Contents

Preface vi
Acknowledgements vii
List of Terms viii
The Key Concepts 1
Notes 162
References 165
Index 172
Accent

This term has two distinct meanings. The first and more common one refers to how an individual pronounces a language; the second refers to the emphasis a speaker places on an individual word. Technically speaking, the first meaning of accent can be defined as the cumulative auditory effect of those features of pronunciation that identify where an individual is from regionally or socially, or identify the speaker’s occupation or social class. As every single speaker must pronounce a word in a particular manner, which allows his/her hearer to identify either where the speaker is from or what the speaker’s social class is, it is untrue to claim that accentless speech is possible (Esling 1998). The belief that one accent of British English is accentless and more prestigious will be discussed under the entry received pronunciation, as will discussion of social accents.

In the English speaking world there are clear differences between the English spoken between and within the main English speaking counties. Most American and most British accents differ, for instance, through the distribution of /a/ within a syllable. In most American accents the consonant /a/ is pronounced before and after vowels within syllables. In most British accents the consonant /a/ can only be pronounced before a vowel within a syllable. Thus the typical American pronunciation of car is /kɑ:/ while the typical British pronunciation is /kɑː/.

Because no two individuals speak in an identical manner, it is difficult to state how many regional accents of English exist. We can and do classify accents by country. Thus we speak of American, British, Canadian, Irish accents and so on. What is clear is that regional accents are more prevalent in the countries that have had the longest length of settlement by English speakers. In other words, there are fewer regional accents in North America and Australia, and more regional accents in Britain and Ireland. Within the United States the region with the largest number of regional accents is New England, which is also the region of earliest settlement (Kurath 1949; Nagy and Roberts 2008).

Different scholars have proposed differing typologies of accent classification, which has resulted in proposals classifying American regional accents into between three and ten types. For instance, Cassidy (1982) classifies American English as Eastern, referring to the non-rhotic accents of Boston, New England and New York City; Southern referring to the non-rhotic accents of the South, an area largely coterminous with the physical territory of the Confederate States; and General American the rhotic accent found in the remainder of the United States. Thomas (1958) argues that there are ten accent areas in the United States. He subdivides the Eastern accent into Eastern New England and New York City, the Southern accent into Southern and Southern Mountain; and General American
into Middle Atlantic around Philadelphia, Western Pennsylvania, Central Midland centred around the middle band of states running from Ohio to Utah, Northwest, Southwest centred around California, and North-Central centred around Chicago. More recently Labov, Ash and Boberg (2006: 146) have classified North American accents of English, including Canadian English, into 17 accent regions.

In Australian English, following Mitchell and Delbridge (1965), it is useful to classify three accent categories: Cultivated, General and Broad. The overwhelming majority of the population speaks General Australian English. Cultivated Australian English is phonetically close to Received Pronunciation, with General Australian closer to the English spoken in the South East of England. Broad Australian English is noticeable for having long vowels initially in diphthongs. A noticeable difference between Australian and British accents is the raising of the vowel /æ/ (as in trap) to a vowel with the quality of a British /E/ (as in dress). Thus, an Australian speaker’s vocalization of the word pan may sound like a British speaker’s vocalization of the word pen.

Within Britain and Ireland, far more accents are found within a more confined geographical area. In England alone Trudgill (2000: 66–8) identifies 16 accent regions. It must be understood that speakers are sensitized to hear significant local accent differences within the wider speech regions identified by linguists. For instance, Trudgill classifies speakers from Newcastle, Durham, Sunderland and Middlesbrough as speaking in a North East accent. Yet speakers from Newcastle have little difficulty in distinguishing their accents from those of Sunderland speakers.

The major historic split in English accents is between North and South, with Northern accents generally representing more historically conservative speech varieties. The category of historic Northern accents can be divided into modern day Northern accents and modern day Central accents, while modern Southern accents can be divided into those found in the South East and in the South West. Some key features that distinguish Southern accents from Northern and Central accents are that, in Northern and Central accents, words containing the strut vowel /ʌ/ are homophonous with words containing the foot vowel /ə/. Thus, for Northern and Central speakers the words put [pʊt] and putt [pʌt] are identical. This is not the case for Southern speakers. Northern and Central speakers pronounce the following set of words with a short front /æ/ vowel while Southern speakers pronounce them with a long back /ə:/ vowel: path, glass, raft, grasp, past, after, basket, laugh, dance, grant, branch, graph, etc. (for further information see Wells 1982a: 135). There is, however, a further complication in describing accents in terms of /ə:/ and /æ/, namely that speakers of South Western forms only use a single <a> vowel [a] which is intermediate between the fronted /æ/ vowel and the back /ə:/ vowel.

