1 *Voice and Voice Quality*

This chapter presents voice quality as a long-term component of accent. We describe in detail the Laryngeal Articulator Model (LAM), defining key terminology and providing a unified account of laryngeal anatomical structure and physiological function. We present observations of lower-vocal-tract phonetic phenomena that have often been portrayed in inconsistent, confusing, and incorrect ways in the literature on speech physiology and in describing the sounds of the world's languages. Prior to the methodology described in this text, the structure and functions of the laryngeal articulator, and hence of voice quality, were poorly understood, partly because of the difficulty in visualizing the articulator producing its full range of sounds as they occur in the world's languages. Categories are presented in order from the lower to the upper vocal tract, paying specific attention to laryngeal categories.

1.1 Voice Quality Defined

Voice quality, in its broadest theoretical sense as a phonetic descriptor of accent, refers to the long-term characteristics of a person's voice – the more or less permanent, habitually recurring, proportionately frequent characteristics of a person's speech patterns (Abercrombie 1967, Laver 1980). A parallel term denoting how we recognize a person's voice is 'long-term quality' (Nolan 1983). As a property of accent, voice quality refers to all of the habitual, long-term background, or holistic characteristics perceived as the most constant or persistent over time in a person's speech. Perceptually, it is the longest-term phonetic strand of the aural medium for language. The other two strands, the voice dynamics (prosodic) strand and the segmental strand, are progressively shorter term. Segments (vowels and consonants) last for tens or hundreds of milliseconds, voice dynamics (e.g. intonation) components can occur over a stretch of syllables or words, while voice quality, as Abercrombie (1967: 91) put it, '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 sound' of a person's speech. Elements of all three strands can be shared characteristics acquired as a part of the person's regional or social upbringing. Individual anatomy has a part to play, and some characteristics are learned and adopted idiosyncratically by the individual, but most characteristics of all three strands of accent are acquired remarkably similarly by members of the same linguistic community in a given generation. Still, the 'accent' of a place should not be assumed to be absolutely uniform across all speakers. In assessing the voice quality common to a regional or social group, the identification of those voice quality traits that are preponderant within selected sociolinguistic samples of speakers determines the extent of their distribution.

The term 'voice quality' has been used in another sense. In what could be called colloquial phonetic usage, 'voice quality' is applied narrowly to the kind of voicing produced at the glottis, specifically by the vocal folds. In voice quality theory, this narrower usage is more precisely termed the description of 'phonation type' or 'phonatory quality'. In somewhat intermediate phonetic usage, the term has also been used to designate the short-term effects, or 'register' effects, that originate within the larynx, in the lowermost part of the vocal tract in general. They are generally syllabic in duration and linguistically contrastive in the particular sound system or phonology in which they occur. We shall often use the cover term 'vocal quality' in referring to these shorter-term auditory effects produced by vocal-tract configurations, especially those produced within the laryngeal articulator.

Whether in its broader or narrower sense, the sound generated at the glottis, primarily by the vocal folds, is considered the source or the first element that shapes a person's speech. When we recognize a person by the sound of their voice, voice quality in the broader sense is generally what we mean. We are responding to speech cues that are the most persistent, habitual or long-lasting in the accent of the person whose voice we are hearing. These speech cues can be perceived in short portions of speech precisely because they are properties incorporated to varying degrees into the rapidly changing short-term vowels and consonants that make up the rapidly fluctuating segmental level of speech. In the broader sense of long-term voice quality, the vibrations originating within the larynx constitute the phonation type that is the background heard throughout the voiced sounds that a speaker produces. This provides a holistic unity to the auditory character of the voice despite the variation found in the segmental and voice dynamic strands. Components of the speech signal generated in the laryngeal articulator are particularly amenable to long-term functionality, since vibrations are inherently sustained relative to more momentary articulatory gestures and because articulations in the lower vocal tract are generally slower than those in the oral vocal tract. The laryngeal source signal is then modified through the shapes created by the positions of the articulators in the rest of the laryngeal articulator and through the upper vocal tract, producing characteristic resonances. These spectral characteristics are more telling when the source is voiced (due to periodic vibrations) rather than voiceless (when the glottis is open); but laryngeally voiceless sounds can also have spectral frequencies associated with them as a result of turbulence (friction/noise) generated by articulatory postures in the laryngeal vocal tract before the speech signal is further modified by actions of the oral articulators. To the extent that the laryngeal source signal and consequent vocal tract modulations persist throughout the sounds that an individual speaker produces, they can be identified auditorily as quasi-permanent (Abercrombie 1967) or long-term (Laver 1980, Nolan 1983) elements of voice quality. Because these characteristics are present more or less all the time that a person is speaking, they come to typify the sound of the person's voice.

