7 Linguistic Phonetics

JOHN LAVER

1 Introduction

Linguistics and phonetics are often characterized as *the linguistic sciences*, implying both similarities and differences. Both linguistics and phonetics are grounded in a basic interest in the nature of human communication. If the subject of *linguistics* is the scientific study of the nature, use, and variety of all aspects of language, the subject of *phonetics* is the scientific study of the nature, use, and variety of all aspects of speech. These are broad definitions of both subjects, and not all linguists and phoneticians would accept such a breadth of scope. But the two subjects have developed so widely in the second half of the twentieth century that a broad view probably better represents the modern diversity of both subjects. What is less controversial is that linguistics and phonetics share a common if partial domain in *phonology*, the study of communicative aspects of spoken language. (In the text below, the first significant mention of a technical term is printed in italics.)

The intersection of linguistics and phonetics in the study of spoken language is visible in the perspectives that each borrows from the other for phonological purposes. Linguistics contributes to phonetics its phonological understanding of the distinctive patterns that make up the coded, conventional aspects of speech which differentiate individual words and other units of spoken language. Phonetics contributes to linguistics its phonetic understanding of the production and perception of the detailed artefacts of speech that embody those significant phonological patterns. Each contribution is complemented by the other. To study formal patterns alone risks becoming over-abstract, and losing touch with the physical realities of spoken language. To study the artefacts of speech without due regard for their identity as conventionally coded signals risks losing sight of the communicative motive of spoken language. The name usually given to the study of spoken language from a phonetic perspective, following the example of Ladefoged (1971, 1997), is *linguistic phonetics*.

2 Linguistic Phonetics and General Phonetic Theory

The objective of linguistic phonetics, which most phoneticians would regard as the center of their professional domain, is to describe the phonetic correlates of phonological units of spoken language and their interactions. Another way of putting this is to say that the ultimate task of linguistic phonetics is to give a comprehensive account of speech patterns and their pronunciations in all languages and dialects of the world. To achieve this task, linguistic phonetics draws on *general phonetic theory*, which is the foundation for the phonetician's understanding of how speech is produced, transmitted, and perceived, against a background of a *general phonological theory* of spoken language. The aim of this chapter is to give a compact account of the shape and content of a model of linguistic phonetics within this framework of a general phonetic theory. Within the current volume, Cohn (chapter 8) presents a summary view of the shape and content of phonological theory, and the reader is referred to her chapter for definitions of basic phonological concepts used here, such as "phoneme," "allophone," "phonological feature," and "phonological syllable."

More extensive presentations of linguistic phonetics than is possible here are available in Abercrombie (1967), Catford (1977, 1988, 1994), Clark and Yallop (1995), Ladefoged (1993, 1997), Ladefoged and Maddieson (1996), and Laver (1994a). Hardcastle and Laver (1997) offer a comprehensive account of the phonetic sciences, including both linguistic and nonlinguistic aspects. A branch of phonetics with particular relevance to both linguistic phonetics and phonology is *acoustic phonetics*. Recommended publications in acoustic phonetics for readers interested in linguistic communication are Kent and Read (1992), Ladefoged (1971, 1993), and Stevens (1998). Laver (1994b) surveys nonlinguistic interests in phonetics, including *paralinguistic* interests in communication of attitudinal and emotional information through tone of voice, and *extralinguistic* interests in matters such as speaker-characterization. Coulmas (1993) provides a comprehensive account of phonetic and linguistic variation in different *sociolinguistic* speech communities. Goldsmith (1995) gives a wide-ranging review of many different approaches to phonological theory.

3 The Scope of Linguistic Phonetics

A comprehensive approach to linguistic phonetics might entail addressing at least four complementary objectives:

1 describing the phonetic basis for differentiating all contrastive (*phonemic*) and contextual (*allophonic*) patterns in speech which signal the identities of linguistic units in any given language;

2 describing the phonetic regularities which distinguish the speech-styles of a given sociolinguistic community from those of others within any given language;
3 describing the idiosyncratic but characteristic phonetic events which distinguish the speech of one member of any given sociolinguistic community from that of other members;
4 describing all recurrent aspects of speech that make one language sound different from others.

All four objectives could be thought relevant to capturing the full extent of the behavioral substance of spoken linguistic communication. Most linguistic phonetic accounts of languages, however, have almost entirely restricted themselves to the first objective. Research by a number of other specialisms has used this first objective as a foundation for pursuing one or more of the other objectives. Sociolinguists interested in the way that speech acts as an index of membership of different communities have investigated the second objective, usually in an urban context. Speech pathologists, and those interested in speaker characterization for other reasons, such as a focus on forensic phonetics, have addressed the third objective. Speech technology has successfully developed automatic systems for speech production, speaker recognition, and language identification (Laver 1994c); but the methods used mostly exploit automated machine learning about hidden statistical patterns in the acoustic waveforms of speech, which doesn't involve explicit "description" in the same sense. No language investigated so far has been comprehensively and explicitly described against all four objectives (though general phonetic theory could in principle be applied to each of these tasks).

Within the first objective, linguistic phonetic accounts also often limit themselves to specifying only the phonetic basis for distinguishing the patterns that contrastively identify one phonological unit from another, for example the consonant or vowel phonemes that discriminate minimally different words in English such as *call* and *tall*, or *seal* and *sill*. The contextual patterns associated with the incidence of contrastive linguistic units in different *structures* and in different *environments* are less often described in detail, rich in phonetic regularity though they are. These aspects of sound-patterning often ignored by linguistic phonetic accounts include the wide range of allophonic realizations of phonemes in different syllable structures and in different contextual environments within syllables.

The limiting of linguistic phonetic accounts of languages to a description chiefly of distinctive phonological contrasts is no doubt because it is seen as a means to a different end. An account of phonological contrasts is all that is normally felt by linguists to be needed for further discussion of linguistic behavior at higher levels than phonology, in morphology, lexis, syntax, and semantics. From the phonetician's perspective, however, once these contrastive patterns have been identified, it is in the phonetic detail of the contextual allophonic interaction of linguistic units that some of the most interesting and challenging phenomena in speech production and perception are to be found.

The presentation of a model of linguistic phonetics in this chapter will give priority to describing the phonetic basis for differentiating the contrastive and contextual patterns in speech which signal the identities of linguistic units in the different languages of the world, but will touch in passing on the other objectives as well. The phonetic symbols used in transcription, enclosed in square brackets [ ], will be those of the International Phonetic Alphabet (1993) of the International Phonetic Association (IPA), set out in what is usually called the IPA Chart. The chart is attached as an appendix to this chapter, for consultation about transcriptional symbols and their classificatory phonetic identification.

4 The Coverage of a Linguistic Phonetic Theory

When the full range of the vocal sound-making capabilities of the human species is considered, it becomes apparent that only a restricted subset of the range is used as the basis for contrastive and contextual patterns in spoken language. To offer a few examples, no language makes distinctive use of the percussive noise of the teeth colliding as the jaw is snapped shut. Nor is the noise of air being squeezed between the cheek wall and the outer surface of the teeth and gums used in language by normal speakers (though it is sometimes used as a substitute for the voice by speakers who have had their larynx removed by surgery). The ability to simulate a snoring sound is not used contrastively, nor is a falsetto voice used deliberately to contrast one sound against another, in any known language.

There is a further degree of constraint. Not only is the range of sounds that is used in language limited to a relatively small subset of those physiologically possible, but within that subset there is a core of frequently used sounds that turn up repeatedly in widely different language-families, within a broader range of less frequent sounds. As part of that core, most languages use [t], [n] and [s] as consonants, as in the pronunciations of English *tea*, *knee*, and *sea*. Relatively few, on the other hand, use consonants such as the initial sounds [f] in English *fin*, [θ] in *thin* or [ð] in *then*. A very large number use the vowels [i], [a], and [u], as in English *peel*, *pal*, and *pool*. But very few use the vowels [y], [ø] or [œ], as in French *lune* ("moon"), *yeux* ("eyes"), or *peur* ("fear") respectively. Only about one-third of all known languages use diphthongs, such as [aʊ], [eɪ] and [ɔɪ], in the word-final syllables of the English verbs *allow*, *allay*, and *alloy* (Lindau et al. 1990).

There seem to be five interactive principles that may explain this human tendency to use a somewhat restricted number of sound-types for purposes of linguistic communication (Lindblom 1983, 1986, Ohala 1989, Stevens 1972). These are:

1 perceptual stability;
2 adequate perceptual contrast;
3 ease and economy of articulatory performance;
4 ecological robustness;
5 ease of modifiability to the needs of the communicative situation.

Perceptual stability is achieved by languages tending to use sounds for which small articulatory adjustments make little auditory difference. Maintaining adequate perceptual contrast entails avoiding sound-differences close to the limits of human discrimination. Ease and economy of articulation are the outcome of choosing sound-types which do not unduly tax the capabilities of the speech production system. Ecological robustness reflects the ability of sounds to resist the perceptual masking effects of other sounds likely to be heard in the environment (especially speech from other speakers). Finally, given that the relative speed, loudness, and articulatory precision of the speech of a given speaker change frequently in response to variations in the social and physical circumstances of the conversation, it is helpful if parameters of speech control are used which can be appropriately modified without damaging intelligibility.

Different languages, and a given language at different times, reach differing solutions to the trading relationships between these five principles. That these solutions are not always optimal is one potential basis for the sound patterns of languages changing over time.