A key difference between Northern and Central accents is that Northern accents retain some long historic monophthongs whereas Central accents have undergone diphthongization. This results in the following potential pronunciation differences between Northern and Central speakers with Northern realizations first: made [me:d], [meid] boat [bot], [bəut]. South Eastern and South Western are chiefly divided by the fact that South Western accents are unique – excluding a shrinking relic area in the Southern rural areas of Lancashire – among English accents in
retaining /ʌ/ after a vowel within the same syllable. Thus, while a typical South Eastern speaker would pronounce *arm* as [ɑːm], a typical South Western speaker would pronounce it as [əm].

Scottish, Irish and Welsh accents are widely heard distinctive and recognizable accents found within the individual nations and more generally across the UK and Ireland. As with Northern English accents, they are essentially conservative. All three have to a greater or lesser extent been influenced by Celtic languages. Scottish accents have, in addition, been influenced by *Scots*, a distinct English dialect that was introduced into Scotland from Northumbria in the seventh century and was subsequently itself influenced by contact with Scottish Gaelic. Until the political union between Scotland and England, *Scots* was the de facto national language of Scotland. Some of the key defining features of present day Scottish accents are that it is rhotic; there is no distinction between the *foot* /ʊ/ and *goose* /uː/ vowels that are pronounced /uː/; vowel duration is dependent on the phonetic context with vowels being generally longer before /ʌ/ or a morpheme or word boundary (Stuart-Smith 2008), and the presence of the consonants /w/ and /j/ – a voiceless labial-velar approximant, which in most other forms of English has merged with /w/ resulting in the loss of a distinction between the following pairs of words: *which*/*witch, whine/wine* and /x/, a velar fricative that has been retained in words such as *loch* [lɒx]. In most English accents *loch* would be homophonous with *lock* and pronounced /lɒk/. The Scottish accent also differs from most English accents in that it does not favour H dropping.

There are two main accents found in Ireland, with the accent of the North of the island Ulster English being in a sense intermediate between a Scottish and an Irish accent. However, Ulster English accents have some unique features such as the *palatization* of /k/ and /g/ before low vowels. Cat is pronounced as [kjat] (see Hickey 2008 for further details). Unlike Scottish and Southern Irish accents, Ulster English accents do not have the *phoneme* /ʍ/. All Irish accents, like Scottish accents, do not tend to favour H dropping. The chief distinguishing features of Southern Irish English accents are that it is rhotic; it retains the /ʍ/ phoneme; and, uniquely, the *th* sounds in words such as *thin* and *this* are pronounced as *dental stops* – for example, [tɪn] not /θɪn/ and [dɪs] not /ðɪs/ (see manner of *articulation* and place of *articulation*). Unlike most accents of English, but not Welsh accents, /l/ is traditionally pronounced in the same manner before and after vowels. However, Wells (1982b: 431) and Hickey (2008: 92) report that post-vocalic /l/ is becoming dark (see *velarization* and *dark L*).

Welsh-accented English, unlike the other Celtic accents, is typically non-rhotic though there are three distinct areas of rhoticity within Wales. The first borders the rhotic English South West, the second is the site of the first historical English speaking settlement and the third is the traditional Welsh-speaking areas (Wells 1982b; Penhallurick 2008). In some Welsh accents there is a possibility of contrasts that are not found in other varieties of English, notably *blue/blew, [bluː], [bluː] pain/pane, [pɛn] [pɛn] no/know [nɔː], [nou]. Somewhat surprisingly, a widely commented-upon feature of Welsh accented English – its lilting or sing-song intonation – has not yet been systematically investigated. In a study aimed at producing a phonetic description of Welsh-accented English in the South Wales valleys, Walters (2003) points out that the most notable feature that contributes to
the melody of the Rhondda valley is intonation. Unlike other varieties of English the most common pitch movement prior to the tonic syllable was downwards, while after the tonic syllable the most common initial tone movement was rising.

**Accent 2**

An accented syllable refers to a syllable that is realized by auditory prominence within a word. This auditory prominence enables the word to stand out in a stream of speech. Accenting results from a combination of pitch, loudness and duration. The accent within a word is normally placed on the primary stressed syllable within the word. But if the speaker wishes to highlight a contrast he/she may move the accent on to another syllable, for instance Bolinger (1962: 83) reports a speaker saying: *This whiskey wasn’t EXported from Ireland it was DEported*. The speaker, in order to signal the contrast, has moved the accent from the primary stressed syllable por. The communicative value produced by accented syllables is discussed under intonation.

