# Introduction

In his *Institutes of Oratory*, Quintilian wrote that 'The voice of a person is as easily distinguished by the ear as the face by the eye' (c.III, Book XI). The importance of an individual speaker's voice in everyday social interaction, as an audible index of his identity, personality and mood, could hardly be overstated. Yet we know now only a little more about the factors that give rise to different qualities of the voice than Quintilian did. Abercrombie was recently still able to write, with justification, that 'voice quality is the least investigated' of the different strands in the production of speech (Abercrombie 1967: 91). This book is an attempt to apply the principles of phonetic analysis to the description of voice quality.

Voice quality is conceived here in a broad sense, as the characteristic auditory colouring of an individual speaker's voice, and not in the more narrow sense of the quality deriving solely from laryngeal activity. Both laryngeal and supralaryngeal features will be seen as contributing to voice quality. Perceptually, voice quality in this broad interpretation is a cumulative abstraction over a period of time of a speaker-characterizing quality, which is gathered from the momentary and spasmodic fluctuations of short-term articulations used by the speaker for linguistic and paralinguistic communication. Following Abercrombie, the term 'voice quality' will be taken to refer 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 sound that issues from his mouth' (Abercrombie 1967: 91).

It has sometimes been maintained that the analysis of voice quality is quite external to the proper study of language (Sapir 1921: 47). One might then ask, given that phonetics shares with linguistics the responsibility for describing how spoken language works, why phonetics should take any professional interest in an apparently extralinguistic area such as voice quality. One obvious justification lies in the attitude that the semiotic function of linguistic communication can be better understood

1

## 2    *Introduction*

when seen in a wider semiotic context. Voice quality, as a major vehicle of information about physical, psychological and social characteristics of the speaker, has a vital semiotic role to play in spoken interaction (Laver 1968; Laver and Trudgill 1979).

There is, however, a further justification for a linguistic interest in voice quality. It springs from the essential nature of phonetic data, and involves a distinction between phonetic segments and phonetic 'settings'. Considering phonetic segments first, it is striking that despite the alternative approaches that are possible, most general phonetic theories that have been put forward have given major prominence to one particular way of dissecting the continuum of speech. They nearly all segment the continuum into short stretches approximately correspond-ing to the basic linguistic units of consonants and vowels. It is not difficult to understand why phonetic theory should organize its descriptive taxonomy in such a way as to facilitate the correlation of phonetic segments with phonological units. Phonetic theory is properly pre-occupied with linguistic uses of the vocal apparatus. In addition, the intellectual roots of modern phonetics are embedded in an orthographic culture whose alphabetic influence is very pervasive.

The time-domain of phonetic segments of this sort is typically short, and there is considerable variety in the articulatory activities involved. Because phonetic description finally has a linguistic motivation, it is the differences between the segments that tend to be emphasized, rather than the similarities. There is an alternative approach to the task of articulatory description, however, that concerns itself with both differences and similarities in vocal performance in speech. In such an approach, individual segments are seen as being articulatorily related to other segments in that a particular articulatory feature could be abstracted from the chain of segments as a shared property of all or most of the segments. A recurrent feature of this sort constitutes in effect a tendency for the vocal apparatus to be subjected to a particular long-term muscular adjustment (Abercrombie 1967:93) or 'articulatory setting' (Honikman 1964:73). One example of such a setting would be a quasi-permanent tendency to keep the lips in a rounded position throughout speech. Another would be a habitual tendency to keep the body of the tongue slightly retracted into the pharynx while speaking. Another would be the persistent choice of a characteristically 'whispery' mode of phonation. Settings give a background, auditory 'colouring' running through sequences of shorter-term segmental articulations.