5 The Shape of a General Phonetic Theory

The obedience of spoken language to the five principles described above has an impact on the desirable shape of a general phonetic theory. A well-designed general phonetic theory is one whose posited features and organizational units cover the maximum range of data with the simplest descriptive constructs. If spoken languages in general tend most frequently to favor a core of speech sounds which are perceptually stable, adequately contrastive, relatively easy to articulate, ecologically robust, and intelligible in variable circumstances, then the basic constructs to be set up in general phonetic theory should be the ones whose nature and relationships give the simplest and most economical account of such sounds. The theory is then completed by adding a minimum set of more elaborate constructs, to cover the less frequent and usually more complex sounds.

6 Organic and Phonetic Aspects of Speech

Within the model of general phonetic theory to be offered here, it will be convenient first to distinguish *organic* versus *phonetic* factors in speech. *Organic* *factors* are those which are "to do with anatomical structure or morphology, and with the constraints which that structure imposes on the potential for physiological action" (Mackenzie Beck 1997: 256). *Phonetic factors* are those which arise from any learnable aspect of use of the vocal apparatus, with its acoustic and perceptual correlates (Laver 1994a: 28). The interplay between organic and phonetic factors in speech is one of the major sources of acoustic variation between different speakers. The recovery of relatively invariant properties in speech data from different speakers, to aid the decoding of linguistic messages (Perkell and Klatt 1986), can only be achieved by resolving the relative contributions of organic and phonetic factors.

Many theoretical and practical consequences arise from the fact that any two speakers of normal anatomy must be treated as capable of producing phonetically identical utterances, despite the often very substantial organic differences between them. The fact that the vocal organs of different speakers can be of very different sizes means that speech from two individuals can be acoustically very different, in absolute physical terms. Comparability of pronunciation therefore arises from considering not the absolute values of acoustic parameters, but their values relative to the individual speaker's own acoustic potential. So the intonational value of the pitch of a large adult male speaker's voice can be compared to that of a small female child by considering in each case whether the pitch should be counted as high (or mid, or low), in relation to the speaker's own pitch range (Ladd 1996). In absolute terms, the voice pitch ranges of these two speakers would be very unlikely to show any physical overlap at all. In relative terms, however, they can be brought into comparability, and when heard as the same in these terms they can be regarded as phonetically equivalent.

The same situation applies to comparisons of the phonetic quality of different speech sounds. Vowel-sounds, for example, are acoustically characterized by patterns of resonant frequencies of the vocal tract (Ladefoged 1993). The absolute values of the resonant frequencies depend on the overall length and shape of the tract. These frequencies change as the organs of the vocal tract manipulate it into different configurations, within organic limits set by individual anatomy. The configurations of two vocal tracts can be thought to be phonetically equivalent when the ratios of the lowest resonant frequency to higher resonant frequencies in each of the two cases are closely similar. In absolute terms, given that the resonant-frequency ranges for two such organically different speakers as the large man and the small girl would once again show virtually no overlap, it would not be feasible to say that these two speakers were producing comparable sounds. In relative terms, however, they can both be perceived as producing the same vowel [uː] in their pronunciations of the English word "boot" [buːt], for instance, when the resonant frequencies of each of them show appropriately similar ratios.

Phonetic equivalence is one end-point of a more general scale of *phonetic similarity*, which is a metric for comparing the phonetic characteristics of any two sounds. The concept of phonetic similarity is hence a necessary basis for the whole of general phonetic theory. In addition, the view that organically different speakers can produce and perceive phonetically equivalent sounds has profound implications for describing normal use by native speakers. Equally profound are the implications for understanding the articulatory and perceptual processes of spoken language acquisition by infants, foreign-language learning by non-native speakers, and pathological use in speech disorders.

The dimension of phonetic similarity is relevant, finally, not only to comparing speech sounds from all different speakers of normal anatomy, but also to two further situations. The first is as the basis, within a single speaker, for grouping phonetically similar allophonic variants into a single phoneme, as a family of phonetically related sounds fulfilling the same contrastive phonological role. The second applies to decisions about the range of phonetic segment-types that can be represented by a given character in alphabetically based writing systems for whole language communities. The decision, for instance, about what speech sounds in different languages are eligible to be written with the letter "r" depends in part on the comparability of the phonetic and perceptual qualities of the candidate sounds concerned.

7 Articulatory, Acoustic, and Perceptual Levels of Description of Speech

Emerging from the discussion in the section above is a second general distinction, between three different aspects of the phonetic description of speech. These are related to the three links in the chain of speech, from the speaker's generation of an utterance, to its transmission through the air, to its reception by the listener. The first is the *articulatory* level of description, which accounts for the changing configurations and other actions of the speaker's vocal apparatus. The second is the *acoustic* level, which consists of statements about the physical consequences of articulatory actions in terms of vibratory patterns of air molecules within the vocal apparatus and in the air between the speaker and the listener. Finally, the third level of description concerns the *perceptual* impressions of the listener receiving the acoustic information.

The rest of this chapter will focus on phonetic aspects of speech, and will be concerned chiefly but not only with the articulatory level of description.

8 Linear and Non-linear Units of Speech Organization

The phonetic events that make up the time-course of speech tend to be continuous, with only relatively few steady states or sharply defined breaks that could serve as the boundaries of natural, serial units of speech organization. Obvious natural breaks do occur, however, in two circumstances in the linear production of speech by a single speaker. One is at the beginning and end of a *speaking-turn* by one participant in a conversation. The other is at the beginning and end of an individual *utterance*, bounded by silence, within the individual speaking turn. Exhaustively dividing the rest of the stream of speech into a sequence of units smaller than the utterance involves appealing to a number of convenient assumptions. A key traditional assumption is that the continuum of speech can be appropriately handled, for analytic purposes, as if descriptive categories were discrete, not continuous. On this basis, it becomes reasonable to set up smaller-scale phonetic constructs such as the feature and the segment.

8.1 The relationship between phonetic segments and phonetic features as units of speech production

Phonetic features are collectively the ingredients of *phonetic segments*. In the minimum case, two segments may differ from each other by the presence or absence of just one phonetic feature. A feature exploited in every human language in this way is the phonetic feature of "voicing." Voicing is caused by vibration of the vocal folds in the larynx. Whether the vocal folds vibrate or not will be determined by the interaction of airflow from the lungs and the tension-states of relevant laryngeal muscles. The word-initial consonant-sounds in the two English words *zeal* /zil/ ⇒ [ziːl] and *seal* /sil/ ⇒ [siːl] differ in their voicing state, in that the vocal folds are being made to vibrate in the first case (making [z] a "voiced" segment) and not in the second (making [s] a "voiceless" segment). The transcriptional conventions in the example above are that slant brackets // show the phonemic status of the symbols; "⇒" means "is phonetically pronounced as"; square brackets [ ] show the phonetic status of the pronunciation of the words concerned; and [ː] after a segment means that the sound is produced "with relatively longer duration."

While segments can be thought of as linear units following one another sequentially in the chain of speech without interval, features are non-linear. They can overlap each other in time, and have start-points and end-points which do not necessarily align with those of the chain of segments. Phonetic segments, representing phonological vowels or consonants, are temporally anchored in the chain of speech by the cooccurrence and mutual timing of their constituent features.

8.2 Phonetic and phonological features

The constructs of a general phonetic theory should include a supposed *universal set of phonetic features*, whose comprehensive coverage of spoken language remains provisionally true until shown by further research to be inadequate. The general phonetic theory summarized here is based on these principles (Laver 1994a). It tries to include a set of phonetic features capable of describing the phonetic basis of all phonological contrasts, and of all the contextual patterns of their interaction, so far discovered in the spoken languages of the world. The set of phonetic features proposed in a general phonetic theory would nevertheless, in the ideal, always be larger than the set proposed to cover the languages of the world. This is because of the need to extend its coverage to the differentiation of sociolinguistic communities and the characterization of individual speakers.

It is important to appreciate the difference of technical status between descriptive phonetic features and distinctive phonological features. An example was quoted earlier of the phonetic feature of voicing providing the phonetic basis for a minimal contrast between two consonantal sounds in English, /z/ and /s/. Opportunities for conceptual confusion are rife at this point, in that the presence or absence of "voicing" can be seen in two quite different perspectives. Phonetically, the difference between [z] and [s] as physical speech sounds is described in terms of the presence or absence of vibration of the vocal folds, as mentioned briefly above and described in more detail in section 10.2 below. To expand on the phonetic example mentioned briefly earlier, /z/ and /s/ as consonants in English are phonologically differentiated by the distinctive presence or absence of a single *distinctive feature*, often represented as +VOICE versus –VOICE. (Capitalization of the name of the feature, with "+" and "–" indicating presence versus absence, is a useful way of distinguishing the status of phonological features from that of phonetic features, which often – potentially confusingly, as in this case – have the same or similar names.)

Viewed as a phonetic feature, "voicing" is part of the descriptive, objective vocabulary of phonetics. Viewed as a distinctive feature, VOICE is part of the formal vocabulary of phonology. The purpose of phonetic features is to describe the articulatory, acoustic, or auditory characteristics of speech sounds as events in the real, physical world, independently of the language concerned. The purpose of distinctive features is to focus on the role of the features as part of a conventional, semiotic code for identifying phonological units particular to a given language. The term "distinctive feature" is thus reserved for use as a contrastive phonological concept.

