance of a single syllable, a much longer stretch of utterance would be necessary for a listener to be able to recognize the speaker by his characteristic pitch and loudness ranges. However, partly because impressionistic labels for voice quality nearly always seem to contain an implicit reference to these features, and partly because of the quasi-permanent rôle of such features in characterising the speaker, pitch and loudness ranges are included in this outline model as features integrally associated with the overall quality of a speaker's voice. In a theoretically more rigorous analysis one would abstract such 'dynamic' features separately from features of 'quality' (Abercrombie, 1967, Ch. 6).

B. SETTINGS OF THE SUPRALARYNGEAL VOCAL TRACT

Supralaryngeal settings of the vocal tract can be divided into four groups, referring to different sorts of modification of the shape and acoustic characteristics of the tract:

(a) longitudinal modifications
(b) latitudinal modifications
(c) tension modifications
(d) nasalisation
(a) Longitudinal Modifications. Longitudinal modifications of the vocal tract can result from vertical displacements of the larynx upwards or downwards, from a neutral position, to give 'raised larynx voice' or 'lowered larynx voice'. Pouting forwards of the lips also effects changes in the length of the longitudinal axis of the vocal tract, and has auditory and acoustic correlates in voice quality.

(b) Latitudinal Modifications. Latitudinal modifications of the vocal tract involve quasi-permanent changes in the cross-sectional area at a particular location of the tract, which result in local constrictive or expansive tendencies. These modifications include: different sorts of labialisation, with the space between the lips being narrowed either vertically or horizontally, or both; two types of pharyngalisation, involving muscular constrictions of the pharynx, or in narrowing of the pharynx by backward displacement of the tongue from its central position in the mouth; and thirdly, settings of the tongue that result in a constrictive or expansive tendency somewhere in the oral cavity.

The settings of the tongue in the oral and pharyngeal cavity are parallel to many of the 'secondary articulations' of traditional phonetic theory. One could speak of overall tendencies towards maintaining a particular 'secondary articulation', but of course such settings are in no sense 'secondary' as far as voice quality, as distinct from segmental features, is concerned (Abercrombie 1967, p. 93). A less prejudicial conceptualisation of the general principle underlying these settings of the tongue, whether in the mouth or in the pharynx, is achieved if any local constrictive or expansive tendency is thought of as resulting from a shift, along one or both of the horizontal and vertical axes of the mouth, of the centre of gravity of the tongue away from the neutral position in which it would lie in an unmodified vocal tract. In this way, pharyngalisation can be said to result from backing in the mouth of the centre of gravity of the tongue; velarisation from backing and raising; palatalisation from raising; alveolarisation from fronting; and a quality which currently lacks a conventional phonetic term, but which is sometimes impressionistically called "hot potato voice", as if the speaker literally had a hot object in his mouth, is the result of lowering, and perhaps backing, of the centre of gravity of the tongue.

For convenience in the labelling system, instead of the more cumbersome 'tongue-raised-and-backed voice', et cetera, the more usual phonetic labels such as 'velarised voice' can be retained, provided that it is remembered that no implications of secondary status enter their definition, which would be in terms of the relative position of the centre of gravity of the tongue.

It is probable, because the tongue is of a relatively fixed volume, that any constrictive tendency in the mouth has a corresponding compensatory expansive tendency in other parts of the mouth and pharynx. Similarly, some longitudinal modifications of the vocal tract probably involve a latitudinal component as well; raised larynx voice, for example, has a component of slight pharyngalisation. It is also probable that longitudinal modifications involving vertical displacements of the larynx and oral and pharyngeal latitudinal modifications of the vocal tract can affect the fine detail of the mode of vibration of the vocal cords within any phonation type, because of the interactions of the different muscle systems involved.

(c) Tension Modifications. In a detailed model of voice quality, account would have to be taken of the effect of variations in the degree of overall muscular tension of the vocal organs on the acoustic damping characteristics of the vocal tract, through factors of radiation and absorption of sound energy in the tract walls. Different degrees of muscular tension may contribute importantly, for
example, to the auditory differences between the qualities of the impressionistically-labelled "metallic voice" and "muffled voice". More research is needed in this area, and factors of overall muscular tension will be omitted from further discussion in this article.