**Acoustic phonetics**

Acoustic phonetics is the subfield of phonetics, which studies the physical properties of speech. It complements articulatory and auditory phonetics. Advanced study of acoustic phonetics requires some mathematical knowledge and familiarization with the relevant equipment such as spectrographs or relevant software, such as Praat. Acoustic analyses reveal the physical facts of a sound wave, which can be used to corroborate articulatory analyses and provide a firm phonetic underpinning for phonological analyses. However, acoustic analyses of intonation contours may conflict with our perception of intonation contours; what we perceive as a rising tone movement may be measured instrumentally as neither an increase nor decrease in fundamental frequency. In cases where acoustic analyses conflict with human perception it is by no means clear which analysis a phonetician should prefer, nor how a phonetician should reconcile the difference (see Wichmann 2000: 2).

Acoustic phonetics describes the physical effect of speech on the air molecules surrounding the speaker. When people speak they disturb the air molecules near their mouths, which in turn bump into other molecules and displace them. This results in a chain reaction, which continues until the energy imparted by the vocal organs dissipates and the sound dies out some distance from the speaker. When an air molecule is displaced it rebounds back to its starting point and continues to oscillate to and fro until it runs out of energy, at which point it eventually stands still again. The sounds we hear are related to the characteristic vibrations of molecules around their places of rest. It is these cycles that acoustic phoneticians plot in order to measure the frequency of a sound. There are two ways in which the oscillation of an air molecule can vary. The first is if the molecule is given a harder initial bump: it will travel further from its place of rest. We perceive the increased amplitude as an increase in loudness. The second way to vary the oscillation is to increase or decrease the speed in which the cycle can be completed. A faster cycle with amplitude kept constant is perceived as a higher pitch, and a slower cycle as a lower pitch.
Sounds whose periods remain almost the same for cycle after cycle, such as a note on a piano, are known as periodic sounds. Other sounds are known as aperiodic sounds. Only periodic sounds give rise to a clear sensation of pitch. The sound waves of some speech sounds, chiefly **vowels**, can be decomposed into a combination of more than two periodic sounds, known as harmonic components. The waveform of an aperiodic sound, such as a consonant, does not repeat itself and cannot consequently be broken down into harmonic components. However, aperiodic sounds can be identified instrumentally in terms of the overall shape of their spectra. This is because the amplitude of any individual aperiodic sound is greater in some frequency regions than others. It is these differences in the overall amplitude profile of the sound over the frequency range that enable us to distinguish one aperiodic sound from another (see also **formant**). Phoneticians can identify all individual speech sounds either through the combination of fundamental frequency and harmonic structure or the overall spectra shape, and can produce spectrographic illustrations of each individual speech sound. For accessible accounts of acoustic phonetics see Ball and Rahilly (1999: ch. 9); Clark, Yallop and Fletcher (2006: ch. 7); Ladefoged (2001: ch. 8); Harrington (2010); and Lodge (2009: ch. 9).

**Acute**

This term has three distinct meanings all of which contrast with **grave**. The first is slightly old-fashioned and is used in descriptions of **tone** languages to refer to a rising lexical tone. The second meaning is used in some accounts of **intonation** to describe a rising tone. The third meaning refers to one of the features of a sound in **distinctive feature** theory (see Jakobson and Halle 1956), where it defines sounds involving the combination of a medial articulation in the mouth and a concentration of acoustic energy in the higher frequencies. Sounds that contain the feature acute are front **vowels**, and **dental**, **alveolar** and **palatal consonants**.

**Advanced tongue root**

In most languages differences in **vowel** quality can be described in terms of variations in height, backness and lip-rounding. However, in some African languages, such as Akan and Igbo, there are sets of vowels that cannot be distinguished solely by variations in these three dimensions. In one set of vowels, the root of the tongue is drawn forward and the larynx is lowered. These vowels are described as advanced tongue root (ATR) vowels. In the other set, the root of the tongue is not drawn forward and the larynx is not lowered. This set of vowels is described as retracted tongue root (RTR) vowels. However, Ladefoged (2001: 211) notes that not all speakers of Akan make the difference between ATR vowels and RTR vowels by bunching-up their tongues. Instead, what seems to signal the distinction is either tongue root movement or larynx lowering, which results in an enlargement of the middle and lower pharynx. In English, while no sets of vowels are distinguished solely by this articulatory gesture, the vowels in **FLEECE** and **GOOSE** have a more advanced tongue root than do the vowels in **KIT** and **FOOT**. See **tense** and **lax**.
Affricate