Part of a phonological interest in distinctive features is the exploration of the degree to which different phonological features fall into putatively *natural classes*, where the members of the class share some phonetic and / or distributional property that distinguishes that class from other classes. This often entails grouping classes into more abstract, superordinate classes, such as the phonological class of "sonorant." This superordinate class is normally taken to include the subordinate classes of English vowels, liquids (such as /r, l/), glides (/j, w/) and nasal stops (/m, n, ŋ/). For further discussion of distinctive features and natural classes, see the chapter on phonology in this volume by Cohn (8).

8.3 *The phonological syllable*

The *syllable* is not identified here as a unit of phonetic description. Many phoneticians have tried to develop a robust definition of the properties of a phonetic syllable, but no objective correlate that would link phonetic performance on a one-to-one basis to the phonological syllable has yet emerged (Laver 1994a: 113–15).

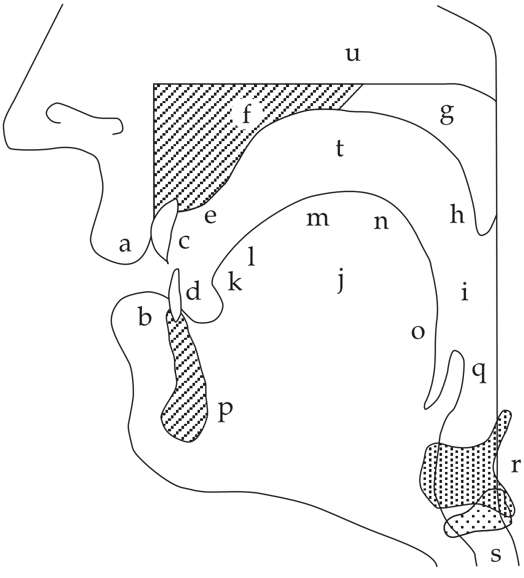
The term "syllable" is perhaps best reserved for use at the phonological level, where it is useful (though not itself unproblematic) for two purposes: for the location of word-identifying patterns of stress; and as an organizing concept for the mutual distribution of vowels and consonants. This organization is reflected in the traditional phonological view that *vowels* are nuclear in the syllable, with all syllables containing one and only one vowel.

Consonants are marginal in syllables, being either syllable-initial or syllable-final. Using "C" to mean "a consonant," and "V" to mean "a vowel," the structure of an English monosyllabic word like "strikes" /straɪks/ would be formulaically represented as CCCVCC. Languages differ in the syllable structures they allow. English allows both *open* and *closed* syllables (that is, syllables without and with, respectively, one or more final consonants), as in /aɪ/ "I" V, /saɪ/ "sigh" CV, /saɪd/ "sighed" CVC, and /saɪzd/ "sized" CVCC. Hawaiian allows only open syllables, as in the di-syllabic word /ola/ "life" V + CV.

9 The Componential Organization of Speech Production

The success of phonetics in developing an objective, replicable, internationally standard method of describing all speech sounds in all spoken languages lies in part in a componential approach to phonetic description. Each discriminable sound is regarded as the composite product of the action of a number of sub-processes of the speech production system. These are described in more detail in section 10 below. A schematic view of the *vocal organs* which make up the sub-processes, including the *lungs*, the *larynx*, the organs of the *mouth* and the *pharynx* in the vocal tract, and the soft palate (technically called the *velum*), is shown in figure 7.1.

This componential analysis underlies the conventions of phonetic transcription of the International Phonetic Association. As an illustration of this approach, and to inform the explanation offered below of descriptive phonetic categories, a typical (abbreviated) label for the sound represented in the phonetic transcription of the IPA's International Phonetic Alphabet (1993) as [b] would be "a voiced, labial, oral stop." The four elements of this label constitute individual phonetic features and identify independently controllable components of the production of the sound:



a. Upper lip	h. Uvula	o. Root of the tongue
b. Lower lip	i. Pharynx	p. Lower jaw
c. Upper teeth	j. Body of the tongue	q. Epiglottis
d. Lower teeth	k. Tip of the tongue	r. Cartilages of the larynx
e. Alveolar ridge	l. Blade of the tongue	s. Trachea (windpipe)
f. Hard palate	m. Front of the tongue	t. Oral cavity
g. Velum (soft palate)	n. Back of the tongue	u. Nasal cavity

Figure 7.1 Schematic diagram of a cross-section of the vocal organs
Source: After Laver 1994a: 120

- "*voiced*": the vocal folds in the larynx are vibrating (superimposing aerodynamic pulses on the moving column of air flowing out of the lungs);
- "*labial*": the lips are involved as articulators;
- "*oral*": the velum is in a raised position, sealing off the exit at the back of the mouth to the nasal cavity, causing any airflow to pass through the mouth alone;
- "*stop*": the closure of the lips momentarily seals off the escape of the air in the mouth and pharynx to the outside atmosphere, causing a short-term rise in air-pressure in the vocal tract. As the lips open again, the compressed air is then released through them with a small, audible explosion.

An assumption in such abbreviated labeling is that the flow of air is generated by the action of the lungs, with the flow being out of the body. (In a fuller label, this would add an explicit element "with pulmonic egressive airflow" – see section 10.1.) By identifying the activities of different sub-processes in this way, and with an underlying understanding of the activities thus represented, a componential labeling system in effect offers a set of instructions to informed readers about what to do phonetically with their own vocal apparatus to generate a phonetically equivalent or near-equivalent sound.

10 Speech Production Processes

There are only two basic ways in which a speaker can perceptibly differentiate one segmental speech-sound from another – by changing the *phonetic quality* of the sound, or its *duration*. Variation of pitch and loudness play their part at a suprasegmental level when speech is continuous, and may result in differences of meaning, but matters of the prosodic and metrical control of speech production will not be addressed in this chapter. For interested readers, these topics are discussed in Laver (1994a: 450–546), together with issues to do with continuity and rate of speech.

The control of phonetic quality and duration depends on the interaction of five major sub-processes in the production of speech:

1 initiation and direction of airflow;
2 phonation type;
3 articulation;
4 inter-segmental co-ordination;
5 temporal organization.

The remainder of this chapter is devoted to an explanation of the way that the activities of these different sub-processes in the control of speech can generate different sounds.

The traditional phonetic approach to the *segmental classification* of speech-sounds is said to be a classification by "place and manner of articulation." "Place" will be seen to be straightforward, but "manner" will turn out to be a complex of a range of different types of activity. Segmental classification by place and manner draws on all the factors in the list above, with their interaction producing segments of different phonetic quality and duration.

Description in the sections below will concentrate on the typology of phonetic features, rather than on exemplifying every cell of the resultant matrix of categories. For a comprehensive account of both segmental and suprasegmental categories of speech sounds, the reader is referred to Laver (1994a: 95–546). For the interpretation of specific phonetic symbols, the IPA Chart in the appendix to this chapter should be consulted.

10.1 Initiation and direction of airflow

There are three categories of *initiation of airflow* used for speech, and two of *direction of airflow*. The means of setting a column of air moving can be classified in terms of the initiating mechanism used. By far the most frequent initiator of airflow in speech is the pulmonic mechanism, setting lung air flowing in an egressive direction to the external atmosphere. This *pulmonic egressive airflow* is then modified in turn by the actions of the larynx, the vocal tract and the velum. Speech made on a *pulmonic ingressive* mechanism, on an inflowing breath, seems to be used only paralinguistically, for example in Scandinavian cultures to express sympathy or commiseration.

The second initiator of airflow used in speech is the larynx, in the *glottalic airstream mechanism*. As the name suggests, the *glottis* (the space between the vocal folds) is involved. With the glottis acting as a valve, and closing off the flow of air from the lungs, the larynx can be abruptly raised or lowered in the throat by muscular action, like a piston in a cylinder. The effect is to compress or rarefy the volume of air in the vocal tract, causing a sharp explosion (on release of a compressed *glottalic egressive* airstream) or an abrupt implosion (on the release of a rarefied *glottalic ingressive* airstream). Sounds made on a glottalic egressive airstream are called *ejectives*, and those on a glottalic ingressive mechanism *implosives*. In Zulu, the word [k'ɑːk'ɑ] "surround" involves two ejectives, symbolized by the apostrophe ['] after the stop symbols.

If voicing is added to an implosive, by pulmonic egressive airflow making the vocal folds vibrate as the larynx descends during a glottalic ingressive initiation, a *voiced implosive* segment is the result. The contrastive difference between the two Hausa words [ɓaɓe] "estrangement' and [babe] "grasshopper" relies on the two stop segments in the first word being voiced labial implosive stops (hence involving two airstream mechanisms, glottalic ingressive and pulmonic egressive), and in the second on the two stop segments being voiced labial pulmonic egressive stops (involving only one airstream).

The third initiator of airflow used in speech is the tongue, in the *velaric airstream mechanism*. Because the tongue is involved, it is therefore also sometimes called the "lingual" mechanism. Velaric sounds are made by the body of the tongue trapping a volume of air between two closures in the mouth, one at the velum, and one further forward. The tongue then retracts the velar closure by sliding backward along the soft palate while maintaining the closed stricture, thus rarefying the air pressure enclosed in the expanded, sealed cavity. When the front closure is then released, the air implodes into the relative vacuum. Sounds made on this *velaric ingressive* airstream are called *clicks*. Since the velaric mechanism is confined to actions within the mouth, the rest of the vocal apparatus is free to add voicing and / or nasality to click sounds.