Change in language is endemic, and the different phonetic components of speech interact in the process of sound change. This may apply especially to voice quality as a vehicle for sound change, because the longest-term strand of accent carries fewer attributes of contrastive linguistic meaning than voice dynamics patterns or segmental articulations. We do not yet know exactly how the other strands depend on voice quality characteristics to carry their contrastive meaning. To know this, we should have to discover how changes to voice quality settings affect each particular segment or intonational pattern in a phonology. Of course, it remains possible that changes in voice quality setting will have no appreciable effect on the nature of segments or prosodic patterns as they express linguistic meaning. Still, we hypothesize that subtle changes in parallel strands of accent are a potential agent of sound change; that is, altered longer-term qualities on a given segment, segmental string or prosodic pattern can shift the meaning-carrying attributes of that sound sequence. For example, combinations of secondary qualities influencing primary segmental (or syllabic) articulations can facilitate the development of a new primary category. Secondary pharyngealization on uvular segments has been described as a potential vehicle for uvulars to diffuse lexically to pharyngealized uvulars and to become primary pharyngeal segments, as they are reported to have done in Wakashan languages (Carlson and Esling 2003). Such processes are inherently auditory – a function of acquisition in context. With the requisite experience, it is common for listeners with regional linguistic familiarity to be able to identify what particular region a speaker comes from or social group a speaker belongs to by their accent. Our heuristic process is equally auditory – developed through the association of speech articulation with the auditory identification of resulting sound patterns. Our aim in this volume is to present the background for those associations, to characterize the auditory quality of these sounds, and to offer illustrations of cases (individuals, linguistic communities, and linguistic contrasts) where those auditory qualities occur.

The comprehensive description of voice quality presented by Laver (1980) reviews the concept of voice quality in phonetic theory, presents the categories and labels for analysing voice quality settings, and summarizes the articulatory, acoustic, and auditory research relevant in defining each category. Laver also makes a pertinent observation about voice quality as auditory background – the relatively stable 'ground' against which the 'figures' of rapidly fluctuating consonants and vowels move in a 'figure-to-ground' relationship (1980: 5). He describes a triad of levels of meaning: linguistic (the figures), paralinguistic data, and voice quality (the extralinguistic background). His approach reflects Abercrombie's (1967) partitioning of the aural medium, in which the timing of auditory/physiological processes is emphasized. Voice quality, as the shifting 'overall timbre' of the voice, was also explored by Pike (1967: 525–527); but he considered it a 'subsegmental' phonemic unit for signalling contrastive meanings, in the sense of attitudinal expression, rather than a suprasegmental layer. What the theories share is the view of articulatory postures and movements over varying time periods creating auditorily recognizable settings of the vocal tract. Therefore, it is of critical importance to define a model that relates all the elements of the vocal tract in a coherent way. A linear 'source-filter' vocal tract model, where the larynx is only a sound-source modulator and the tongue, jaw and lips the principal consonantal and vocalic articulators, has its limitations. It has no way of elegantly linking glottal shape and oral articulation with the postural configurations (of whatever duration) that occur between them – all of which are interdependent. It is in effect missing a piece – the complex of sources and resonators that is the larynx.