It is important that the analytic relationship between settings and segments should be stated as clearly as possible from the very beginning. It is not proposed that settings and segments are complementary divisions of phonetic quality. The standard attitude that phonetic quality should be fully exhausted by a comprehensive segmental analysis is maintained. The analysis of phonetic quality into settings is a second-order analysis, abstracting data from a prior segmental analysis. It is true that it will often be analytically convenient to discuss the relation between settings and segments as if a given setting had a perturbing effect on the articulation of some particular segment, and therefore had some notionally independent existence. It would be extremely tedious to have to spell out the analytic priority of segmental analysis at every mention of the relationship between segments and settings. Let this discussion stand, then, as a general caveat. Having said that, it is also true that a phonetic theory which incorporates an account of settings as well as segments is demonstrably a richer theory, with a wider application, than one which focuses merely on the first-order description of segmental performance.

Settings can have different time-domains. Being by definition poly-segmental, a setting must be a property of a stretch greater than a single segment. But there is no upper bound to its extent in time. A setting can thus be used for phonological purposes, usually as a relatively short-term activity. The phonological use of settings has been described by many writers in the prosodic school of phonology, and also by Zellig Harris, in his article on simultaneous components in phonology (Harris 1944). As Lyons points out, Harris, like the prosodists, showed that 'in many languages the "simultaneous components" recognized in the analysis extended over more than one segmental phoneme and could frequently be associated with the whole of a higher-level structure: e.g. tongue-retraction in Moroccan Arabic, nasality in Swahili, etc.' (Lyons 1962).

Settings can also be used for paralinguistic purposes, usually on a somewhat longer-term basis, in signalling affective information through tone of voice, and regulating the progress of conversational interactions. On a quasi-permanent basis, it can also be used for extralinguistic purposes, as a phonetic component of voice quality identifying the individual speaker.

A useful example is offered by the different signalling functions of nasality. Nasality can be exploited phonologically both as a segment and as a setting. As a segmental quality in vowels, for example, it is used for

4    *Introduction*

contrastive lexical identification in many languages, including French, Portuguese and Yoruba. As a setting, nasality is used phonologically in Sundanese, a language of Java, as a marker of verb forms. Once initiated by a nasal consonant in any position in the syllable, nasality in Sundanese runs forward through all syllable boundaries until checked by a word boundary or a supraglottal consonant (Robins 1953, 1957). Nasality as a setting is also manipulated for paralinguistic purposes. Crystal cites Key (1967) as reporting that in Cayuvava, a language of Bolivia, nasality is used stylistically with an honorific function: 'an individual of lower social or economic status addresses one of higher rank with a prominence of nasalization for all vowels of the utterance; and similarly with a woman being polite to her husband, or a man asking a favour' (Crystal 1970: 191). Finally, nasality is a very common setting component of voice quality, either idiosyncratically or as an indicator of membership of particular sociolinguistic groups. It characterizes most speakers of Received Pronunciation in England, and many accents of the United States and Australia.

One fundamental principle emerges from these illustrations: if we leave aside matters of time-domain as such, then it is impossible to establish from considerations of auditory quality alone what the communicative function of some given vocal setting might be, on any *a priori* basis. Classification of a particular setting as having a linguistic function, or a paralinguistic function, or merely an extralinguistic function, has to await later knowledge, not available to the analyst solely on a phenomenal basis. If the phonetic settings which can be exploited in voice quality all have potential use for phonological and paralinguistic purposes also, then a descriptive phonetic model of settings in voice quality is directly available for application to each of these two other areas. In fact, the settings which will be discussed have almost all been discovered to be used in the phonological repertoire of different languages, or in conventional paralinguistic communication of different cultures. Although discussion will be directed chiefly to the occurrence of the settings as components of voice quality, comment will therefore also frequently be made about their use in individual languages, or in paralinguistic communication.

The study of voice quality has a further relevance to linguistics, springing from perceptual considerations. In order to decide which aspects of vocal performance count as linguistically pertinent in a particular language, the linguistic phonetician must be able to isolate

these aspects as figures against the perceptual ground of the other aspects of vocal performance. The perceptual relationship between figure and ground is reciprocal: the figure gains its clarity by virtue of the relative definition of the ground against which it is set. The question of what counts as linguistic data is thus the obverse of the question of what counts as voice quality data. Neither question can be answered independently of the other. Paralinguistic data completes the triad of complementary relations with linguistic and voice quality data. A phonetic theory which proposes an adequate account of the data of spoken language therefore cannot avoid taking voice quality into account.