(d) Nasalisation. It is customary, in phonetic usage in this area, to distinguish only between 'nasalised voice' and 'denasalised voice'. In a more detailed model, it would perhaps be necessary to re-scrutinize these two categories, and set up some finer sub-divisions. Speech pathology, for example, has shown that there are a number of auditorily distinguishable kinds of nasalisation arising from a variety of organic causes (Luchsinger and Arnold, 1965). Similarly, the term 'denasalised voice' seems to cover at least two different phenomena: firstly, a quality resulting from habitual velar closure, or, more accurately, from a complete lack of nasal resonance; secondly, the quality which is produced by a speaker with a severe cold in the head and nasal catarrh. The first can legitimately be called 'denasalised voice'; the second, speculatively, might be better thought of as a special case of 'nasalised voice', resulting from a partial or complete blockage of catarrhal mucus somewhere in the nasal cavity or nasopharyngeal sphincter, which would allow the cavity to resonate, but in a highly damped manner. An auditory quality rather like that of velarisation seems also to be a component of such 'cold-in-the-head' voices, perhaps because of some feature of muscular adjustment in the vocal tract walls near the soft palate in this special condition.

In this exploratory article, such sub-divisions will be ignored, and the terms 'nasalised voice' and 'denasalised voice' used in the customary phonetic sense mentioned above.

Features of habitual nasalisation may conceivably affect the setting of the vocal tract as a whole, and thereby the laryngeal settings also; this is because some of the depressor-relaxer muscles of the velum, glossopalatinus and pharyngopalatinus (Kaplan, 1960, p. 188), which serve to open the velic valve, have their point of origin in the tongue and pharynx, and these muscles, to be effective, have to exert a pull, in contraction, against their points of origin.

The use of a quantitative scale is helpful when describing supralaryngeal modifications of the vocal tract. At least, three degrees of modification can usually be auditorily distinguished—'slight', 'moderate', and 'severe'. Thus one may choose to refer to 'slightly pharyngalised voice', 'severely nasalised voice', 'moderately raised larynx voice', and so on.

**Labelling System for Voice Quality**

We are now in a position to offer more specific comments on the suggested labelling system. Instead of single-term impressionistic labels such as "beery", "brassy", "sepulchral", et cetera, we can use composite labels made up of a number of phonetic terms, each specifying a physiologically meaningful component of the voice quality in question. The labelling system concentrates, as a beginning to the problem, on the voluntary muscular settings rather than on the limitations imposed by the basic anatomy and physiology which underlie the settings.

As a convention, features of pitch range and loudness range are described first, then features of supralaryngeal modification, and lastly features of phonation type. Typical labels might then be 'deep, loud, nasalised, harsh voice', 'high, soft, velarised, raised larynx, falsetto voice', 'very soft, nasalised, whispy, creaky voice',
'very high, pharyngalised, harsh, falsetto voice', 'deep, nasalised voice', and so forth.

Medium pitch, medium loudness, an unmodified vocal tract, and the 'normal voice' phonation type can be left to be assumed if no contrary specification is explicitly made. Scalar quantities of different degrees of vocal tract modification can be incorporated, as in 'high, loud, severely nasalised voice', or 'deep, soft, slightly velarised, whispery, creaky voice'. If the degree of vocal tract modification is left unspecified, it could be assumed that a moderate degree was applicable.

It becomes possible to communicate fairly reliably about voices, with a phonetically meaningful descriptive system of this sort. The translation of some impressionistic labels can be attempted, and some illustrative examples are suggested in the following list:

