# Pronunciation Considerations in ESL: Voice Quality Settings[1]

John H. Esling
University of Victoria

Voice quality settings are the various composite postures or long-term features of speech. They include the long-term position of the larynx, pharynx, jaw, tongue, velopharyngeal system and lips, as well as long-term laryngeal configurations reflected in the diverse phonation types described by Catford (1964). Voice quality may function linguistically, to characterize the particular language or dialect or social group to which a speaker belongs; or it may function paralinguistically to signal mood or emotion in conversational contexts; or it may function extralinguistically to characterize or identify the individual speaker. It is the voice quality settings that function linguistically that are of most interest to the ESL teacher, for they constitute a part of accent -- together with features of segmental phonology, rhythm, and intonation -- and serve as indicators of the speaker's regional or social affiliations.

An awareness of the voice quality settings -- the articulatory and phonatory postures -- that characterize and indexically mark a person's speech as a part of accent can be a valuable asset to all language teachers, including ESL teachers, and their students. Using this approach, students can practice

---

[1] The author would like to thank Rita Wong (San Francisco State University) and Will Vroman (University of California, Santa Cruz) for their contribution to a workshop at the Fifteenth Annual TESOL Convention, Detroit, March 3-8, 1981, in which the concepts of this article were demonstrated.

and develop their pronunciation and intelligibility, improve their comprehension of a greater variety of English speakers, and become more aware of the image that they project when speaking English.

One aim of this article is to draw attention to the investigation of aspects of setting in several languages by Beatrice Honikman (1964). Her objective is to inform teachers and students of the fundamental differences in pronunciation between ESL students' native languages and English, which can facilitate the acquisition of a satisfactory accent in English. More recently, John Laver (1975, 1980) has proposed an articulatory phonetic terminology for describing voice quality. This thorough investigation of an aspect of phonetic description which is often neglected in phonological analysis and sociolinguistic studies is based largely on the teaching of David Abercrombie in the British phonetic tradition. As a descriptive system, it can be applied effectively in the process of diagnosing, illustrating, and studying voice quality settings in ESL students' native languages, and practicing settings in English as the target language.

Laver's approach reflects an alternative to the tendency to emphasize the differences between segments in the chain of speech. Instead, much of our effort in describing speech should concentrate on the similarities running through all the segments in the phonetic substance of the language we are describing.

> 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 (Laver 1980:2).

Examples of these voice quality features are available on a cassette tape (Laver 1980). They are illustrated for a variety of languages below.

The ESL teacher should be aware of a number of voice quality settings in languages other than English. Honikman (1964) discusses setting features for French, Russian, Indian and Pakistani languages, German, Turkish and Persian. She describes a typical setting of French as rounded, with tongue blade and fronted articulation, with slightly open jaw setting. German is also characterized as lip-rounded. Russian, in contrast, is close in jaw setting, with spread lips and fronted (palatal) articulation. Indian and Pakistani languages are described as having open lips and jaw, with retroflex articulation of the tongue. Turkish and Persian are cited as examples of languages where articulation is performed primarily by the tongue tip. These descriptions are restricted to features which can be identified both auditorily and visually. Only labial, mandibular, and front lingual settings are evaluated.

Using the descriptive framework proposed by Laver (1975, 1980), it is possible to elaborate on these descriptions, identifying features associated with articulations which are not necessarily visible. All descriptions in this article represent phonetic analyses by the author. In addition to *open rounding*,

French accents often have a habitual backing of the tongue -- *uvularization*, or *pharyngalization* in some cases. *Nasal voice, lowered larynx, breathy* or *whispery voice*, and often a relatively high pitch range are common features of French. Since some of these features are also found in a variety of accents of English, it may be that they pose no problem for the French speaker learning English. In fact, features such as lowered larynx and breathy or whispery phonation may be a positive asset, ranking relatively high in sociolinguistic prestige in English, and giving what is recognized as a French accent higher status among foreign accents in English.

German accents, because of their historical proximity to English both since Anglo-Saxon times and in North American colonial development, may also share many of the setting features found in dialects of English. Therefore, lip rounding may be one of the few noticeable differences between German and English settings, although accents of Durham or Northumberland in the north of England also have this feature in common with German. Dialects of German also vary considerably, but characteristic features often include uvularization and combinations of degrees of *raised larynx* and *faucal constriction* in many northern accents, or lowered larynx and *expanded pharynx* in many southern accents. As in the case of French, these settings may carry varying degrees of prestige in an English-speaking community, presumably improving the image of the speaker in areas where the same features are found in familiar, socially prestigious varieties of English. Extreme open rounding, fronted, *palato-alveolarized* tongue position, and *whispery creaky voice* in Norwegian or Swedish are another example of this. Only extreme rounding is uncharacteristic of English taught in ESL classes, whereas a *palatalized* or palato-

alveolarized tongue setting and whispery creaky phonation are common, and even prestigious in some varieties of English, according to sociolinguistic studies of Norwich (Trudgill 1974) and Edinburgh (Esling 1978b).