The above argument stands even on a narrow definition of the concept of 'language', as might be used by those linguists whose interest in phonetic data is limited to the phonologically distinctive aspects of spoken language. A broader view of language will be adopted in this book, however, in which the phonetic material of spoken language and the phonetic component of voice quality overlap to some degree. In this broader approach, the view that is taken of the linguistic accountability of phonetic theory is that phonetic theory should be responsible for describing all recurrent, patterned, phonetic activity that characterizes the spoken language of the speech community concerned. In a famous passage in his article on 'Modes of meaning', Firth wrote that

Phonology states the phonematic and prosodic processes within the word and sentence, regarding them as a mode of meaning. The phonetician links ... this with the processes and features of utterance. Such processes are characteristic of persons, of social groups, even of nations. Moreover, the general feature of voice quality is part of the phonetic mode of meaning of an English boy, a Frenchman, or a lady from New York. Surely it is part of the meaning of an American to sound like one. (Firth 1951)

If the speech patterns of a community such as Liverpool, say, are characterized by persistent velarization and denasality, then this is a most important phonetic index of sociolinguistic and regional group membership, and should not be omitted from a phonetic account of the phonology of English as spoken in Liverpool. A formal linguist might of course object that these settings play no part in signalling distinctive linguistic contrasts in Liverpool English, and are therefore of no interest to his version of linguistics. To oppose that view would be merely to set one value judgement against another about preferable motives for constructing a linguistic theory. It is perhaps sufficient simply to say that in the perspective of this book language is seen partly as a social

instrument, and that a rich analysis of all the material of spoken language, including distinctive and non-distinctive aspects, allows a closer approach to the social texture of language.

A particularly good example of this approach, showing how an interest in settings can enrich a phonological analysis in a sociologically relevant way, is offered by Trudgill (1974). He showed, in an investigation of the speech of Norwich, that the articulatory patterns of working-class speakers, compared with those of middle-class speakers, are marked by the habitual use of a number of settings. These include: creaky phonation, a high pitch range, a loud loudness range, a fronted and lowered tongue position, a raised larynx position, a particular type of nasality and a relatively high degree of muscular tension throughout the vocal apparatus (Trudgill 1974: 186–7; Laver and Trudgill 1979). The relevance to phonological description here is that Trudgill found that by incorporating a setting rule very early in his sequence of phonological rules, he could use it to generate multiple surface phonetic data which in a more conventional segment-focused approach could not easily have been seen as related. Perhaps the chief benefit of this approach is its success in relating linguistic to sociological factors. As Trudgill comments:

> 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 the speaker. It is in any case this feature which most clearly distinguishes WC from MC speakers. This point, of course, did not emerge at all from our atomistic analysis of the co-variation of linguistic and sociological phenomena. (Trudgill 1974: 190–1)

The advantage of setting rules in the phonology of Trudgill's comparative sociolinguistic description is thus 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 whole series of individual rules' (Trudgill 1974: 190).

In a similar vein, Esling (1978) has shown that in Edinburgh social class correlates with laryngeal settings, in that higher social status corresponds to a greater incidence of creaky phonation, and lower status to a greater incidence of whisperiness and harshness.

As a general point, then, to the extent that settings characterize social and regional groupings of speakers, a descriptive model of such settings should be of value to comparative sociolinguistic investigation of those

groups. By extension, the analysis of code-switching within the speech of a single socio-linguistic group should also be facilitated.

A basic attitude in this book is thus that the study of phonetic components of voice quality is not irrelevant to the study of spoken language. But of course the study of voice quality also has a great deal to offer to the many other disciplines that take a professional interest in speech. For all these disciplines, speech is a partial interest. For phonetics, the entirety of its subject is defined by speech, in all its aspects. Before beginning to discuss the descriptive model in more detail, it may therefore be appropriate to comment briefly on the characteristics that it should possess in order to qualify as a *phonetic* model.