<table>
<thead>
<tr>
<th>Impressionistic Label</th>
<th>Phonetic Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ginny voice</td>
<td>Deep, (harsh), whispery, creaky voice</td>
</tr>
<tr>
<td>Husky voice</td>
<td>Deep, soft, whispery voice</td>
</tr>
<tr>
<td>Golden voice</td>
<td>Deep, soft, slightly nasalised voice</td>
</tr>
<tr>
<td>Piping voice</td>
<td>High, falsetto voice</td>
</tr>
<tr>
<td>Bleating voice</td>
<td>High, loud, (severely) nasalised voice</td>
</tr>
<tr>
<td>Light blue voice</td>
<td>High, soft, raised larynx voice</td>
</tr>
<tr>
<td>Hoarse voice</td>
<td>Deep, (loud), harsh/ventricular, whispery voice</td>
</tr>
<tr>
<td>Gruff voice</td>
<td>Deep, harsh, whispery, creaky voice</td>
</tr>
<tr>
<td>Sepulchral voice</td>
<td>Very deep, pharyngalised, tongue-lowered, lowered larynx, (whispery), (creaky) voice</td>
</tr>
<tr>
<td>Adenoidal voice</td>
<td>Soft, denasalised, velarised voice</td>
</tr>
</tbody>
</table>

Because of the unreliability of reference of the impressionistic labels, not everyone might agree with the suggested translations, but at least the phonetic system, while not yet offering a complete specification, does allow positive statements to be made about assumed components. Communication in writing about voice qualities becomes much more feasible, because the phonetic labels in effect convey instructions for attempts at imitation of the voluntarily controllable components referred to.

**Indexical Information in Voice Quality**

The descriptive model of voice quality put forward here can now facilitate the discussion of the sources of indexical information in voice quality; the information falls broadly into three categories:

A. Biological information
B. Psychological information
C. Social information

A. Biological Information in Voice Quality. Biological information about the speaker, derived from the effects, outside his control, of his anatomy and physiology, itself falls into three sub-categories:

(a) size and physique
(b) sex and age
(c) medical state

(a) Size and Physique.—There seems to be a general correlation between a person's size and physique and the size of his larynx and vocal tract. If we hear a
very deep loud voice over the telephone, we confidently expect the speaker to turn out to be a large strong male; and in general our expectations are fulfilled, within fairly wide margins of error (Moses, 1941, Fay and Middleton, 1940a). Exceptions to this rule are not uncommon, but they take one aback when they occur.

(b) Sex and Age.—One usually forms fairly accurate impressions, from voice quality alone, of a speaker’s sex and age (Mysak, 1959, Tarnaud, 1941, Zerffi, 1957). Deviations from ‘normal’ expectations about the correlation between a speaker’s voice and his sex and age seem to have a powerful effect on impressions of personality.

(c) Medical State.—Voice quality supplies a surprisingly varied amount of indexical information about the medical state of the speaker. It is useful to distinguish between permanent and slightly more ephemeral, although still relatively long-term, medical states. Information about permanent aspects can include details of general health, with crude correlations between, for instance, phonaesthesia, or soft, whispery/breathy voices, and poor health, and between so-called “strong”, “resonant”, or “rich” voices, (deep, loud, (nasalised) voices), and good health. Permanent abnormalities of anatomy and physiology can be revealed by voice qualities associated with cleft palates, deafness, and even exceptional singing voices, which are sufficiently rare to be thought abnormal in this sense.

More ephemeral, but still quasi-permanent states of health can be signalled by voice quality when the speaker is suffering from conditions of local inflammation of his vocal organs, as in laryngitis, pharyngitis, and tonsillitis, and from nasal catarrh, adenoids or a cold.

Other ephemeral factors in voice quality derive from changes in the hormonal state of the speaker, where, for example, these result in changes in the copiousness and consistency of the supply of lubricating mucus to the larynx, and in the characteristics of the mucous membrane covering the actual vocal cords. Such changes occur in the pregnant and premenstrual states in women (Perelló, 1962), and in conditions of sexual arousal in both men and women. These changes often seem to cause slight harshness, and whispery or breathy voice.

Clues in voice quality to other more permanent, but occasionally reversible, hormonal states are sometimes found in the case of voice disorders resulting from diseases of the thyroid, adrenal, and pituitary glands (Luchsinger and Arnold, 1965). Systematic research into the possible use of voice quality as a diagnostic index to these and other medical states would be extremely valuable; so far, the area has only sporadically attracted investigation (McCallum, 1954; Palmer, 1956; Punt, 1959; Sonninen, 1960; Canter, 1963).