Affricate is a term used to refer to the classification of a **consonant** on the basis of its manner of **articulation**. An affricate results from the sequential articulation of a **stop** that is released slowly into a **fricative** within the same syllable in the same place in the mouth. Affricates, like all consonants, may be **voiced** or **voiceless**. Voiceless affricates may be **aspirated** or **unaspirated**. Phonetically, affricates differ from a sequence of a stop followed by a fricative in that the duration of the fricative component is shorter than the duration of an independent fricative. There are two affricates found in English, namely the first sounds in *chip* /tʃ/ and in *judge* /dʒ/. The **voiceless palato-alveolar** affricate /tʃ/ is a very common speech sound. It is present in 45 per cent of the world’s languages. Other well-known affricates are the voiceless labiodental affricate /pf/ found in German and the **alveolar** affricate /ts/ found in numerous languages such as Chinese and German (see Ladefoged and Maddieson 1996). In theory, an affricate can be articulated anywhere in the mouth. The International Phonetic Association (IPA) provides symbols for voiced and voiceless **velar** and **uvular** affricates but, in practice, languages tend to favour affricates produced between the **alveolar** ridge and the hard palate.

In many languages, including some accents of English, stops are often affricated. This means that the speaker releases the stop very slowly, which causes a short but audible phase of friction. In effect, by not lowering the tongue cleanly and swiftly, the speaker releases the stop into a fricative. Affrication is a noted feature of Liverpool speech, where some speakers often realize the following stop phonomes: /p/, /t/, /d/ and /k/ as [pʰ], [tʰ], [dʰ] and [kʰ].

Phonologically, affricates realize one **phoneme**. Sounds that may be pronounced phonetically as affricates may be classified on **phonological** grounds as either affricated **allophones** of stops, or as a sequence of two consonant phonemes. In many accents of English the initial consonant cluster in the words *train* and *drain* are pronounced as affricates with the onset of the /ɜ/ producing audible friction. Despite this phonetic reality, the initial consonant cluster is not classified as an affricate. Three justifications may be advanced for not analysing *[tɹ]* and *[dɹ]* in strict accordance with the phonetic facts. The first is that native English speakers feel that words such as *train* and *drain* begin with two consonant sounds, whereas words such as *church* and *judge* are felt to begin with only one consonant. The second reason is that the affricate /tʃ/ can be found in all phonetic environments, for example, word initially in words such as *chip*, word medially in words such as *richer* and word finally in words such as *match*. The sequence *[tɹ]* can only be found word initially in words such as *train* and word medially in words such as *metrical*, it is never found word finally. However, this second argument is not as convincing as it initially seems. Some English consonant phonemes are restricted in their distribution, for example, /tɹ/ is only allowed medially and finally but not initially, and /h/ and /w/ are only allowed word initially and medially.

The final reason is that the initial sound in the consonant sequences *[tɹ]* and *[dɹ]*, and in some rare cases the second consonant may be substituted by another consonant, and this substitution leads to the formation of a new lexical unit (see **phoneme**), for example:
/tʰeɪn/ train: /t/ replaced by /b/ becomes /bəɪn/ brain.
/θæɪn/ train: /θ/ replaced by /w/ becomes /twæɪn/ twain.
/ʤæɪn/ drain: /ʤ/ replaced by /ɡ/ becomes /ɡæɪn/ grain.
/dæɪn/ drain: /d/ replaced by /w/ becomes /dweɪn/ Dwayne.

In affricates such as /tʃ/ it is not possible to substitute either the stop or the fricative component alone. The entire phonological unit contrasts with other English consonant phonemes; for example, if the initial affricate /tʃ/ in chip is replaced by the following consonants /p/, /t/, /d/, /n/, /l/, /s/, /ʃ/, /k/ and /h/, the following set of words results: /pɪp/, /tɪp/, /dɪp/, /nɪp/, /lɪp/, /sɪp/, /ʃɪp/, /kɪp/, /hɪp/.

Airstream mechanism

Speakers need a source of energy to power their speech. The airstream mechanism is the physiological process, which provides the energy used in speech production. The air used to power speech may be produced by an outward flow of air known as an egressive airstream or, in less usual circumstances, by an inward flow of air known as an ingressive airstream. Most speech is powered by an outwards flow of air initiated by the lungs. This is known as a pulmonic egressive airstream. However, as readers who have talked while simultaneously engaging in strenuous physical activity will realize, it is perfectly possible to articulate a speech sound or a sequence of speech sounds when inhaling. Technically this airstream mechanism is known as a pulmonic ingressive airstream.