Firstly, the model should rest on a scientific base, and not rely on impressionistic description which is idiosyncratic to the individual analyst. The advantages of standardization are self-evident. Secondly, the descriptive analysis of a given voice must be communicable in writing, without requiring an audible demonstration of the qualities to which reference is being made each time particular labels are used (Laver 1974). Part of this requirement is a useable transcription system. Thirdly, the analysis the model provides must be replicable. Judges must be able to be trained in the use of the system, and be able to make judgements which are consistent both with those of other judges and with their own judgements of the same material on different occasions. Lastly, as an optimal requirement, the model should be integrative, drawing on work from the same areas as segmental phonetic analysis – namely, auditory and articulatory analysis, acoustics and physiology.

On these criteria, the descriptive system for voice quality offered in this book is a general phonetic model, applicable to the vocal performance of all human beings of normal anatomy and physiology.

The descriptive system that is proposed stands on an auditory foundation. But the auditorily-identified components all have correlates specified at each of the three other levels of analysis, all capable of instrumental verification – the articulatory, physiological and acoustic levels. These specifications are offered on my own responsibility, but I have tried wherever possible to draw upon the authority of specialists in the relevant disciplines, particularly in physiology and acoustics. Attribution to individual sources is made throughout.

A comment on level is necessary at this point: although the book was written as a monograph, it is hoped that it might also be of interest to students. With their needs partly in mind, a deliberately schematic

8    *Introduction*

account of the muscular anatomy and physiology is given, and physiological detail is kept to the minimum necessary to understand the hypothesized working of the muscular apparatus as a mechanical system. No claim is made to be advancing the state of knowledge in physiology as such, and the criterion for judging the degree of detail in the physiological presentation was one of phonetic relevance, at a level capable of being assimilated by senior undergraduate and beginning postgraduate students. The same is true of the acoustic presentation.

The prescription of acoustic correlates of supralaryngeal and laryngeal settings that is offered is based on two types of sources. The first is the available literature on acoustic phonetics. The second is instrumental investigations of my own performance of the settings. The analysis of the acoustic correlates of supralaryngeal settings was carried out by Francis Nolan, of the University of Cambridge Department of Linguistics, on recordings made by myself of the first paragraph of the Rainbow Passage (Fairbanks 1960) in the Phonetics Laboratory of the University of Edinburgh. The analytic method involved a computer program using a linear prediction pole-finding routine, developed in the Engineering Department of the University of Cambridge by Stephen Terepin, supplemented by visual examination of spectrograms. Some of the computer-based results are presented in Chapter 2, in Table 1, and some of the spectrograms in Figure 19.

The analysis of acoustic correlates of laryngeal settings was carried out in the Speech Science Laboratory of the University of California at Santa Barbara, by Robert Hanson, using the Sondhi-tube method of analysing glottal volume–velocity waveforms (Sondhi 1975; Monsen and En-gebretson 1977), with myself as subject. Data on the spectral slope of glottal waveforms from this experiment is included in Chapter 3.

Illustrative recordings of glottal adjustments in laryngeal settings were made by Peter Roach, with both himself and myself as subjects, using a Fourcin laryngograph in conjunction with a PDP-8 computer, in the Phonetics Laboratory of the University of Reading. Some of these are included in Chapter 3.

The account of each type of setting will follow a standard pattern. First, a schematic account is given of the muscular physiology of the setting; then the acoustic characteristics are discussed; and finally, phonological and paralinguistic uses of the setting in question are briefly mentioned.

The point of departure for the descriptive system was Abercrombie's account of voice quality (Abercrombie 1967: 89–95). More sporadic

comments on voice quality are to be found scattered through a wide range
of phonetic literature. I have tried to incorporate these insights wherever
relevant, because I believe that a subject should be thoroughly conscious
of its historical roots. There was not enough space, however, for more
detailed comment on historical aspects, and a further commentary on this
topic can be found in Laver (1975, 1977, 1978). A classified bibliography
of research into voice quality is available in Laver (1979).