1.2 The Laryngeal Articulator

In the model we have developed, the 'Laryngeal Articulator Model' (LAM), the lower part of the vocal tract contains a complex articulator equivalent to the tongue in the oral vocal tract. This replaces the ideas that the pharynx is an empty space controlled only by the muscles around its walls, that the tongue is the articulator that produces pharyngeal sounds, and that the epiglottis is an active articulator or the closing mechanism of the airway. All 'pharyngeal' and 'epiglottal' sounds, pharyngealization effects, epiglottalization effects, laryngealization, glottalization, and an array of lower-vocal-tract volume effects are produced by the action of the aryepiglottic constrictor mechanism together with associated tongue retraction and larynx raising. The locus of stricture is the upper border of the epilaryngeal tube. The aryepiglottic folds are the active articulator, and the epiglottis is the passive articulator. The model elaborates the larynx as an articulator rather than just a source of vocal fold vibration (Esling 2005, 2010). It also redefines the relationship between vibratory structures in the laryngeal space and identifies a set of resonating cavities that shape spectral energy in ways that are predictable based on the inherent folding properties of the articulator. The tongue is given a new relationship to the laryngeal articulator, as the tongue is no longer the primary articulator of pharyngeal/epiglottal sounds.

Our research has focused on observing what occurs within the lower vocal tract, in what we generally refer to as the 'laryngeal constrictor'. This mechanism as a whole is responsible for constriction at the glottis (which cannot be said to constrict on its own) and for increasing degrees of constriction that affect the shape of resonating cavities in the lower vocal tract and create new points of vibration. We have examined the principles that link the laryngeal region (which the airstream passes through first) with the oral region of the vocal tract and how this laryngeal region is used in various languages to signify meaning. Figure 1.1 presents a graphic two-dimensional representation of the two-part Laryngeal Articulator Model. The laryngeal articulator itself is depicted within the laryngeal vocal tract – the space beneath the dotted line, from the glottis (vocal folds), through the epilaryngeal tube, in front of the aryepiglottic folds at the top of the tube, up behind the hyoid bone, through the upper pharynx to the back of the tongue and beneath the uvula. Retraction of the tongue is related to laryngeal articulation, as is larynx raising. Other movements of the tongue (raising and fronting) define states of the oral vocal tract, as do movements of the velopharyngeal, mandibular, and labial articulators. Full constriction of the laryngeal articulator (also called the laryngeal constrictor mechanism or the aryepiglottic sphincter) defines the physiological process of airway closure (Esling 1996). It is important to note that while the laryngeal articulator could be viewed as a 'pseudo' or 'functional' sphincter (and we often refer to it, and the aryepiglottic folds specifically, as such), it is not an anatomical sphincter (Fink 1974a, 1974b).