Examples of temporary states which can become more permanent, and which can be detected in voice quality, are the effects of intoxicating agents like alcohol and tobacco smoke. In excess, these agents tend to damage the vocal cords. “Whisky voice”, “ginny voice”, and “rummy voice” are popular labels for the deep, harsh, whispery voices that tend to signal one result of excessive habitual consumption of alcohol (Luchsinger and Arnold, 1965) and “smoker’s larynx” is a fairly frequently used medical label for the pathological effect of excessive tobacco smoke on the vocal cords (Myerson, 1950; Devine, 1960).

Lastly, information about temporary states such as fatigue can sometimes be found in voice quality.

B. Psychological Information in Voice Quality. We seem to be prepared, as listeners, to draw quite far-reaching conclusions from voice quality about long-term psychological characteristics of a speaker, in assessments of personality. In
Western culture, we are ready to believe, for example, that a harsh voice is correlated with more aggressive, dominant, authoritative characteristics, and a breathy voice with more self-effacing, submissive, meek personalities. The belief that personality characteristics, both normal and psychopathological, are correlated with voice quality, has been tested experimentally by many writers, mainly in the medical and psychological fields (Allport and Cantril, 1934; Cohen, 1961; Diehl, White and Burle, 1959; Eisenberg and Zalowitz, 1938; Fay and Middleton, 1939b, 1940b; Froeschels, 1960; Goldfarb, Braunstein and Lorge, 1956; Mallory and Miller, 1958; Moore, 1934; Moses, 1954; Pear, 1957; Sapir, 1926-27; Starkweather, 1964; Taylor, 1934). Some controversy remains, but in general writers seem to agree that some such broad correlations do exist. Intuitively, one would agree with them, but one major obstacle in the way of reliable scientific statements has been the lack of any standard system of labelling the voice qualities concerned, and a related inability to attain more than a fairly crude quantification of the voice quality variables which act as the experimental stimuli.

If it is true that information about personality is conveyed by voice quality, then the information must be chiefly carried by aspects of the habitual muscular settings, rather than by the basically invariant anatomy and physiology of the speaker.

C. Social Information in Voice Quality. Social behaviour is largely learned behaviour. Because of this, clues in voice quality to social information must lie mainly in those features of voice quality which can be acquired by imitation. In this sense, a particular accent often has a special voice quality associated with it, and the voice quality can thereby act as a partial clue to any social characteristics that are typical of speakers of that accent. Thus voice quality may serve as an index to features of regional origin, social status, social values and attitudes, and profession or occupation, where these features characterise speakers of the particular accent in question. Nasalisation is a voice quality component very commonly associated with particular accents. It characterises most speakers of Received Pronunciation in England, and many accents of the United States and Australia. Velarisation is a regional marker in the speech of speakers from Liverpool and Birmingham, England, and some parts of New York.

Voice quality can also act as an index to membership of a group which is not necessarily an accent-group (Fay and Middleton, 1939a). For example, some British male stage actors used to seem consciously to strive to attain a voice quality like that of Sir Laurence Olivier; similarly, military drill sergeants typically have harsh voices, and these are not necessarily the result of habitual vocal abuse, but rather acquired by imitation, in the hope of projecting the characteristic persona of their profession.

STEREOTYPED JUDGEMENTS IN VOICE QUALITY

We all act, as listeners, as if we were experts in using information in voice quality to reach conclusions about biological, psychological and social characteristics of speakers. Long experience of inferring such characteristics from voice quality, presumably often successfully confirmed by information from other levels, invests our implicit ideas about the correlations between voice quality and indexical information with an imagined infallibility. It is worth questioning the validity of this judgement process. We make judgements, and we act on them, but is the information we infer accurate, or is there a possibility that it is quite false? Since the correlations
must be statistical in nature, and not always of a very high degree of statistical confidence, obviously listeners will sometimes be wrong in the conclusions they draw from particular voice qualities. There is a good deal of evidence that in such subjective judgements we operate with stereotypes (Cantril and Allport, 1935; Eisenberg and Zalowitz, 1938; Fay and Middleton, 1939b, 1940a, 1940b; Starkweather, 1964). Listeners, if they are from the same culture, tend to reach the same indexical conclusions from the same evidence, but the conclusions themselves may, on occasion, bear no reliable relation to the real characteristics of the speaker.