The movement of air used to power a speech sound may be initiated other than by the lungs. An airstream mechanism can be initiated by the movement of the glottis. As with a pulmonic airstream the flow of air known as glottalic may be egressive or ingressive. A further power source for speech sounds found in language is an airstream known as velaric, which is initiated by the pressing of the back of the tongue against the velum. The airstream initiated by the velum is ingressive and produces a sound similar to the ‘tut tut’ paralinguistic sound produced by English speakers to signal displeasure.

In principle, speech can be powered by any airstream mechanism initiated at any point in the vocal tract. Patients who have undergone a laryngectomy have been taught to power their speech by swallowing air into their stomachs and then pushing the air out from their stomachs as a substitute for an egressive pulmonic airstream. Chapter 6 of Laver (1994) and Ladefoged and Maddieson (1996: 77–89) provide a comprehensive and reader-friendly description of airstream mechanisms.

Allophone

An allophone is a phonological term for an audibly distinct variant of a phoneme within a language which does not affect the phoneme’s functional unity. Allophones within a language system are found in complementary distribution. They do not exist in the same phonetic environment. For instance, in most dialects of English there are two variants of /l/ clear <L> [l] and dark <L> [l]. Clear <L> is normally produced by contact between the tip of the tongue and
the upper teeth, which allows the air to escape over the sides of the tongue. Simultaneously, the front of the tongue is raised in the direction of the hard palate. In contrast for dark or velarized <L> the front of the tongue is somewhat depressed, while the back of the tongue is raised towards the soft palate. Dark <Ls> consequently have an auditorily perceptible back vowel resonance which clear <Ls> lack. Yet despite the clear phonetic dissimilarity between clear and dark <Ls> in English they are considered to be allophones and not separate phonemes. This is because, in English, clear and dark <Ls> are in complementary distribution with clear <Ls> preceding vowels and /j/, while dark <Ls> follow vowels. In other words, substitution of a clear <L> by a dark <L> does not lead to the formation of a new lexical item. For instance, regardless of whether a Southern English speaker pronounces a word as [lɪt], a Southern Irish speaker as [lɪlt] or a Northern English speaker as [lɪlt], all three speakers have articulated the lexical item lilt.

In English the nasal /n/ is normally realized in alveolar position (e.g. /pæn/) but in the word panther – see assimilation – it is realized as a dental nasal (e.g. ['pʰæn]). In English the realization of a nasal as dental before /θ/ is predictable. Both variants of the nasal are in complementary distribution and can be classified as allophones. However, it is worth noting that complementary distribution alone is not sufficient to identify allophones. The phones /h/ and /ŋ/ occur in complementary distribution in English with /h/ present only in syllable onset position and /ŋ/ restricted to syllable coda position. If we relied solely on complementary distribution to identify allophones we would have to argue that the two phones were allophones of the same phoneme. However, this is clearly not a satisfactory solution as English speakers intuitively feel that the two phonemes are not allophonic realizations of a single phoneme. Indeed I cannot even imagine what kind of phoneme they could be allophones of!

In deciding whether a particular phone is a phoneme or an allophone we must also consider the notion of phonetic similarity. Phonetic similarity is a vague concept that differs across languages. For instance, in Irish Gaelic, dental nasals are not in complementary distribution with alveolar nasals. Instead their distributions overlap and the choice of a dental or alveolar nasal realizes lexical meaning. Irish contrasts nasals with six different places of articulation, or, in other words, has six nasal phonemes (Ní Chasaide 1999: 112); English has only three nasal phonemes, one of which /ŋ/ is restricted to coda position. Hence what is phonemic in Irish will be allophonic in English.

To an English speaker it is intuitive that the first, second and third phones in the following words are all realizations of the /p/ phoneme pot [pʰɔt], spot [spɔt] and top [tʰɔp], yet speakers of some other languages do not have the same intuition. Hindi, Korean, Mandarin and Cantonese contrast /p/ and /pʰ/ phonemically (for an enlightening discussion see Silverman 2006: 69–80). The following illustrative examples are from Ladefoged and Maddieson (1996: 58) and Bok Lee (1999: 121).