Russian accents, in contrast, often combine the features mentioned above, *close jaw*, *spread lips*, and palatalized tongue position, with faucal constriction -- tightening of the upper pharynx. If this setting is unfamiliar to English speakers, it may prove an obstacle to intelligibility or to social interaction.

Any individual feature which figures prominently in the setting of a foreign language but which does not occur commonly in English is a potential obstacle to intelligibility or social favour. Examples of accents which illustrate characteristic features include: extreme *retroflexion* and *open jaw* in some accents of India; close jaw, nasal voice, uvularization, and tongue *blade articulation* in Chinese; extreme uvularization in Hebrew and some dialects of Arabic; lowered larynx, faucal constriction and uvularization, with labial spreading in Japanese; and *tip articulation*, nasal voice, and breathy voice in Persian.

The voice quality features used in these descriptions are adapted from Laver (1980:158-161,165) and listed below.

Supralaryngeal voice quality settings

| | |
|---|---|
| labial | open rounding |
| | close rounding |
| | spread lips |
| mandibular | open jaw position |
| | close jaw position |
| | protruded jaw |

| | |
|---|---|
| lingual tip/blade | tip articulation |
| | blade articulation |
| | retroflex articulation |
| lingual tongue body | dentalized |
| | alveolarized |
| | palato-alveolarized |
| | palatalized |
| | velarized |
| | uvularized |
| | pharyngalized |
| | laryngo-pharyngalized |
| faucal | faucal constriction |
| pharyngeal | pharyngeal constriction |
| velo-pharyngeal | nasal |
| | denasal |
| longitudinal | labial protrusion |
| | labiodentalized |
| | raised larynx |
| | lowered larynx |

Laryngeal voice quality settings

| | |
|---|---|
| simple phonation types | modal voice |
| | falsetto |
| | whisper |
| | creak |
| compound phonation types | whispery voice |
| | creaky voice |
| | breathy voice |
| | harsh voice |

Teaching pronunciation is an important part of any ESL teacher's

program. A teacher's awareness of differences in voice quality and an ability to present features of pronunciation in terms of the long-term configurations of the vocal tract musculature can be economical for learning and beneficial to the students' performance in spoken English. Students benefit (a) in their ability to communicate in English and (b) in the realization of the personality or image that they project when they converse in English. Teachers, for their part, are better able (c) to understand students' pronunciation difficulties and (d) to provide concrete tools helpful in improving pronunciation performance.

The most important benefit for the ESL student who acquires an awareness of voice quality settings in English is improved spoken communication. A number of segmental pronunciation difficulties may all result from the learner's inability to grasp the generalization that a particular setting -- or long-term configuration -- represents. The vowel and consonant phonemes of a language share features which can be taken together to constitute the habitual articulatory posture of that language. If the voice quality of the learner's native language differs from the setting normally found in the target language, interference will necessarily be long-term and the intelligibility of individual vowels and consonants may be reduced. It is not yet clear how great a role voice quality settings play in comprehension or intelligibility, but we can assume that if multiple accent characteristics are made to sound closer to a native pronunciation of the language (over the entire period that the individual is speaking in the case of settings), the speaker's intelligibility can be expected to improve.

The second major benefit of practice with voice quality settings

is to improve the image that students project when they speak English. Settings have been shown to differentiate individuals according to social background (Trudgill 1974:190; Esling 1978a, 1978b). ESL learners can find it helpful and of general interest to be presented with setting features that reflect the range of social status in an English-speaking community. Since the social background, and even notions of the intelligence or ability of the individual are communicated to some extent through voice quality, attention drawn to these long-term aspects of pronunciation can be revealing, and result in progress for many students.

Teachers will find the concept of settings helpful in recognizing what is causing many students' pronunciation problems. Several vowels and consonants, for example, could all be difficult because of one or two features of voice quality which are very unlike English. Attention to a single higher-level feature may account for a number of collective difficulties encountered with lower-level segmental features. For example, in learning French, it is helpful to keep the lips rounded and tense throughout speech. For a learner of French who has difficulty articulating the individual segments /y/, /ø/ and /œ/, it may be more efficient to concentrate on a habitually rounded setting than to ponder the individual segments as they occur. One might also discover in this process that the /i/ of much colloquial French is often more rounded than an /i/ of English as a result of the roundedness of the setting.

Finally, although students may be aware that their pronunciation differs from the target phonology presented to them, they may be unable to identify the reason for this. Here, a teacher's awareness of voice quality can help identify a formerly unrecognized

or unconscious element of a student's speech. Presenting components of common or preferred voice quality settings in English provides the student with the necessary model to practice (and to listen for) to improve performance in spoken English.