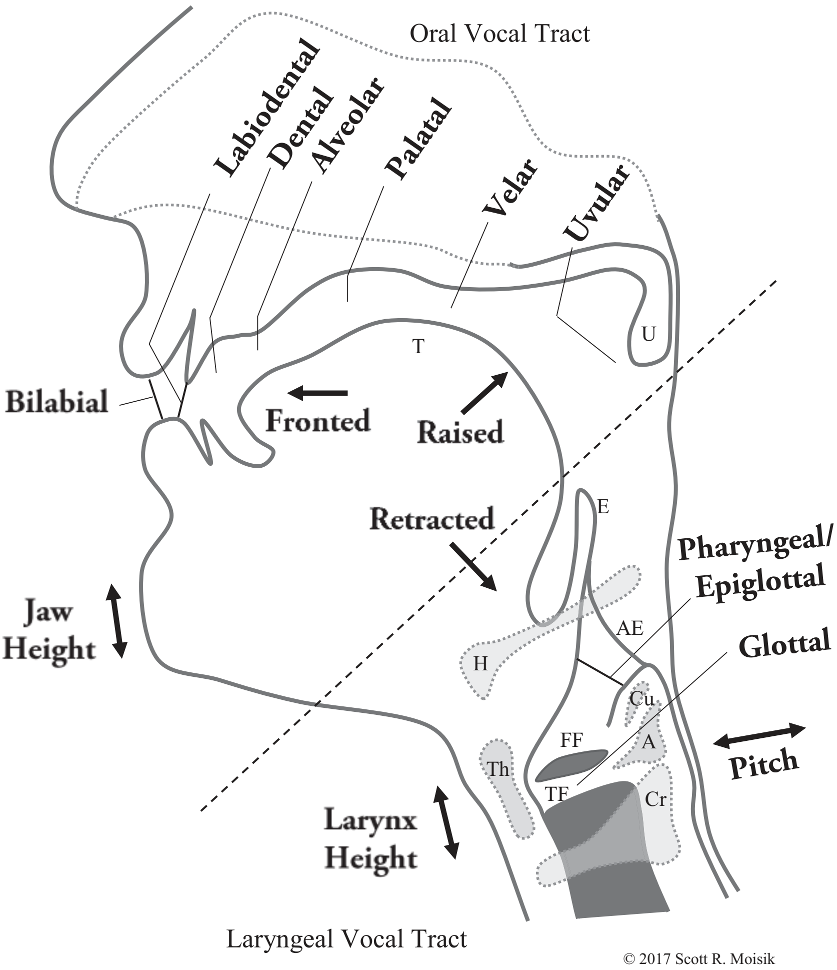

Figure 1.1 *The Laryngeal Articulator Model, or the two-part vocal tract*
T = tongue; U = uvula; E = epiglottis; H = hyoid bone; AE = aryepiglottic folds; Cu = cuneiform cartilage; A = arytenoid cartilage; Th = thyroid cartilage; FF = ventricular (false) folds; TF = vocal (true) folds; Cr = cricoid cartilage.

Figure 1.1a in the accompanying media files illustrates the three stages of the compression or ‘folding’ behaviour of the vocal tract model in Figure 1.1. The laryngeal mechanism is shown, from left to right, from open, to closing, to almost closed. With progressive closure, as the constrictor mechanism folds in tighter on itself, tongue retraction and larynx raising also increase.

The model serves as the background guiding the interpretation of the voice quality strand of accent. Our designation of laryngeal activity as articulation reconciles the narrow notion of voice quality as phonation or vibratory register with the broader description of voice quality as a set of configurational postures throughout the vocal tract. The multitude of vibrations the laryngeal articulator can generate is a function of its pattern of constriction, related to its articulatory shaping of the lowest resonating chambers in the vocal tract. These resonances are a function of the volume of the epilaryngeal tube, of the degree of tongue retraction into the pharynx induced by aryepiglottic fold tightening and consequent vertical compaction, of the height of the larynx (distance below the hyoid bone) during these events, and of changes in the dimensions of the piriform fossae. Normally, the larynx is raised during laryngeal constriction, compressing the epilaryngeal tube and shortening the vocal tract through the pharynx. With accompanying tongue retraction, the pharynx is also compressed vertically, and resulting pharyngeal volumes are smaller. These effects are produced deep in the vocal tract, earlier than other resonance or noise components generated in the upper (oral) vocal tract. As a result, the laryngeal constrictor mechanism exerts a profound effect on the production and perception of voice quality throughout the entire vocal tract. Understanding that the larynx is an articulator and not just the location of vocal fold vibration makes it easier to grasp the ambiguity of voice quality as laryngeal phonation in counterpoint to voice quality as the long-term setting of the articulators. The laryngeal articulator contains and combines both of these actions at the same time, and there is no clear-cut physical boundary between the posture of the mechanism for generating a particular phonation type and for generating a resonance with spectral (formant) characteristics that might otherwise be termed vowel quality and attributed solely to a dominant lingual posture in the oral vocal tract. In fact, the interlaced relationship between vowel quality, tonal quality, and phonatory quality as components of voice quality is a fundamental principle underlying the Laryngeal Articulator Model of speech production in phonetic theory.