The languages that use click sounds contrastively are confined to southern and eastern Africa. Ladefoged and Maddieson (1996: 246–80) offer a comprehensive account of these sounds, with many examples from languages such as Nama, Zulu, and Xhosa. In English, clicks are used only paralinguistically, to indicate annoyance (usually written "tsk, tsk" in the English writing system), or to encourage horses to accelerate, or onomatopoeically to simulate the clopping-sound of their hooves.

10.2 Phonation type

The biological function of the larynx is chiefly to act as a protective and regulative valve for the airway to and from the lungs. The valving mechanism that has evolved is a delicate and complex muscular structure within a supporting framework of cartilages (Dickson and Maue-Dickson 1982, Laver 1980). The so-called *vocal folds* are two shelves of muscular tissue which run horizontally from front to back of the larynx, capable of separation at the back to leave a flat, triangular space with its apex at the front. This space was identified earlier as the glottis, and there are six modes of *phonation* used in spoken language to distinguish different segments, involving different adjustments of the glottis.

When pulmonic egressive air flows upwards from the lungs, a *voiceless* sound is produced if the triangular space of the glottis is left wide open, as if for breathing out. Examples of voiceless consonant-sounds widely used in languages are the word-initial sounds in English *see* [siː], *tea* [tiː], and *she* [ʃiː]. If the vocal folds are brought close enough together to make the continuous airflow through them turbulent, either through a gap left at the back or through a narrowed glottis, the result is called *whisper*.

Voicelessness can be heard in the pronunciation of some vowels in a number of languages. As an allophonic process before pauses, (described below as an outcome of the coordinatory process called "devoicing"), French vowels often lose their voicing. An example would be [wi̥ː] *oui* ("yes") at the end of an utterance, where [̥] below the symbol indicates voicelessness. Alternatively, this devoicing is often substituted by whisper, rather than strict voicelessness. English also exploits allophonic voicelessness, in optional pronunciations of unstressed vowels between two voiceless consonants, as in the first syllable of *potato*/pəteɪtəʊ/ ⇒ [pə̥teɪtəʊ] in Received Pronunciation of British English. Further examples of voicelessness or whisper on vowels in Amerindian, Sudanic, Sino-Tibetan and Australian languages are given in Ladefoged and Maddieson (1996: 315) and Laver (1994a: 295–7).

In the third type of phonation, vibration of the vocal folds is the basis for *voiced* sounds, as mentioned briefly in sections 8.1 and 9 above. Examples of voiced sounds widely used in languages are the word-medial consonant-sounds [z, d, g, m] in English *easy* [iːzɪ], *aiding* [eɪdɪŋ], *again* [əgeɪn], and *seeming* [siːmɪŋ], as well as the vowel-sounds in these words [iː, ɪ, eɪ, ə]. In voiced sounds, the vocal folds are brought lightly together by muscular action, blocking off the outflow of pulmonic air, and air pressure below the closed folds building until it is sufficient to blow the folds apart against the muscular tension holding them closed. Once airflow is re-established through the glottis, an aerodynamic effect is produced within the glottis, with the egressive pulmonic flow creating very local suction as it passes at high speed through the relatively small gap between the vocal folds. This local force sucks the vocal folds towards each other, and combines with the muscular tension to restore the closed position of the vocal folds. The abrupt restoration of closure sends a small shockwave traveling on the outflowing breath through the vocal tract, and acoustically excites it into resonance.

The cycle from closure of the vocal folds to separation and renewed closure typically happens very fast (in a range from 60 to 240 times per second in adult male voices in normal conversational English). The frequency of the vibration corresponds to the auditory *pitch* of the voice. The contour of pitch in the successive, intermittent voiced sounds of a whole utterance is in effect heard as a melody, and functions as the *intonation* of the utterance.

The fourth type of phonation used in spoken language is *creak* or *creaky voice* ("creak" is also sometimes called *vocal fry* or *glottal fry* in American publications). In this mode of phonation, the front part of the glottis vibrates, at a considerably lower frequency than in normal voicing, while the back part is pressed more tightly together. Pairs of Danish words can be distinguished by the presence of syllable-final creak (sometimes also called *laryngealization*) versus its absence, for instance in [dṵ] "tablecloth" versus [du] "you" (Laver 1994a: 330–3).

The fifth type of phonation is *whispery voice* (also sometimes called *breathy voice* or *murmur*). As in whisper, the vocal folds do not completely seal off the trans-glottal escape of the pulmonic airflow while vibrating, but leave a gap – either at the back of the glottal triangle, or along the length of the approximated but vibrating vocal folds. The result of the continuous leakage of air is to superimpose audible whisperiness on the pulsed voicing throughout the phonation. Whispery voice in English is used phonetically in English as an optional allophonic feature to replace the normal voicing of [h] when that consonant occurs in inter-vocalic position in some accents of British English, as a whispery voiced resonant [ɦ]. Examples are *ahead* /əhɛd/ ⇒ [əɦɛd] and *perhaps* /pəhaps/ ⇒ [pəɦaps], with phonetic voicing running right through these words, becoming momentarily whispery during the "h". Whispery voice of this sort is also used in English paralinguistically throughout an utterance to signal secrecy or confidentiality.

In a range of other languages, whispery voice is used contrastively to distinguish one consonant phoneme from another. An example is [bəla] "a snake" versus [bɦəla] "good" in Sindhi, using [ɦ] in association with the [b] symbol to indicate a whispery-voiced beginning to the syllable in the second word. Section 10.4 below classifies this as an inter-segmental coordinatory instance of "voiced aspiration" (Laver 1994a: 354).

Finally, closure of the vocal folds may itself constitute the medial phase of a stop segment, in which case it is called a *glottal stop* [ʔ]. Glottal stops are used only allophonically in English, for example as a phonetic realization of the final /t/ consonants in London Cockney *eat that pizza* /iːt ðat piːtsə/ ⇒ [əiʔ ðaʔ pʰəiʔsʌ].

10.3 Articulation

A key part of appreciating how descriptive phonetic classification works is understanding the relationship between segments and features. This section on articulation begins with a clarification of this relationship, and then discusses principles of classification by place of articulation, degree of stricture, multiple degrees of stricture and aspect of articulation. The technical vocabulary introduced in this section is then used in the discussion of intersegmental coordination.

10.3.1 Featural phases of the structure of segments

The complex relationship between segments and features can be clarified by appeal to the concept of three internal *phases* of a segment – the onset phase, the medial phase and the offset phase (Laver 1994a: 112–13). The configuration of the vocal tract during speech changes dynamically from moment to moment between variably greater and lesser degrees of local constriction of the airflow. These constrictions are created by a mobile, *active articulator* (such as the tongue, or lower lip) moving towards a fixed or less mobile, *passive articulator* (such as the hard palate, soft palate, or upper lip). The time occupied in maintaining the maximum degree of articulatory constriction (or degree of *stricture*) reached by the vocal tract during the production of an individual segment delimits the *medial phase* of the segment's performance.

During the *onset phase* of a segment the active articulator is approaching the maximum stricture, and in the *offset phase* is moving away from this towards the configuration for the medial phase of the next segment. One segment's offset phase overlaps with the onset phase of the next segment, in an *overlapping* phase.

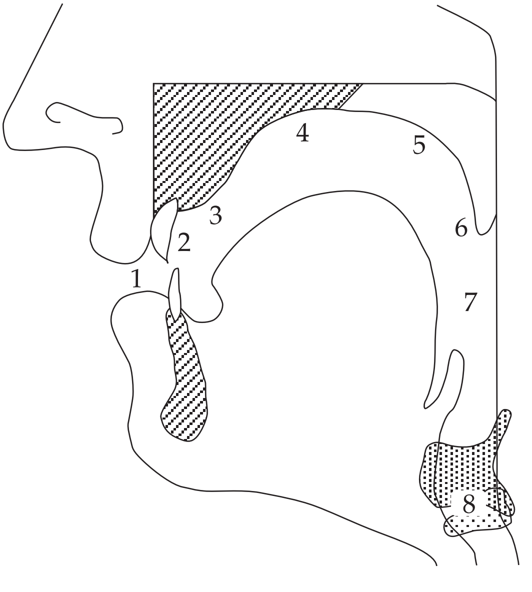
The concept of the phasal structure of segments is important for two reasons to do with the temporal distribution of phonetic features. The first is that a given feature may start or finish within a particular segmental phase. For example, in English syllables anticipatory nasality begins relatively early within the medial phase of a vowel-segment before a nasal consonant-segment, as in *calm* /kam/ ⇒ [kʰãːm]. (In the IPA transcription here, superscript [ʰ] means "aspiration," or "voice onset delay," [˜] means "is nasalized," with the soft palate open to allow airflow into the nasal cavity.)