The analytic approach used in this book is one of componential analysis
rather than holistic identification. Each given voice is analysed into the
independently controllable settings whose composite auditory effect
characterizes the overall voice. In this way, with some forty basic settings
available in the descriptive repertoire, a very large number of different
composite voice qualities can be described. Attaching a scalar label to the
effect of a given setting makes the system yet more delicate: these scalar
labels are discussed in Chapter 5. In a holistic approach one would have to
have a single label identifying each of this very large number of different
voices, giving an impossibly cumbersome system. Holistic labels of an
impressionistic sort, such as 'heavy voice', 'light voice', 'tinny voice', or
'thick voice', are therefore avoided.

The learnability of the system depends on the possibility of the user
being able to discriminate and identify the different auditory elements.
Demonstrations of the settings can be heard on the tape-recording
accompanying this book, and a suggested transcription system, based on
International Phonetic Association conventions, is given in Chapter 5.
The descriptive system was initially taught to a group of postgraduate
and postdoctoral researchers in 1976 (see Esling 1978). The preparation
of a full set of self-instructional materials is the subject of a current
project funded by the Medical Research Council in the Phonetics
Laboratory at the University of Edinburgh. These are primarily intended
to be used by medical personnel, but should be capable of wider
usefulness.

The possibility of learning to use the descriptive system is enhanced by
the fact that the vocal qualities concerned are all capable of being imitated
by all anatomically and physiologically normal speakers. This is because
voice quality derives from two distinct factors in vocal performance. The
first of these is to do with the nature of the individual speaker's own vocal
apparatus. The particular anatomy of the speaker constrains his voice
quality by the effect of such physical features as the dimensions, mass and
geometry of his vocal organs. Thus, *organic* features such as the length of

10    *Introduction*

his vocal tract, the size of his tongue, velum, pharynx and jaw, the shape of his laryngeal structures and the volume of his nasal cavity, will all contribute their effect to the overall quality of the speaker's voice.

The second factor is to do not with the nature of the vocal apparatus at a speaker's disposal, but the use to which he puts it. Each speaker, as part of his habitual style of speaking, tends to use particular settings of his vocal apparatus. The descriptive system offered here largely excludes consideration of the first, organic type of influence on voice quality, except as a ground against which the figures of individual settings are perceptible. The concern of the descriptive system is primarily the second, *phonetic* type of feature.

Since these phonetic setting features are all by definition a matter of a mode of control of the muscular apparatus for speech, then all normal speakers should be able to learn to imitate the articulatory basis of the settings, and to recognize their auditory correlates. The descriptive model therefore refers to settings of an idealized vocal apparatus, and ignores inter-speaker differences of anatomy. The generality that this assumption permits is one of the essential attributes of a *general* phonetic theory.

Phonetics is inherently a synthesizing subject, drawing its methodology and its descriptive concepts from a wide range of different disciplines. It would be pleasant to think that this book might hold some appeal for a correspondingly wide range of readers. Linguists, phoneticians and speech scientists are amongst the primary groups of readers envisaged, as are speech therapists and speech pathologists. It may be worth emphasizing that the book concerns itself only with the 'normal' voice; but it is assumed that an indication of the enormous variety of potential qualities that the normal voice can achieve will be of interest to those concerned with speech and voice disorders.

Given the wealth of speaker-characterizing information conveyed by the voice, it is hoped that the descriptive model outlined here will also be useful to those researchers interested in the complexities of nonverbal interaction in face-to-face communication, including psychologists, psychiatrists and ethologists. In addition, since voice quality can signal membership of many types of social and regional groupings, sociologists, social anthropologists, social psychologists and sociolinguists may find the descriptive system of interest.

Spoken communication between men and computers is almost certain to be a widespread social reality in the not-too-distant future. There are