Although voice quality in its narrower sense is taken to refer only to phonation produced by the vocal folds at the glottis, sources of periodic vibration are not restricted to the glottal level alone. There are three main sets of folds within the laryngeal mechanism that can generate periodic energy: the vocal folds, the ventricular folds, and the aryepiglottic folds (Edmondson and Esling 2006). The vocal folds are the most efficient source of voicing, but the ventricular folds can add to the aerodynamic/mechanical action of vocal fold vibration, as in harsh types of phonation (cf. Esling 2005) or in some Tibetan chanting or throat-singing styles (Fuks et al. 1998). The third level of stricture in the larynx – the aryepiglottic folds – represents the sphincteric pursing of the top of the epilaryngeal tube. It is the primary place of articulation at which pharyngeals are produced. Sounds labelled as epiglottals are also produced at this same place of articulation, since there is no other 'place' that forms stricture in the lower vocal tract. Distinctive periodic signals, different from vocal or ventricular vibration, are generated when a particular configuration of the laryngeal constrictor mechanism allows the aryepiglottic folds to trill (vibrate) against the epiglottis. All three levels of laryngeal folds generate periodic signals that can be taken as a 'source'. Besides the three main sets of folds within the larynx, other tissues can also be made to undulate when aerodynamic flow is sufficient, including tissues in the walls of the epilaryngeal tube and the pharynx, the tip of the epiglottis, and even mucosal accumulations that build up in narrowed passages of the airway.

As an articulator, the larynx adds a complex of possibilities to the formulation of how each of the three strands of accent is constructed. It also lends a new phonetic interpretation to how linguistic, paralinguistic, and extralinguistic meaning relate. Inherent in the model is the principle that pharyngeals are primary in the formation of voice quality, both in its narrow sense and in its broader sense. That is, vibratory patterns that can be associated with voicing (phonation) reflect the same vibratory possibilities associated with pharyngeal articulatory contrasts, for example, pharyngeal/epiglottal trilling. This has important consequences for voice quality theory. It means, firstly, that the articulatory mechanism for producing pharyngeal sounds is more complex than previously considered. The auditory distinctions and descriptive categories that have proliferated to account for the many distinctions in this region have been justifiable and insightful (IPA 1989) but were not well explained by the vocal tract model available at the time. Pharyngeal articulations are essentially laryngeal, since epilaryngeal-tube stricture and the aryepiglottic fold constriction that occurs at the top of the epilaryngeal tube are at the same time responsible for protective airway closure and for the generation of sounds labelled 'pharyngeal' or 'epiglottal'. This implies that there are not two separate, linearly organized places of articulation for the category pharyngeal/epiglottal, which is not independent of the larynx (see Esling 1999). Furthermore, instead of laryngeal sound production being restricted to glottal events (attributed to the vocal folds and shaping of the glottis), laryngeal sound production includes both glottal events and pharyngeal/epiglottal articulatory possibilities through the epilaryngeal tube and past the aryepiglottic folds at its superior border. These structures are depicted in Figure 1.2. Describing

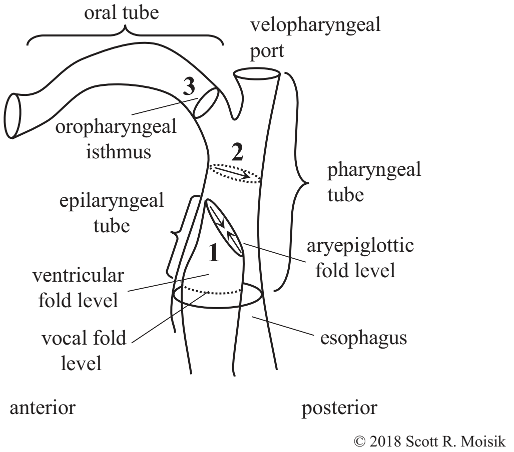

Figure 1.2 *The epilarynx as a tube-within-a-tube*
The tube-shaped epilarynx is found at the bottom of the pharynx tube; together these two regions define the lower vocal tract (LVT). The broad, mostly independent actions of these tubes are (1) epilaryngeal constriction and (2) pharyngeal constriction. The anterior boundary of the LVT is delimited by (3) velic sphinctering at the oropharyngeal isthmus.