A given feature may alternatively be co-terminous with the medial phase of the segment, as in the case of audible friction in [θ] in English *thin* /θɪn/ ⇒ [θɪ̃n]. Or the feature may run through the medial phases of two or more adjacent segments, as in the English word *soon*, /sun/ ⇒ [sʷũːn], where lip-rounding runs through the first two segments, relaxing to a neutral position towards the end of the word. (The vowel-segment [u] is inherently lip-rounded, and in the case of consonant-segments lip-rounding is phonetically symbolized by the attachment of the diacritic [ʷ] – see also section 10.3.4 below on multiple degrees of stricture.) A feature running through adjacent segments can be called a *setting* (Laver 1994a: 115, 391–427), and an analysis of features into settings is useful not only for linguistic phonetics, but also for *paralinguistic* analysis of affective or emotional communication through tone of voice, and *extralinguistic* analysis of speaker-characteristics (Laver 1980, Nolan 1983, Pittam 1994).

10.3.2 Place of articulation

Classification by *place of articulation* identifies the location of the articulatory zone in which the active articulator is closest to the passive articulator during the medial phase of a segment. An enabling concept for approaching this classification is to distinguish between neutral and displaced places of articulation. In its *neutral configuration*, the vocal tract is as nearly as anatomy allows in equal cross-section along its full length from lips to pharynx. (If a vowel-sound were to be produced in such a configuration, it would have the quality of the "neutral" vowel [ə] in the pronunciation of the first (unstressed) syllable of the English word *canoe* [kənu]; and acoustically the resonant frequencies would be such that the ratio of the higher frequencies were odd multiples of the lowest.)

In the neutral configuration, the potential active articulators (the lower lip and the tip, blade, front, back and root of the tongue) lie in their natural anatomical position opposite their passive counterparts along the longitudinal axis of the vocal tract. A segment whose place of articulation is neutral is made by an active articulator moving towards its neutral, passive counterpart. The neutral configuration of the vocal tract, and some labels for *neutral places of articulation*, are given in figure 7.2.



1. Labial
2. Dental
3. Alveolar
4. Palatal
5. Velar
6. Uvular
7. Pharyngeal
8. Glottal

Figure 7.2 Schematic diagram of some of the neutral places of articulation
Source: After Laver 1994a: 135

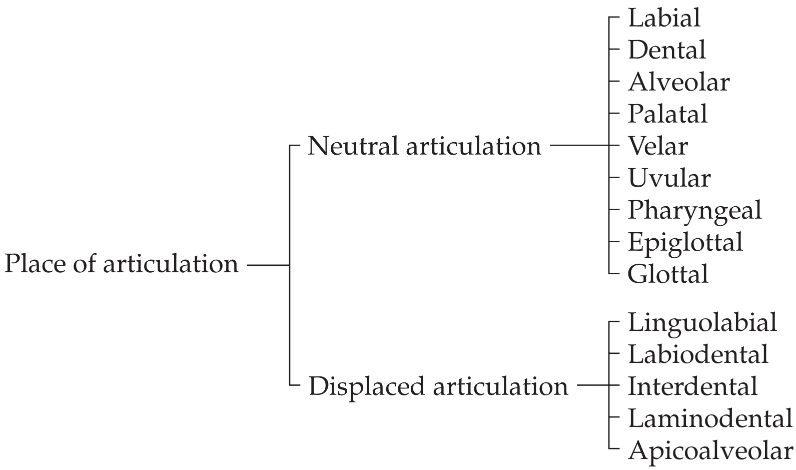


Figure 7.3 Labels for neutral and displaced articulations

Neutral places of articulation are thus involved when the bottom lip moves up against the top lip to create a *labial* articulatory narrowing or closure; when the tip of the tongue touches the inner surfaces of the central upper incisors to make a *dental* closure; when the blade of the tongue articulates against the alveolar ridge behind the teeth to make an *alveolar* closure; or when the back of the body of the tongue contacts the soft palate to create a *velar* closure.

When the bottom lip is retracted from its neutral place, however, to articulate instead against the central upper incisors, or the tip of the tongue is retracted to articulate against the alveolar ridge behind the teeth, the resulting *labiodental* and *apicoalveolar* strictures are classified as *displaced articulations*. Similarly, if the blade of the tongue is protruded between the lips, and makes a *linguolabial* closure against the upper lip, that too is a displaced articulation. The labels for neutral and displaced articulations are shown together in figure 7.3.

Examples of neutral articulations in English are all vowels (except those in some accents where the tongue-tip is curled upwards in anticipation of a following /r/), and the word-initial consonant-sounds in *pea* [piː] and *bee* [biː] (both labial); *theme* [θiːm] and *thee* [ðiː] (both dental); *teal* (tiːl] and *deal* [diːl] (both alveolar); *cash* [kaʃ] and *gash* [gaʃ] (both velar); and *he* [hiː] (glottal). Instances of displaced articulations in English are the word-initial consonant-sounds in *feel* [fiːl] and *veal* [viːl] (both labiodental).

Setting up the classificatory distinction between neutral and displaced articulations amounts to a claim about the relative frequency of incidence of different sounds in the languages of the world. The simpler, less elaborate concept of neutral articulations underpins a broadly sustainable assumption that neutral labial, dental, alveolar, palatal, velar, and glottal sounds are more frequently encountered, for instance, than the displaced linguolabial, labiodental, and apicoalveolar sounds. However, this claim becomes less successful when one considers the relative infrequency in the languages of the world of neutral uvular and pharyngeal sounds, for whose relative rarity more specific reasons would have to be advanced.

10.3.3 *Degree of stricture*

Classification by *degree of stricture* answers the question: "In the medial phase of the segment, to what degree is the gap between the active and passive articulators narrowed?" Languages exploit three types of segments defined by the criterion of degree of stricture – stops, fricatives, and resonants. In the medial phase of *stops*, the degree of stricture is one of *complete articulatory closure*. Examples from English are the word-initial consonant-segments [p, b, t, d, k, g] in *post*, *boast*, *toast*, *dosed*, *coast*, and *ghost* respectively.

In *fricatives*, the articulatory stricture in the medial phase is one of *close approximation*, with the airflow made turbulent by passing through a very narrow gap between the active and passive articulators, generating an audible hissing noise ("friction"). Examples of fricatives are the word-initial consonant-segments [f, v, θ, ð, s, z] of English *fan*, *van*, *thigh*, *thy*, *sink*, *zinc*, or the word-medial consonant-segments [ʃ, ʒ] in English *mesher* and *measure*.

In the medial phase of *resonants* (which can involve sounds representing both consonants and vowels), the stricture is one of *open approximation*. This is a stricture which is sufficiently open to allow the airflow to pass smoothly without turbulence. Open approximation is optimal for allowing the pulses of voiced vibration from the larynx to set the vocal tract into resonance as an acoustic tube.

Examples in an accent of British English of resonants which act as consonants are the word-initial segments representing /j, w, r, l/ in *yield* [jiːld], *wield* [wiːld], *raw* [ɹɔ] and *law* [lɔ]. Examples of resonants acting as vowels from the same accent are the word-final segments in *bee* [biː], *Shah* [ʃɑː], *paw* [pɔː] and *two* [tuː]. The IPA chart subclassifies the open-approximation degree of stricture of such resonants in terms of three further articulatory dimensions. The first two are divisions vertically and horizontally of the *vowel-space* in the mouth within which the highest point of the regularly curved tongue is located for the resonant in question. The vertical division is subdivided into *close, close-mid, open-mid*, and *open* resonants. The horizontal division is subdivided into *front, central*, and *back* resonants. The third classificatory dimension for resonants acting as vowels describes the lip-position of the segment, divided into *rounded* and *unrounded*. The resonant in *bee* [biː] is close, front, unrounded; in *Shah* [ʃɑː] is open, back, unrounded; in *paw* [pɔː] is open-mid, back, rounded; and in *two* [tuː] is close, back, rounded.

10.3.4 *Aspect of articulation*

The concept of *aspect of articulation* extends the concepts of "neutral." It is suggested that the majority of stops, fricatives, and resonants in the languages of the world will be performed with the tongue in a regularly curved shape (convex both longitudinally and laterally), with the velum closed, and with a stricture maintained more or less as a steady state throughout the medial phase in a single, neutral place of articulation. This set-up will be treated as a neutral reference against which three non-neutral groups of aspects of articulation can be described. These are the conformational, topographical, and transitional aspects (Laver 1994a: 140–7).

The *conformational aspects* deal with the routing of the airflow channel. There are three distinctions to be drawn. The first is between *oral* airflow versus *nasal airflow*. The second is between *central* versus *lateral airflow*. The third is between *single* versus *multiple strictures*.

As instances of differences between oral and nasal sounds, neutral voiced *oral stops* include [b, d, g], as in English *bib* [bɪb] (*oral labial stops*), *did* [dɪd] (*oral alveolar stops*) and *gig* [gɪg] (*oral velar stops*) respectively. Their non-neutral *nasal stop* counterparts are [m, n, ŋ], as in English *mum* [mʌm] (labial), *none* [nʌn] (alveolar) and *sung* [sʌŋ] (velar). An allophonic difference between an oral and a *nasal fricative* at the same place of articulation is in Igbo "to wedge in" [ɪfa] versus "to shriek" [ɪ̃fã] (Williamson 1969: 87), from Nigeria. Here both are non-neutral in a different respect, in that they share a displaced labiodental place of articulation. A phonemic difference between an oral and a *nasal resonant* can be found in Sioux "sun" [wi] versus "woman (abbreviated form)" [wĩ] (J. Harris, personal communication).