A baby's earliest vocalizations begin to illustrate some of the characteristics, particularly features of voice quality, that will distinguish that individual as an individual. This is evident from the ability of mothers to recognize their own infants' cries (Murry et al 1975, 1977). Such voice quality features are idiosyncractic or personal indices. As units of linguistic contrast are acquired -- a segmental system of vowels and consonants, and voice dynamics features including intonation and rhythm -- the baby is also learning the higher-level linguistic features of voice quality that set off the baby's family and neighbours' families from other dialect groups. These features function as regional and social indices, or indicators in Labov's terminology. The child will also learn contrastive paralinguistic uses of voice quality settings. It is not well documented whether these settings are acquired before, after, or at the same time as minimally distinctive phonological contrasts, but it seems reasonable to assume that by the time the rudiments of the local phonology are mastered by the child, the settings characteristic of the local community are also present. Esling (1978b) has shown that 8-year-olds possess most of the voice quality setting features of adults in the community.

It may be that there is a mutual relationship in the acquisition of setting features and segmental features in that the long-term setting provides the background posture required for the accurate rendering of the segments in the language. Linguistic contrast between phonemes reflects the sometimes small

number of features which distinguish one phoneme from another, while at a slightly higher level of analysis it is the sameness or continuity of the phonetic substance throughout the string of phonemes in speech which defines the particular setting. To a large extent, voice quality settings signal membership of the larger group -- the local region or city or country, or social class.

This leads us to two important points about second language learning and instruction. The first is that voice quality setting features, perhaps because they are associated largely with individual speaker recognition or paralinguistic emotional colouring, may be unrecognized in the extent to which they signal regional or social information. Such distinctions would be particularly difficult for a foreign learner of the language to recognize, lacking the opportunity or ability to observe the extent of distribution of the phenomenon. Segmental features may be more apparently different in a target language from those of the student's native language, while setting features may be harder to recognize as having linguistic significance in the target language. This may cause learning problems for the ESL student. Whereas the child acquires setting and segmental features as a mutually combined system, the foreign learner may impose the new phonemes of English on the old background posture of a non-English, and perhaps inappropriate, voice quality setting. As a result, the identity of segmental contrasts may be obscured or masked by the old posture, and the student's English may often be unintelligible.

The second important point is that the ESL student may not recognize the difference in acceptability between various settings in English, with their contrasting social or regional implications.

The student's own native setting may contain features which, without the speaker's knowing it, evoke an unfavourable response from English speakers. In cases such as these, it should be possible for the ESL instructor to introduce setting features of English which would be more familiar to a greater number of hearers. Practicing English setting features should also provide the background vocal posture necessary for more fluent articulation of the phonology of English.

It is difficult to speak of learning pronunciation as if only one setting were used by all native speakers of the language. All languages have regional and social dialects with distinct settings that function as indicators. Beyond this level of generalization, there is still a considerable amount of individual variation among native speakers. This diversity is characteristic of English, although a combination of setting features common to a wide variety of American English speakers can be identified and presented to ESL students in the United States in the same way that the distinctive segments -- the vowels and consonants of a representative variety of American English -- are presented for improving pronunciation.

First, it will be useful to review some of the different voice quality settings that characterize dialects of English. These include: (1) nasal voice, close jaw and creaky voice; where a quasi-permanent nasality, habitually close position of the jaw, and constant or intermittent creaky, or very low-pitched, phonation combine to produce a voice quality characteristic of the accents of many British speakers; (2) velarized tongue body position; which may be thought of as a characteristic of Liverpool speech, or other parts of Lancashire, where the tongue assumes a velarized position throughout the whole of a person's speech; (3) uvularized

tongue body position and lowered larynx; characteristic of many Yorkshire accents, where the tongue is held further back in the mouth than in much Lancashire speech, and the larynx is kept slightly lower; (4) protruded jaw and harsh voice; which characterize the mandibular setting and phonation type of vernacular Edinburgh dialect; (5) breathy voice and nasal voice; characteristic of Glasgow, where phonation is typically breathier than in other varieties of Scots, and the voice is nasalized throughout; (6) retroflex articulation and spread lips; found in accents of Northern Ireland, where the tongue tip is often retroflexed slightly and the lips are spread, that is, the space between the lips is expanded horizontally, and often vertically as well.

What kind of setting, therefore, are we going to present to ESL students as a model? It should first be remembered that we are not referring to a prescriptive production model, but rather an abstract model made up of a variety of potential components. The purpose of such a model or illustration is sensitization -- to stimulate the students' awareness of the variety of voice quality settings that exist among the languages represented in the classroom, that are found in the local community, and that students might learn to evaluate as they experience and observe more of English-speaking society.