pharyngeal articulations as a function of the laryngeal mechanism also has the consequence that pharyngeals precede other places of articulation in vocal tract geography and occupy a preeminent place in the chronological acquisition of speech sounds, so that the description of pharyngeal/epiglottal articulations figures at the core of voice quality description when infants are first learning to speak (addressed in Chapter 6).

1.3 Origins of Voice Quality Theory

The comprehensive history of voice quality presented by Laver (1975, 1979, 1980, 1991) chronicles the earliest origins of the concept of voice quality in phonetic theory. Laver (1975: 90) credits Wilkins (1668) with being the first in English to identify different linguistic communities as having generalized speech characteristics as a product of the different ways of rendering the basic set of speech sounds and the preponderance of certain sounds. Wilkins called individual speech sounds 'letters' and the generalized characteristics the 'pronunciation' – the differences in 'pronunciation' between languages depending 'upon the nature of those letters, of which they do chiefly consist and are framed' (1668: 381). In the nineteenth century, Henry Sweet at Oxford, whom many credit with being the inspiration for George Bernard Shaw's character Professor Henry Higgins in *Pygmalion* (cf. *My Fair Lady*) but which was more likely directly inspired by Daniel Jones (Collins and Mees 1998), explored 'the quality of the voice' in contrasting languages (1877). Sweet used the term 'organic basis' (1890: 69, Henderson 1971: 182–187) to refer to the long-term quality specific to a given language, although it is likely that others in the nineteenth century, such as Alexander Graham Bell (1908) shared the same concept. Paul Passy and Daniel Jones fostered this view of the 'indexical' nature of accent (see Laver 1968), inspiring Beatrice Honikman at University College London to explore 'articulatory settings' in the accents of various language groups (1964). Abercrombie's (1967: 89–95) definition of 'voice quality' is the most cogent placement of the concept within phonetic theory as having a general, relatively non-fluctuating quasi-permanent character, distinguishing it temporally from rapidly fluctuating segmental (vowel and consonant) articulations and middle-duration dynamic (prosodic) patterns.

A long-standing hypothesis in voice quality theory has been the premise that the long-term postural setting of the articulators predisposes and facilitates the performance of the more rapidly fluctuating segmental articulations. A corollary of this hypothesis is that adopting the long-term postural setting of a new target language makes it easier to perform the shorter-term segmental (and perhaps dynamic) articulations of that language. This was Honikman's goal in describing the articulatory settings most characteristic of the speech of contrasting languages (see also Esling and Wong 1983). As Sweet put it: 'Every 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' (1890: 69). These claims have proven difficult to demonstrate empirically, although they provide a theoretical argument that remains attractive. Chomsky and Halle drew attention to a preparatory state of the articulators before speaking, with the vocal processes adducted in preparation for voicing, which they called the 'neutral position'. As Chomsky and Halle observed, citing cineradiographic studies of speech, 'just prior to speaking the subject positions his vocal tract in a certain characteristic manner' (1968: 300). There was initial scepticism that a physiological posture acquired over time could be universal rather than a function of cultural development (Annan 1972). And research into voice quality has demonstrated that the 'neutral position' of the articulators is language specific, that is, it exhibits a configurational bias depending on the speech patterns in the language variety a speaker has learned (Gick et al. 2004). Kedrova and Borissoff (2013) have summarized the