Stop articulations can show complex aspectual patterns of oral and nasal sequences within the medial phase of a stop. The place of articulation of the oral and nasal elements are *homorganic* – the oral stricture is at the same place of articulation. When the nasal element is minor compared with the duration of fully oral closure, and occurs at the beginning of the medial phase, the stop is said to be *pre-nasal*; when it is final with respect to the oral closure, it is called a *post-nasal stop*. The duration of such nasal elements is shorter than in full segmental sequence of nasal + oral stops, as in English *candor*, for instance. When the nasal element dominates the duration of the oral closure in the medial phase, it is said to be a *pre-occluded* or *post-occluded nasal stop*, depending on the initial or final location of oral closure. Examples of *complex oral / nasal stops* are found in a range of languages, including some in Africa, India, and South and Central America. An instance of pre-nasal stops comes from Kalam, a Papuan language of New Guinea, in "down valley" [^{m}bim] and "sinew" [kɨ^{n}dɨl] (Pawley 1966). A fuller discussion of such complex oral / nasal stops is offered in Laver (1994a: 227–35).

For the sake of economy, sounds will from now on be assumed to be oral unless specific mention is made of their nasal status.

In the case of differences between central and lateral sounds, a neutral example would be the *voiceless alveolar central fricative* [s], as in English *sea* [siː]. A non-neutral instance would be a *voiceless alveolar lateral fricative* [ɬ], as in North Welsh "her ship" [iɬɔŋ], which is in phonemic contrast with a *voiced alveolar lateral resonant* "his ship" [ilɔŋ] (Albrow 1966: 2). In both lateral cases, the air flows through a gap at one or both sides of the tongue behind a central contact between the tip or blade of the tongue against the alveolar ridge.

The active articulators of the vocal tract are sufficiently flexible and versatile to be able to create articulatory strictures in two different places simultaneously (i.e. sharing the same medial phase). When two such strictures are of equal degree the conformational aspect of articulation shows (non-neutral) *double articulation*. Two examples from the West African language Yoruba are the words [k͡pe] "to call" and [g͡be] "to carry" (Bamgbose 1969: 164). The two simultaneous closures in these double stops [k͡p] and [g͡b] are made at the labial and velar places of articulation, and they are therefore called (voiceless and voiced) *labial velar stops*.

An example of a double articulation involving a consonant-sound in English is the initial segment in *well* [wɛl], in which the lips are in a rounded position, and the back of the tongue is raised to a position close to the soft palate, but in neither case close enough to create local friction. The result is a *labial velar resonant*.

When one stricture is of greater degree than the other during the shared medial phase, the narrower stricture is said to be the *primary articulation*, and the more open stricture is called the *secondary articulation*. The auditory effect of secondary articulations is usually to add a modifying "coloring" to the perceptual quality of the primary articulation. Examples include *labialization*, which adds lip-rounding to a segment; *palatalization*, in which the front of the tongue is raised towards the hard palate; *velarization*, in which the back of the tongue is raised towards the soft palate; *pharyngealization*, in which the root of the tongue is retracted towards the back wall of the pharynx; and *nasalization*, in which the soft palate is lowered, allowing air to flow through the nasal cavity and add nasal resonance to the oral resonance of the rest of the vocal tract.

The auditory effect of a neutral, single stricture without secondary articulation is sometimes referred to as "plain." The quality associated with palatalization is sometimes said impressionistically to be "clear," and that with velarization and pharyngealization "dark." In most accents of English, there is a structural allophonic difference between the pronunciations of /l/ in syllable-initial position and in syllable final position, in that both show secondary articulations, with the /l/ of *leaf* [lʲiːf], for instance, being a ("clear") *palatalized voiced alveolar lateral resonant* and that of *feel* [fiːɫ] a ("dark") *velarized voiced alveolar lateral resonant*. Another English example of secondary articulation is the *labialized palatoalveolar fricative* initial in *she* [ʃʷiː], where the primary articulation is the fricative stricture mid-way between the alveolar and palatal places of articulation (hence "palatoalveolar," symbolized by [ʃ]), and the secondary articulation is one of rounding of the lips, symbolized by the superscript diacritic [ʷ]). The use of secondary articulation is discussed further in the section on inter-segmental co-ordination below. Figure 7.4 summarizes the labels for double and secondary articulation.

The *topographical aspects* deal with the shape of the tongue as the active articulator both longitudinally and transversely. Laver (1994a: 141–2) discusses

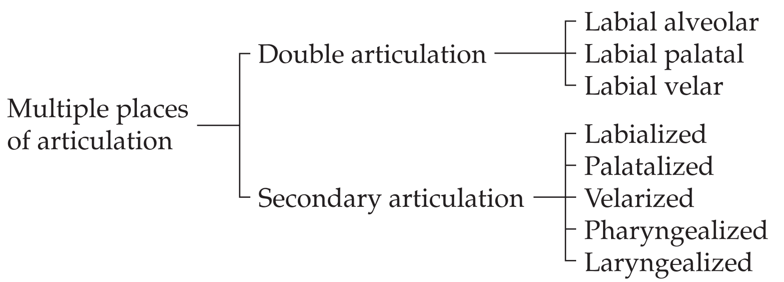


Figure 7.4 Labels for double and secondary articulations

longitudinal processes such as extending or withdrawing the tip of the tongue, and advancing or retracting the root of the tongue. But the most frequently found topographical aspect involving the long axis of the tongue is *retroflexion*, in which the tongue tip is curled up and backwards, sometimes to the extent of presenting the underside of the tip to the roof of the mouth. Margany, a language of South Queensland, shows a phonemic contrast between a (neutral) voiced alveolar stop and a *voiced postalveolar retroflex stop*, in "to cry" [badi] and "maybe" [baɖi] respectively (Breen 1981).

The major transverse aspect distinguishes a flat blade of the tongue from one in which the blade is *grooved*. In English, [s] is produced by most speakers with the air flowing through a very narrow channel in the tongue along the surface of the blade just opposite the alveolar ridge. *Flat alveolar fricatives* occur in Icelandic. In this case it is not clear which category should be treated as neutral, in that grooved alveolar fricatives are far more common than flat alveolar fricatives. It is possible that the higher "pitch" of the fricative noise made through a narrow groove is more audible, and hence ecologically more robust, than the lower pitch of a fricative made through a broader, flat gap. But the articulatory adjustment for creating a central groove is physiologically more complex than for a flat gap, and it may be that the concept of a neutral articulation as the more natural and more widespread sound breaks down at this point.

The *transitional aspects* handle the question of whether the active articulator is static during the medial phase of the articulation, or in dynamic movement. In performing a neutral stop articulation such as the voiced alveolar stop [d], the blade of the tongue rises at moderate pace up towards the alveolar ridge as the passive articulator, makes contact for an appreciable duration, then descends. A (non-neutral) *voiced alveolar tapped stop* is like the neutral version, but moves much faster into contact, makes a very brief closure with the alveolar ridge, and moves away fast. An example is found in many American English accents, as the pronunciation for "t" between two vowels, in a word like *city* [sɪɾɪ].

A tapped stop is sometimes likened to one tap of a trilled stop, another non-neutral example. A trilled stop is one where the active articulator, such as the tip of the tongue, is positioned close to the passive articulator and the airflow through the narrow gap (analogous to the aerodynamic situation in voiced vibration of the vocal folds) brings it repeatedly into full contact. The symbol for a *voiced oral alveolar trilled stop* is [r], and for one made at the uvular location is [ʀ]. A language that contrasts voiced alveolar tapped and trilled stops is Kurdish, as in the pair of words "wound (injury)" [bɾin] versus "cutting" [brin] (A. Ferhardi, personal communication). A contrast between an alveolar tapped stop and a uvular trilled stop is found in European Portuguese, in "dear" [karu] versus "car" [kaʀu] (Parkinson 1988: 138).

When a stop is *flapped*, it strikes the passive articulator in passing. A (non-neutral) *voiced oral alveolar retroflex flapped stop* [ɽ] starts with the tongue-tip curled upwards, and then in uncurling the tip strikes the alveolar ridge very briefly, making a sliding contact that is quickly broken. Westermann and Ward (1933: 76) cite the Sudanese language Gbaya as contrasting a trilled stop with a flapped stop, in "beans" [ere] versus "hen" [eɽe].

Transitional aspects of articulation affect resonants as well. A *monophthong* is phonetically a (neutral) resonant segment with a relatively steady-state articulatory position being maintained throughout its medial phase. A *diphthong* is a (non-neutral) resonant which changes its articulatory position from one position of open approximation towards another during the medial phase. A *triphthong* is a (non-neutral) resonant which changes articulatory position during the medial phase from one position of open approximation towards another and then another. English is unusual amongst the languages of the world in that resonants acting as vowels can show all three transitional aspects of articulation. In some accents of British English, the vowel in a syllable may be represented by either a monophthong (as in *awe* [ɔ]), a diphthong (as in *eye* [aɪ]) or a triphthong (as in *ire* [aɪə]).

10.4 Inter-segmental coordination

Segmental description in this chapter so far has been limited to events within the boundaries of a single segment. Some of the most phonetically interesting events occur in the overlapping phase between two adjacent segments, where the first segment's offset phase is co-terminous with the next segment's onset phase (Laver 1994a: 339–90). When a segment is next to utterance-marginal silence, the onset and offset phases involve transitions from and to the articulatory rest position. Also relevant is the effect of the characteristics of one segment's medial phase spreading, anticipatorily or perseveratively, into part or all of the medial phase of the adjacent segment. Significant phonetic events involving coordination of adjacent segments include the phenomena of devoicing, aspiration, release, affrication and co-articulation.