Of the three types of indexical information in voice quality, biological, psychological, and social, it is the biological information which probably tends to lead to the most accurate conclusions, especially as to sex and age. Biological conclusions are possibly more reliable because of the fact that they derive principally from the involuntary, largely invariant aspects of a speaker's anatomy and physiology. Psychological and social conclusions are much more likely to be erroneous, because of their culturally relative nature, and because they derive from a more variable strand of the speaker's voice quality, the habitual muscular settings of the larynx and vocal tract.

**Experimental Investigation of Voice Quality**

The descriptive model suggested in this article represents no more than an initial structuring of the area, and a good deal of work, both experimental and theoretical, will be necessary before the phonetic description of voice quality can approach adequacy. Happily, experimental phonetics is not lacking in appropriate techniques of investigation. Experimental research can follow two complementary lines of approach, in speech analysis and speech synthesis.

In the speech analysis approach to voice quality, data are needed on a number of different aspects. It would be valuable, for instance, to have anthropometric measurements of typical variations in anatomical dimensions, as well as acoustic and physiological information about long-term articulatory activity. Acoustic techniques currently available include a wide range of analytic devices, from spectrography for discovering the distribution in time and frequency of acoustic energy, to inverse filtering of the speech signal for recovery of the characteristics of the glottal waveform. Physiological techniques which might be utilised include cineradiography, stroboscopic cinelaminography (Hollien, 1964), electromanometry, and electromyography (Cooper, 1965).

Speech synthesis is a useful avenue of research, in that hypotheses about physiological activities and their acoustic and auditory correlates can be easily tested. It is particularly valuable in that voice qualities of narrowly defined specifications can be fairly precisely simulated (Wendahl, 1963; Laver, 1967), and psycho-phonetic perceptual tests and the training of judges in the phonetic labelling system can be correspondingly facilitated.

**The Relevance of Voice Quality for the Disciplines Concerned with Speech**

I have maintained that while it is the business of general phonetics to suggest a descriptive model of voice quality, the study of this area has relevance for a number of disciplines.

Speech pathology has a direct interest in voice quality, and various systems of descriptive terminology are used by workers in this field. These systems, moreover, are not necessarily lacking in implications of phonetic specificity, as in the use of
terms like 'hyperrhinolalia' and 'hyperrhinophobia' as labels for 'severely nasalised voice', and 'dysphonia plicae ventricularis', for 'ventricular voice'. The limitation of such systems is in their emphasis on deviations, for whatever etiological reasons, from socially acceptable norms of voice quality. This is not to disparage such systems; in many areas of speech pathology attention is necessarily focused on abnormality. However, such systems are inherently too partial for general applicability to voice qualities of all kinds. The advantage of the more general type of system outlined in this article is in its being able to function without prejudice as to culturally-assessed factors of 'normality' or 'abnormality'. Both sorts of systems share the characteristic of analysing voice quality on a physiological, componential basis. Either system would be valuable, for instance, in the research area commented on earlier, that of the use of the voice as an indexical diagnostic clue to various pathological states.

Psychology and psychiatry have shown a frequent interest in voice quality research, because of the importance of the voice as an index to affective conditions, and to personality. Psychological experimentation in this area might benefit considerably from the techniques of synthetic simulation of voice qualities mentioned earlier, in that the voice quality stimulus variable in experiments investigating the correlations between voice quality and personality, for example, might be brought under more delicate, reliable and quantified control in this way.

Finally, considerations of voice quality are crucial for some aspects of the specifically linguistic study of speech. The distinction between voice quality and 'phonetic quality' is one of the most fundamental distinctions in linguistic phonetics (Ladefoged, 1962), since phonetic quality is the basic datum of the subject, in its capacity as the vehicle for the meaningful distinctions of phonology. It is sometimes thought that phonetic quality and voice quality are independent aspects of the phonic continuum, and that therefore phonetic quality can be related directly, as a simple abstraction, to the 'real world' of 'phonic quality'. A case can nevertheless be made for considering that phonetic quality and voice quality are not completely independent, and that phonetic quality can in fact only be judged against the previously assessed background of the voice quality of the speaker producing the utterance. Phonetic quality in this view would thus be a more abstract concept than is perhaps often believed.