<table>
<thead>
<tr>
<th>Hindi</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pal/</td>
<td>/pal/ take care of</td>
</tr>
<tr>
<td>/pʰal/</td>
<td>/pʰal/ sucking</td>
</tr>
<tr>
<td>/bal/</td>
<td>/bal/ knife blade</td>
</tr>
<tr>
<td></td>
<td>/bal/ arm</td>
</tr>
<tr>
<td></td>
<td>/bal/ hair</td>
</tr>
<tr>
<td></td>
<td>/bal/ foot</td>
</tr>
</tbody>
</table>

For speakers of Hindi and Korean the presence or absence of aspiration is as phonetically dissimilar as is the presence or absence of voicing. English speakers
by contrast barely recognize the phonetic dissimilarity between an aspirated and unaspirated stop. Conversely for English speakers the initial phones in the words *they* /θeɪ/ and *day* /deɪ/ sound phonetically dissimilar enough for them to be able to distinguish between the phones and to contrast them phonemically. However, cross-linguistically dental fricatives /θ/ and /ð/ are rare – see Table 9 – and for most languages [θ] and [ð] are allophonic realizations of /s/ and /z/.

Yet even establishing that phones are in complementary distribution and that they seem to be phonetically similar is not always conclusive evidence in establishing whether or not two phones are allophones of the same phoneme. Clark, Fletcher and Yallop (2007: 97–8) report that in Italian there are three nasal phonemes: the bilabial /m/, the alveolar /n/ and the palatal /ɲ/. The nasal phoneme, before the velar consonants /k/ and /g/, assimilates predictably to [ɲ] as in [baŋka] bank. However, it is by no means clear what phoneme [ɲ] is an allophone of. The closest sounding phoneme is /ŋ/ but Italian orthography suggests that it is an allophone of /n/. There does not seem to be any principled reason why it could not be an allophone of /m/. Indeterminacies such as this have led some scholars to abandon phonemic analysis and propose alternate analyses based on underlying representations.

**Alveolar**

The term alveolar refers to the classification of a consonant sound on the basis of the place of articulation of the sound. Alveolar consonants are produced by contact or close proximity between the alveolar ridge, which is the bony structure behind the front teeth, and the tip or blade of the tongue. Assuming a normal pulmonic egressive airstream mechanism the following classes of consonants are classified by the IPA as alveolar: plosives [t, d], nasals [n], trills [r], taps [ɾ], fricatives [s, z], lateral fricatives [ɻ, ɭ], approximants [ɹ] and lateral approximants [ɻ].

The voiceless alveolar plosive consonant [t] is one of the most favoured consonants across languages. It is estimated (Ladefoged 2005: 156) that around 98 per cent of the world’s languages contain a variant of [t], though in some languages, such as French, the [t] has a dental place of articulation. Languages including French, Hungarian, Persian and Russian do not contain any alveolar phonemes; instead they have dental phonemes.

Table 1 summarizes the presence or absence of alveolar consonant phonemes across a number of major languages. Table 1 and all other tables that list phonemes in the 25 languages are based on data mostly found in the 1999 IPA handbook. It can be seen that there are eight possible alveolar manner of articulations with plosive, fricative and lateral approximants being the most favoured across the languages surveyed. The most marked manner of articulation is the lateral fricative which is found in Welsh and represented orthographically as ll as in place names such as Llanelli [ɻan'ɛli] and Llangollen [ɻan'goʊlən]

In English, in addition to the alveolar phonemes listed above, in some accents alveolar allophonic variations may be heard. In some forms of received pronunciation an <r> pronounced between two vowels may be realized as a tap, as in [vəɹi] Scottish English is commonly said to favour taps [ɾ], though this is a matter of some dispute. For further information see Wells (1982b: 411)
Table 1  List of alveolar phonemes found in 25 languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Plosive</th>
<th>Nasal</th>
<th>Trill</th>
<th>Tap</th>
<th>Fricatives</th>
<th>Lateral Fricatives</th>
<th>Approximants</th>
<th>Lateral Approximants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amharic</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td>r</td>
<td>s z</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Arabic</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>s z</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Bulgarian</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td>r</td>
<td>s z</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Cantonese</td>
<td>t</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td>s z</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Czech</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td>r</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td>r</td>
<td>s z</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td></td>
<td>s z</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td></td>
<td>s z</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Greek</td>
<td>n</td>
<td>r</td>
<td>s</td>
<td>z</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Hausa</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td></td>
<td>s z</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Hebrew</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td></td>
<td>s z</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Hindi</td>
<td>n</td>
<td>r</td>
<td>s</td>
<td></td>
<td>z</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Hungarian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Igbo</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td>s z</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>r</td>
<td>s</td>
<td>z</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td>t</td>
<td>tʰ</td>
<td>d</td>
<td>n</td>
<td>s z</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Mandarin</td>
<td>t</td>
<td>n</td>
<td></td>
<td></td>
<td>s</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Persian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>r</td>
<td>s z</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>n</td>
<td>r</td>
<td>r</td>
<td>s</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Thai</td>
<td>t</td>
<td>tʰ</td>
<td>d</td>
<td>n</td>
<td>r</td>
<td>s</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Welsh</td>
<td>t</td>
<td>d</td>
<td>n</td>
<td></td>
<td>r</td>
<td>s</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Here and in the other tables the cells that are shaded in grey indicate that the particular language does not contain any phonemes articulated at the relevant place of articulation.