10.4.1 Aspiration

When a segment that in most contexts is fully voiced throughout its medial phase occurs next to a silent pause, say in utterance-initial position, the transition from silence may have the effect of delaying the beginning of voicing for that segment. In most accents of English, voicing for an utterance-initial voiced stop or fricative will start after beginning of the medial phase. Using [#] to indicate silence, and a subscript [̥] to mean "delay in the onset of voicing" the utterance-initial word "zeal" would be transcribed [# ̥ziːl]. Because [z] in most contexts is normally pronounced with full voicing through its medial phase, this delay is usually referred to as *initial devoicing* (though the phonological orientation of such a practice should be noted). Correspondingly, when such a segment is next to utterance-final silence, there may be an early offset of voicing, and the sound is said to be *finally devoiced*. Such utterance-final devoicing would be transcribed, in the English word "lees," as [liːz ̥#].

In both the initial and final cases, the devoicing is partial, in that not all of the medial phase is deprived of vibration of the vocal folds. When there is no voicing at all in the medial phase, the question is prompted of what differentiates a *fully devoiced segment* such as [z̥] from its voiceless counterpart [s]. Some phoneticians and phonologists make appeal to issues of differential muscular tension in the vocal apparatus, and set up the categories of *lax* and *tense* to describe hypothesized factors that continue to differentiate such devoiced and voiceless segments. It is probably more satisfactory, at a phonetic level of description, to accept the non-differentiability of fully devoiced and voiceless segments. Figures 7.5a and 7.5b characterize the timing relationships between the laryngeal and supralaryngeal events in the devoicing process, and relate them to the next category of coordination to be discussed, aspiration.

When a voiceless segment such as an oral stop is initial before a resonant in a stressed syllable in most accents of English, there is an audible delay in the onset of voicing after the end of the stricture of the medial phase, in the overlap phase between the stop and the resonant. This phenomenon is called *aspiration*. An instance is the English word "peat" [pʰiːt], where the aspiration is transcribed as a small superscript "h." The audible quality of the [ʰ] anticipates that of the oncoming resonant, for which the vocal tract is already assuming the relevant articulatory position. Aspiration is reasonably rare among the languages of the world. French, for example, does not aspirate syllable-initial stops in such circumstances, in words such as "paté" [pate]. Aspiration acts as an allophonic process in English, applying to all voiceless stops /p, t, k/, but is exploited phonemically in a number of languages, including Chengtu Szechuanese, in words such as "to cover" [kai] versus "to irrigate" [kʰai] (Fengtong 1989: 64).

Aspiration is perhaps best defined as "a delay in the onset of normal voicing," since a category of *voiced aspiration* is found in a number of languages of the Indian subcontinent and in central and southern Africa, as a relationship between voiced stops and following resonants. In this case, the phonatory quality of the transition from the stop to the following resonant is one of whispery voice, usually becoming normal (i.e. without audible glottal friction) before the end of the resonant. Examples of voiced aspiration, involving

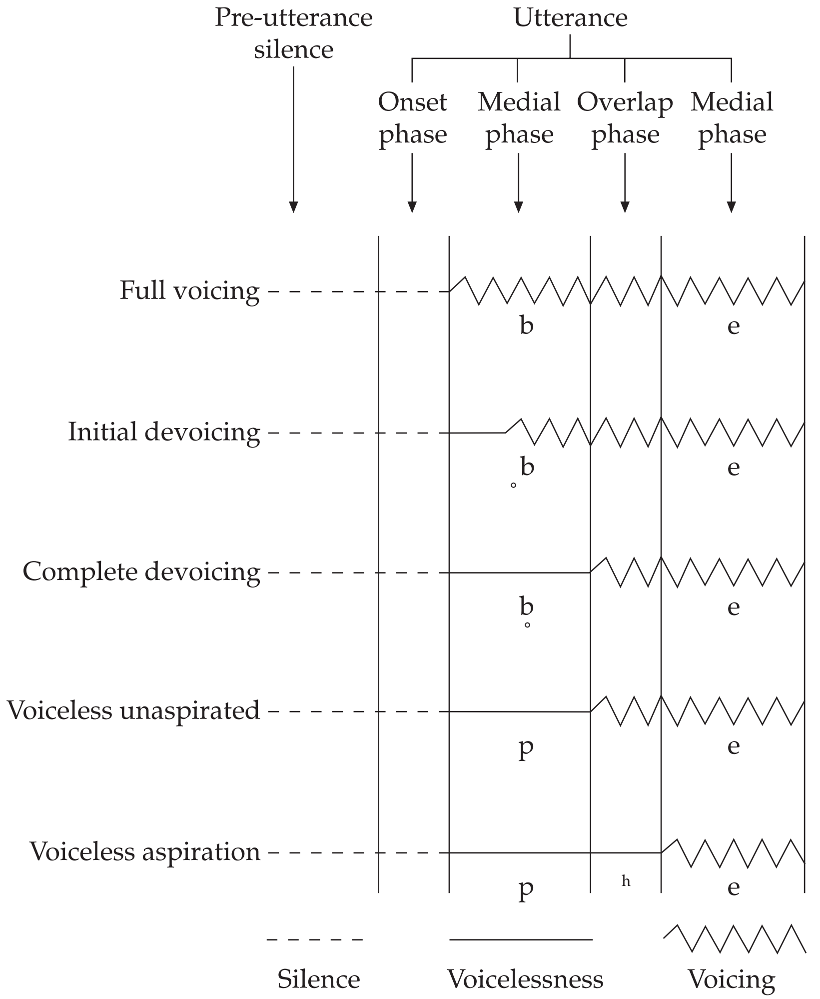


Figure 7.5a Timing relationships between laryngeal and supralaryngeal events in initial devoicing and aspiration
Source: After Laver 1994a: 340

both oral and nasal stops, are the Sindhi words "to speak ill of others" [ɡɪla] versus "wet" [ɡɦɪla], and "in" [mẽ] versus "a buffalo" [mɦẽ] (Nihalani 1975: 91).

Parallel to aspiration as a late onset of voicing in syllable-initial contexts is early offset of voicing in a resonant before a voiceless segment in syllable-final position, which is called *pre-aspiration*. This is a characteristic of many of the circumpolar languages. It can be voiceless, as in an example from Icelandic in "thank" [θahka] (Ewen 1982), or voiced, as in Hebridean Gaelic (of Lewis) in "bag" [p^{h}ɔɦk] (Shuken 1984: 127).

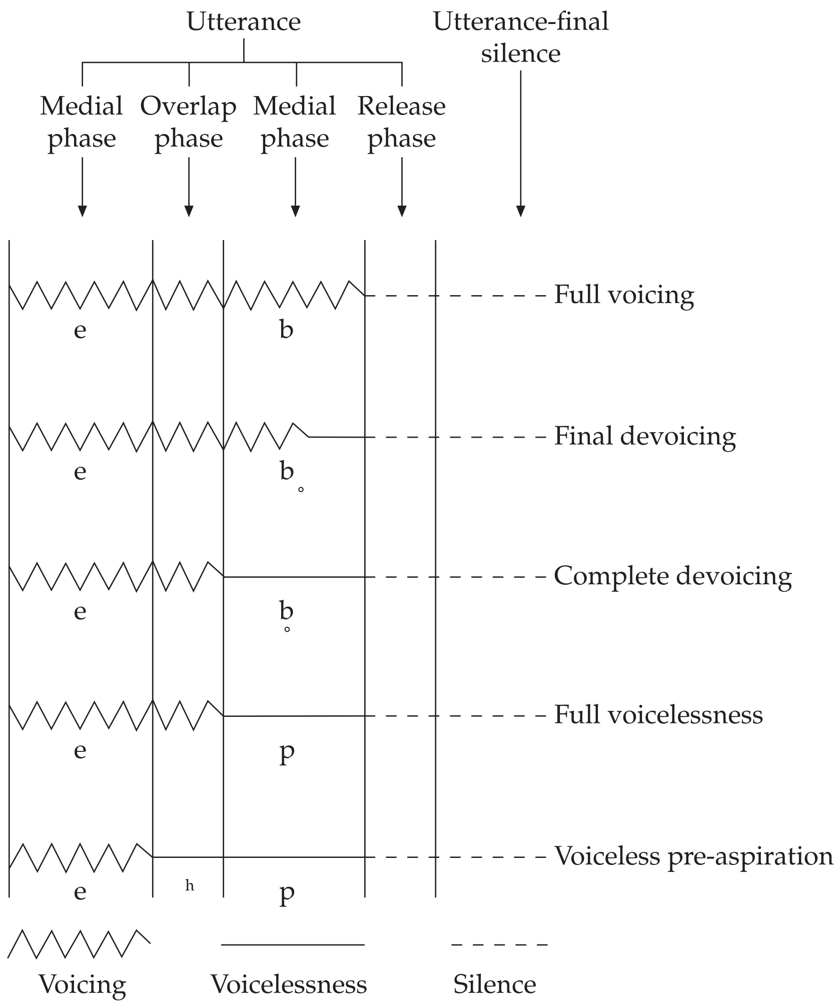


Figure 7.5b Timing relationships between laryngeal and supralaryngeal events in final devoicing and pre-aspiration
Source: After Laver 1994a: 341

10.4.2 Stop release

The offset phase of any oral stop may release the compressed air built up during the medial phase in a variety of ways. Alternatively, the stop may be incomplete, and lack a final release. In this latter situation, the oral closure is sometimes reinforced by a simultaneous glottal stop, as a double articulation. Both modes are found in English, as optional variants. A syllable-final *unreleased stop* can be transcribed for an accent of British English as *top* [tɒp˺] (and the glottally reinforced version as [tɒp͡ʔ˺]), with the released version as [tɒp^{h}].