One of the difficulties facing phoneticians and linguists investigating the phonology of a particular language arises from the general fact that voice quality and phonetic quality can rely for their manifestation on many similar activities of the speech organs. Activities used on a quasi-permanent basis, in some of the habitual laryngeal and supralaryngeal settings, can potentially also be used, on a much shorter-term basis, in the contrastive articulations representing phonological units.

Thus many of the features discussed in this article in their rôle as contributors to voice quality, such as labialisation, palatalisation, velarisation, pharyngalisation, nasalisation, breathy voice, and creaky voice, have also been found in various languages as signals used to differentiate the phonetic quality of the sounds representing phonological units in those languages.

It is because of the potential linguistic utilization of such features that the phonetician or linguist conducting phonological research must take an early decision about the status of these features, when they occur in the speech of his informant.
VOICE QUALITY AND INDEXICAL INFORMATION

CONCLUSION

The study of speech attracts the research attention of a number of different disciplines, each with its own professional interests. For the majority of these disciplines speech is a partial interest, and the main focus of the discipline lies outside speech as such, as in the case of psychology and psychiatry. Speech therapy, on the other hand, takes speech as its principal data, but brings a specialised interest to bear on a restricted area within the wider field of speech as a whole. The one subject which takes as its professional domain the study of all aspects of speech is phonetics. As such, phonetics should be able to offer, to these other interested disciplines, comprehensive theoretical structures for the description of all aspects of speech. This article has outlined a general phonetic approach to the description of voice quality, as one particular aspect of speech, in the hope of facilitating interdisciplinary discussion about this aspect, and about the indexical information conveyed by its different factors.

ACKNOWLEDGEMENTS

I gratefully acknowledge the helpful advice of Professor D. Abercrombie, Mrs. S. Hutcheson, L. A. Iles, J. A. Kemp and W. Lawrence, who read preliminary drafts of this article. Final responsibility for the views expressed is entirely mine.

REFERENCES


FAY, P. J. and MIDDLETON, W. C. (1939). Judgement of occupation from the voice as transmitted over a public address system and over a radio. J. appl. Psychol. 23, 536-560.

FAY, P. J. and MIDDLETON, W. C. (1939). Judgement of Spranger personality types from the voice as transmitted over a public address system. Character and Personality, 9, 144-155.

FAY, P. J. and MIDDLETON, W. C. (1940a). Judgement of Kretschmerian body types from the voice as transmitted over a public address system. J. soc. Psychol. 13, 151-162.

FAY, P. J. and MIDDLETON, W. C. (1940b). Judgement of intelligence from the voice as transmitted over a public address system. Sociology, 3, 186-191.


TAYLOR, H. C. (1934). Social agreements on personality traits as judged from speech. J. Soc. Psychol. 5, 244-248.
Proceedings of the Sixth International Congress
of Phonetic Sciences
Prague 1967

(Separatum)

© Academia
Publishing House
of the Czechoslovak Academy of Sciences
1970
THE SYNTHESIS OF COMPONENTS IN VOICE QUALITY

JOHN LAVER*

It has been said that the distinction between phonetic quality and personal quality, or voice quality in my usage of the term, is one of the basic assumptions of phonetics. Although we take phonetic quality as the basic datum of our subject, we know less than we might about voice quality, and I have been trying to construct a simple descriptive model of this area.

The essence of the approach is that voice quality is analysable into components, in much the same way that a consonant is described by reference to the voicing, place and manner components in its production.

Briefly, voice quality is conceived as deriving, acoustically, from two main sources: firstly, the permanent and ephemeral organic foundation of the speaker's anatomy and physiology; and secondly, the long-term muscular adjustments, or "settings", once voluntarily acquired perhaps, now unconscious, of the larynx and supralaryngeal vocal tract.

This second section, the long-term muscular settings of the larynx and supralaryngeal vocal tract, is an area open to traditional phonetic techniques of imitation and kinesthetic introspection, and one can thereby form hypotheses about the physiological actions necessary to produce particular qualities heard from a speaker. These hypotheses can be tested by acoustic synthesis, and I have carried out some preliminary simulation of some voice qualities, using PAT, the Edinburgh University resonance analogue speech synthesiser designed by Walter Lawrence.