and Stuart-Smith (2008: 64–5). A noticeable feature of American English and an apparently increasing feature in working-class New Zealand speech is what sociolinguists refer to as t-voicing. The speaker substitutes a tap for an alveolar stop. Thus, Ladefoged (2001: 151) claims that a General American pronunciation of *petal* will resemble a Scottish pronunciation of *pearl*. In both cases the speaker will say something like [pʰɛɹ].

**Amplitude**

Amplitude is an acoustic phonetic term that refers to the extent to which an air molecule moves to and fro around its place of rest. The greater the amplitude, or the distance between the points of rest, the greater the intensity. The greater the intensity the louder the sound will be perceived as.

**Aperture**

Aperture is a phonological term that is used to refer to contrasts involving the degree of openness of an articulation in models of phonology such as
autosegmental, metrical and firthian. For consonants, a tertiary classification of aperture is recognized. Complete closure of the oral tract results in stops and nasals. An aperture sufficient to generate a turbulent airstream results in fricatives. A more open aperture that is insufficient to generate a turbulent airstream mechanism results in approximants and semi vowels. For vowels the degree of aperture correlates with tongue height, with close vowels having a less open aperture than do open vowels.

Approximant
An approximant is a term used to refer to the classification of a consonant on the basis of its manner of articulation. Approximants are produced by a narrowing of the opening through which the airstream occurs. The narrowing is not close enough to produce friction. As a result these sounds are also known as frictionless continuants. The IPA recognizes two types of approximants: one where the airstream escapes over the tongue; the other known as a lateral, where the airstream escapes around the sides of the tongue. Approximants can be articulated at the following places of articulation: bilabial, labiodental, alveolar, dental, palato-alveolar, retroflex, palatal and velar.

In English there are four voiced approximant phonemes /ɔ/, /w/, /j/ and /l/. Voiceless approximants are possible in language though there are none in English. Approximants can be subdivided into two further classifications: liquids – in English /ɔ/ and /l/; and semi vowels – in English /j/ and /w/.

Archiphoneme
An archiphoneme is a phonological solution to the problem of phonemic neutralization. Phonemic neutralization refers to cases where the contrast between phonemes within a word is lost. For instance, in coda position in English there is no contrast between nasal phonemes, which are immediately followed by a plosive. The nasal before /p/ must be bilabial (e.g. limp /lɪmp/); the nasal before /t/ must be alveolar (e.g. lint /lɪnt/); and the nasal before /k/ must be velar (e.g. link /lɪŋk/). By choosing to transcribe the nasals in limp, lint and link as /m/, /n/ and /ŋ/, respectively, an analyst attributes a contrast that is not there! The contrast between the coda final plosive determines the place of articulation of the preceding nasal phoneme.

After /s/ in onset position the voicing contrast between plosives is lost. Yet speech is transcribed as /spɪtʃ/ and not /sbiːtʃ/, stitch is transcribed as /stɪtʃ/ and not /sdɪtʃ/ and skate is transcribed as /sketʃ/ and not /sɡeitʃ/. The plosive, though, cannot be voiced because the voicelessness of the initial fricative perseveres and devioes the following segment (e.g. snow [snoʊ] and slow [sloʊ]). The transcription of /p/, /t/ and /k/ after /s/ incorrectly implies the existence of an available voicing contrast. The solution to this problem proposed by the Prague School linguist Nikolai Trubetzkoy was to explicitly signal the phonemic neutralization by notating an archiphoneme, indicated by a capital letter, which is not identified with a particular phoneme but rather with the suspension of contrasts between the relevant phonemes. Thus, limp, lint and link are transcribed as
Articulation

Articulation is a phonetic term that refers to any physiological movement which modifies an **airstream mechanism** in the production of the sound. There are two distinct sets of articulators: the active articulators that move to create the stricture, and the passive articulators that the active articulators move towards. The active articulator is usually the tongue. However, in **bilabial** and **labiodental** sounds, the passive articulator is the lower lip and, in the articulation of /h/, the **vocal folds**.