When the offset phase of a syllable-final stop is released, the channel for the release may be either oral or nasal, central or lateral. In English, the word *button* can be pronounced with *oral release* as [bʌtən], or *nasal release* as [bʌtn̩]. In the latter case, the diacritic below the [n̩] symbol indicates that the nasal segment is "syllabic," acting as the nucleus of the second syllable. The difference between central release and lateral release can be seen in the English word *little*, pronounced with *central release* as [lɪtɔl], or with *lateral release* as [lɪtl̩].

10.4.3 Affrication

Affrication is also a characteristic of the offset, release phase of stops. The stop closure is released more slowly than in a non-affricated stop, so that a brief moment of audible friction is heard as the stricture passes momentarily from complete closure through close approximation. In English, the sounds at the beginning of the words *cheap* [t͡ʃiːp] and *Jeep* [d͡ʒiːp] are *affricated stops* (or "affricates"). The special relationship between the stop and the fricative element is symbolized by the use of the linker diacritic. The fricative element of an affricate is by definition homorganic with the stricture of the associated stop element, and affricates can be made at any place of articulation where stops can be formed. The fricative element may also be lateral or central, as in the first and second affricates respectively in the Nahuatl phrase "sit down please!" [ʃimot͡ɬalit͡sino] (Suárez 1983: 32).

10.4.4 Co-articulation and assimilation

It is not surprising, given the rate that segments follow each other in the stream of speech, that one segment may influence the articulatory characteristics of segments yet to be spoken, or be influenced by those that precede it. When such an influence crosses a word-boundary, it is said to result in assimilation; when it is restricted to word-internal action, it is said to show co-articulation.

An example of *perseverative assimilation* in an accent of British English is the continuing influence of voicelessness across the word boundary in *what's this?* pronounced as [wɒts ðɪs], where the underlying /z/ of *is* loses its voicing under the influence of the preceding [t]. An instance of *anticipatory assimilation* is *that boy* pronounced as [ðap bɒɪ], where the place of articulation of the underlying /t/ is made identical to that of the following [b].

Anticipating the strictural requirements of oncoming segments in the stream of speech often results in secondary articulations. Labialized consonant-sounds are found in English before vowels which have rounded lip-positions, in words such as [pʷuːl] *pool*, [mʷɔː] *maw*, and [fʷʊl] *full*. A further example of a secondary articulation that characteristically anticipates future segmental requirements is nasalization. In *anticipatory nasalization*, when a vowel-sound precedes a nasal consonant, the soft palate opens during the medial phase of the resonant, anticipating the oncoming requirement for nasality. A language-differentiating facet of this process is that in these circumstances, the soft palate opens later in French than in English, presumably because of the need to protect the perceptual distinctiveness of French nasal vowel-sounds in phonemic opposition to their oral counterparts.

Co-articulatory *anticipations of place of articulation* also occur. In English, the stricture of velar stops such as [k] before a front resonant as in *keep* [k^{h}iːp] is made further towards the front of the mouth than [k] before a back resonant as in *calm* [k^{h}ɑːm].

10.5 Temporal organization of speech

The discussion so far has concentrated on matters to do with the phonetic quality of speech sounds. The remaining variable is *duration*. Segments have certain inherent durational constraints which have physiological or perceptual explanations (Laver 1994a: 431–6). This section will concentrate, however, on the contrastive and contextual control of duration for phonological purposes. The terms "length," "long," and "short" will be reserved for use at a contrastive level, and greater or less "duration" for use at the phonetic level of description.

Phonemic distinctions of length in both vowels and consonants have been observed, with vowel-length distinctions predominating. Vowel-length distinctions abound in accents of English, though usually with associated differences of segment-quality. An example of a language using *contrastive vowel-length* (with length signaled by the diacritic [ː]) is Rarotongan Maori, in word-pairs such as "taro bed" [pɑʔi] and "ship" [pɑːʔɪ] (Buse 1966: 52).

Phonemic distinctions of consonant-length are much rarer, but are found occasionally, as in the Eskimo-Aleut languages of the Canadian Arctic. Inuktitut (Inuit) distinguishes short and long consonants in phrases such as "they arrive together" [tikiqɑtɑujut^{h}] versus "they arrive frequently" [tikiqɑtːɑqtut^{h}] (Esling 1991).

An instance of a language (unusually) contrasting both vowels and consonants is Finnish. An example of such a word-pair is "a crease" [rʏpːʏ] versus "a drink" [rʏːpːʏ] (T. Lauttamus, personal communication).

Allophonic adjustments of duration both to structural position and to phonetic environment are very common. In English, the duration of vowel-sounds is greatest in open syllables such as *bee*, and less in closed syllables such as *beat*. In syllables of comparable structure, the duration of a vowel-sound is greater before a voiced consonant-segment such as [d] in *bead*, and less before a voiceless consonant-segment such as [t] in *beat*.

11 Conclusion

The detailed resources of general phonetic theory that have only been able to be sketched in here are probably adequate for the task of describing the segmental make-up of almost all languages known today, though of course some problems of detail remain to be resolved. In the terms introduced at the beginning of this chapter, general phonetic theory is basically fit for the linguistic phonetic purpose of "describing the phonetic basis for differentiating all contrastive and contextual patterns in speech which signal the identities of linguistic units in any given language." Not so evident is whether it is yet fit for the three other purposes identified as relevant to a broader interest in linguistic phonetics – the description of phonetic regularities in the speech-styles of sociolinguistic communities, of the characteristic phonetic events that distinguish the speech patterns of individual members of those communities, and of the ways in which languages sound different from each other. Such questions raise large issues for the future about the nature and motivation of work in phonetics, and about the desirable and useful limits of resolution of the descriptive apparatus used.

APPENDIX 7

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	ʼ	as in:
ǀ	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
!	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
ǁ	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

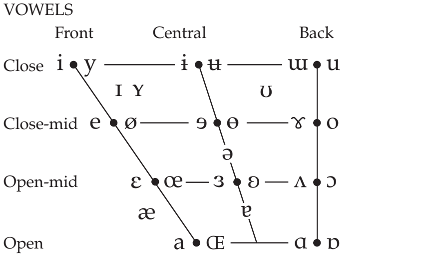
SUPRASEGMENTALS

- ˈ Primary stress
- ˌ Secondary stress ˌfoʊnəˈtɪʃən
- ː Long eː
- ˑ Half-long eˑ
- ˘ Extra-short ĕ
- . Syllable break ɹi.ækt
- | Minor (foot) group
- ‖ Major (intonation) group
- ‿ Linking (absence of a break)

TONES & WORD ACCENTS

LEVEL		CONTOUR	
e̋ or ˥	Extra high	ě or ˩˥	Rising
é ˦	High	ê ˥˩	Falling
ē ˧	Mid	e᷄ ˦˥	High rising
è ˨	Low	e᷅ ˩˨	Low rising
ȅ ˩	Extra low	e᷈ ˧˦˧	Rising-falling etc.
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

- ʍ Voiceless labial-velar fricative
- w Voiced labial-velar approximant
- ɥ Voiced labial-palatal approximant
- ʜ Voiceless epiglottal fricative
- ʢ Voiced epiglottal fricative
- ʡ Epiglottal plosive
- ɕ ʑ Alveolo-palatal fricatives
- ɺ Alveolar lateral flap
- ɧ Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. k͡p t͡s

DIACRITICS

Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

̥	Voiceless	n̥ d̥	̤	Breathy voiced	b̤ a̤	̪	Dental	t̪ d̪
̬	Voiced	s̬ t̬	̰	Creaky voiced	b̰ a̰	̺	Apical	t̺ d̺
ʰ	Aspirated	tʰ dʰ	̼	Linguolabial	t̼ d̼	̻	Laminal	t̻ d̻
̹	More rounded	ɔ̹	ʷ	Labialized	tʷ dʷ	̃	Nasalized	ẽ
̜	Less rounded	ɔ̜	ʲ	Palatalized	tʲ dʲ	ⁿ	Nasal release	dⁿ
̟	Advanced	u̟	ˠ	Velarized	tˠ dˠ	ˡ	Lateral release	dˡ
̠	Retracted	i̠	ˤ	Pharyngealized	tˤ dˤ	̚	No audible release	d̚
̈	Centralized	ë	̴	Velarized or pharyngealized	ɫ			
̽	Mid-centralized	e̽	̝	Raised	e̝ (ɹ̝ = voiced alveolar fricative)			
̩	Syllabic	n̩	̞	Lowered	e̞ (β̞ = voiced bilabial approximant)			
̯	Non-syllabic	e̯	̘	Advanced Tongue Root	e̘			
˞	Rhoticity	ɚ	̙	Retracted Tongue Root	e̙			