Different voice qualities were synthesised as backgrounds to a standard linguistic message, a longish sentence of synthetic speech from "The North Wind and the

* Department of Phonetics, University of Edinburgh.
Sun”, “Then the North Wind blew as hard as he could, but the more he blew, the more closely did the traveller fold his cloak around him, and at last, the North Wind gave up the attempt.”

Three different fundamental frequency ranges were used, which I shall call Voice, (50—250 c/s), Falsetto, (120—480 c/s), and Creak, (32—128 c/s).

Besides the normal unmodified larynx and vocal tract settings for PAT, some extra optional laryngeal and supralaryngeal modifications were superimposed on the three fundamental ranges. Laryngeal modifications were: harshness and whisperiness. Supralaryngeal modifications included nasalisation; a latitudinal distortion of the vocal tract,—velarisation; and a longitudinal distortion of the vocal tract,—raised larynx.

Thus voices with up to five different components can be simulated: optional harshness, optional whisperyness, and optional nasalisation; one of the three configurations of the vocal tract,—unmodified (i.e. normal), velarised, or raised larynx; and one of the three fundamental frequency ranges, voice, falsetto or creak.

All possible combinations of these components give a total of seventy two different voice qualities, such as: Nasalised Falsetto, Harsh Whispeery Creak, Harsh Nasalised Velarised Voice, Whispeery Nasalised Raised Larynx Falsetto, and so on. I'd like to play you some of these, reserving acoustic details of the syntheses for later, should anyone be interested. I would especially appreciate any comments on the naturalness or otherwise of the qualities,—judging naturalness by a readiness, for example, to ascribe characteristics of personality to the voice in question.

**List of items on Demonstration Tape**

1. Normal Voice
2. Normal Falsetto
3. Normal Creak
4. Normal Whispeery Voice
5. Normal Harsh Voice
6. Normal Harsh Whispeery Voice
7. Normal Nasalised Voice
8. Velarised Voice
9. Raised Larynx Voice
10. Whispeery Velarised Creak
11. Harsh Whispeery Nasalised Velarised Falsetto

In conclusion, I'd like to say that a componential description seems to be a useful approach to structuring this area. In the synthetic work so far, only categorical

---

differences between voices have been attempted; the synthesis of scalar, quantitative differences within a component is the eventual aim of the project.

DISCUSSION

Fant:

1. Have you made an acoustic analysis of various pathological voices?
2. Your nasal quality might have benefitted from a simultaneous shift up in F1.

Mazzarella:

The artificial voices used in the demonstration often seemed to suggest a certain age of the speaker.

Laver:

ad Fant: I thank Dr. Fant for his suggestion about raising F1 to give a better nasal quality.

I have been less concerned with pathological voices than with the voice quality variations which are susceptible of imitation by normally equipped speakers — i.e. voluntary long-term muscular settings of the larynx and vocal tract. I have personally made no acoustic analyses yet, relying more on the somewhat restricted literature, and on physiologically-based hypotheses.

Clearly the next step is to proceed to the acoustic analysis of individual voices.
PŘEMYSL JANOTA

Personal Characteristic of Speech

The study is outcome of six years of research dealing mainly with modern phonetic methods. It discusses the personal characteristic of the auditory aspect of oral communication. The criterion for these personal characteristics is the ability of the listener to identify the speaker and to distinguish him from other speakers by the auditory characteristics of his speech, the colour of this voice, the so-called individual timbre.

The work is based on a complex of listening experiments, an analysis of the auditory aspect of spoken texts, and, last but not least, on synthesis of vowels and their auditory evaluation in confrontation with natural vowels. Work making use of synthetic vowels is described at greater length in the study and the author outlines the possibilities of using this method in solving general problems of phonetics of more general character.

BLANKA BOROVIČKOVÁ, VLASTISLAV MALÁČ

The Spectral Analysis of Czech Sound Combinations

The authors present the results of a research in speech acoustics achieved by them in recent years. Their investigation aimed at automatic speech identification with the help of computer. The study offers a detailed description of the acoustic structure of Czech sounds determined for the first time with the aid of modern experimental instruments.

The results achieved by the authors are applicable not only to linguistics and electroacoustics but also to psychology and phoniatry.