One effective method of sensitizing ESL students to their own and each other's native voice qualities is to ask students to prepare a short phrase from everyday conversation, an announcement, or a tongue-twister to produce in their native language to the rest of the class. Even with only one or two representatives of each language, a linguistically heterogeneous class can yield noticeable differences. Particularly salient

voice quality features can usually be assumed, provisionally, to be linguistically motivated, and can be contrasted from language to language. Students quickly learn that voice quality is not only individual, but also a part of one's accent in a language. Rapid presentation also tends to make the students approximate the average setting of their language.

Another technique for building awareness of voice quality in pronunciation is for students to observe and make notes of the settings of various personalities that they see on television. Certain programs might reflect a variety of regional or social dialects in English, whereas national newscasts might present a model which students wish to imitate. Whether or not imitation is used as a technique, it should be pointed out to students that there are voice quality settings which one adopts in increasingly formal or prestigious varieties of English. The features of a socially higher valued setting in English may or may not correspond to the voice quality features that students bring from their native languages. If not, the difference may contribute, along with differences in rhythm, intonation, and segmental phonology, to low intelligibility or unfavourable social judgements against the speaker. It is important for these students to become aware of voice quality and how to observe and recognize different settings. They should also be presented with a model containing salient features which are likely to occur in the pronunciations of English which they are accustomed to hearing.

In the United States, for example, a broad model of voice quality setting might include the following features:

(1) spread lips
(2) open jaw
(3) palatalized tongue body position

(4) retroflex articulation
(5) nasal voice
(6) lowered larynx
(7) creaky voice

Not all dialect groups will share the same features, and some may even be opposite, but settings that combine some or all of these features are very common, and represent articulatory habits that students can easily observe and learn to recognize.

Spreadness of the lips is common in many dialects of English. Students with excessive rounding at inappropriate moments, for example during /s/ or /l/ which are normally unrounded, can practice smiling slightly as they speak. Slightly rounded segments such as /ʃ/ or /r/ must then be thought of as the marked case where a slight labial adjustment is introduced. Openness is common in American English but not, according to Honikman (1964:75), in British English. The stereotype that Americans speak as though chewing gum has its origins in this setting feature. Accents in many American television programs visually reinforce both spreading and openness.

Palatalization -- fronted and raised tongue body position -- can be illustrated by frequent vowel raising in English, for example in the word *yeah* which may be realized as [jæ:], [jɛə], or [jɪɛ̃ə̃]. Retroflexion of the tongue tip, as in the example of Irish English, can also be identified in West Country or south coast dialects of England, and characterizes rhotic varieties of American English. Nasalization as a voice quality setting is common in much American or British English.

Lowering of the larynx, giving the voice a deeper or hollower sound, often characterizes national political figures or news announcers, where the degree of prestige of the setting

can be assumed to be high. This might be an unusual feature in a corresponding British or French context. Creaky phonation, or a low pitch range, is often present in similar American contexts. Neither feature is necessarily confined exclusively to males.

In conclusion, it is desirable to make ESL students aware of the voice quality settings that characterize their own languages, as well as to present voice quality characteristics which they can use as a model of pronunciation in English. Such models can be referred to analytically to identify the settings of speakers of English whom students hear and observe, or for sensitization within a model of an accent of English which is easy to recognize and to practice. Voice quality comprises the constant background of settings that define both (1) the voice of the individual and (2) the accent of the individual's language variety. While the former are personal, the latter are language-specific and should be described and taught within the pronunciation component of the ESL curriculum.

REFERENCES

Abercrombie, David. 1967. *Elements of general phonetics.* Edinburgh: Edinburgh University Press.

Catford, J. C. 1964. Phonation types: the classification of some laryngeal components of speech production. *In honour of Daniel Jones.* 26-37. David Abercrombie, et al (Eds.). London: Longmans.

Esling, John. 1978a. The identification of features of voice quality in social groups. *Journal of the International Phonetic Association 8*: 18-23.

Esling, John. 1978b. *Voice quality in Edinburgh - a sociolinguistic and phonetic study.* Ann Arbor : University Microfilms.

Honikman, Beatrice. 1964. Articulatory settings. *In honour of Daniel Jones*. 73-84. David Abercrombie, et al (Eds.). London: Longmans.

Laver, John. 1975. *Individual features in voice quality*. Ph.D. dissertation, University of Edinburgh.

Laver, John. 1980. *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Murry, Thomas, Harry Hollien, and E. Müller. 1975. Perceptual responses to infant crying: maternal recognition and sex judgements. *Journal of child language 2*: 199-204.

Murry, Thomas, Pamela Amundson, and Harry Hollien. 1977. Acoustical characteristics of infant cries: fundamental frequency. *Journal of child language 4*: 321-328.

Trudgill, Peter. 1974. *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.