The tongue itself is usually segmented into five distinct regions. The tip or apex, which is the front-most extremity of the tongue, is especially flexible and can touch the roof of the mouth as far back as the hard palate. It is the active articulator in **retroflex** sounds and in **trills**. Sounds that are articulated with the tip of the tongue are classified as apical. The blade is the upper surface of the tongue immediately behind the tip. When the tongue is at rest it is opposite the alveolar ridge. It is the active articulator in the production of **dental**, **alveolar** and **palato-alveolar** sounds. Sounds that are articulated with the blade of the tongue are classified as laminal. The front is the region of the tongue below the hard palate when the tongue is at rest. It is the active articulator in the production of **palatal** sounds. It may be used to articulate fronted **velar** stops in words such as *key* and *geese*. The back is the region of the tongue below the soft palate when the tongue is at rest. It is the active articulator in the production of **velar** sounds. Sounds that are articulated with the back of the tongue are classified as dorsal. The root of the tongue is the region opposite the pharynx and it is not used as an active articulator in languages other than Semitic ones. However, the root of the tongue is involved in the production of sounds that are classified as being made with an **advanced tongue root**. It may be involved in the production of **tense** sounds.

Sounds are classified in terms of their manner of **articulation**. This refers to the degree of closure caused by the movement of the active articulator and the passive articulator. If the airstream mechanism is able to flow out of the mouth unimpeded the sound is a vowel or a semi-vowel. Sounds produced with a closure that impedes the airflow are consonants. Those produced with a complete closure of the mouth are known as stop consonants. If the soft palate is raised and the airstream is unable to flow out the nasal passage the sound is a **plosive**, as in /p/, /b/, /t/, /d/, /k/ and /g/. Conversely, if the soft palate is lowered and the airstream is free to flow out through the nose the sound is a **nasal** stop, as in /m/, /n/ and /ŋ/. Sounds where the active articulator approaches the passive articulator so closely that it results in audible friction are **fricative** consonants (e.g. /f/, /v/, /θ/, /ð/, /s/, /z/, /ʃ/, /ʒ/ and /h/). Sounds where the organs do not approach closely together are known as **approximants**, for example, g /l/ and /ɹ/.

The passive articulators define the place of **articulation**. If the upper lip is the passive articulator the sound is **bilabial**. /p/ and /b/ are bilabial plosives, /m/ is a bilabial nasal stop. Sounds produced at the teeth have a **labiodental** or **dental**
place of articulation. /f/ and /v/ are labiodental fricatives, and /θ/ and /ð/ are dental fricatives. Those produced at the alveolar ridge, the bony prominence behind the upper teeth, have an alveolar place of articulation. /t/ and /d/ are alveolar plosive consonants, /n/ is an alveolar nasal, /s/ and /z/ are alveolar fricatives and /l/ and /ɹ/ are alveolar approximants. Sounds produced at the part of the mouth between the alveolar ridge and the beginning of the hard palate have a palato-alveolar place of articulation. /ʃ/ and /ʒ/ are palato-alveolar fricatives. Sounds produced at the hard palate, the arched bony structure behind the front teeth, have a palatal place of articulation. /j/ is a palatal approximant.

Sounds produced at the soft palate or velum, the mobile fleshy area contiguous to the back of the hard palate, have a velar place of articulation. /k/ and /ɡ/ are velar plosives and /ŋ/ is a velar nasal. Sounds that are made at the uvula, the small appendage that dangles from the back of the soft palate, have an uvular place of articulation. There are no English uvular phonemes. The remaining passive articulators, with the exception of the glottis, are not used in the English sound system. These are the pharynx, the tubular cavity immediately above the larynx and the epiglottis, the anatomical structure that closes over the larynx. The larynx itself is made up of cartilage and muscle and contains the vocal folds. The space between the vocal folds is the glottis and sounds that are made by the narrowing of the space between the vocal folds have a glottal place of articulation, for example /h/, which is a glottal fricative. The remaining place of articulation is retroflex. This refers to a sound produced by the touching of the tip of the tongue against the front of the soft palate.

**Articulatory phonetics**

Articulatory phonetics is the subfield of phonetics, which studies how speech sounds are articulated by the vocal organs. It complements acoustic and auditory phonetics.

**Articulatory setting**

An articulatory setting is the medium to long term setting of all the articulators in relation to one another. It varies both between languages and accents, and between individual speakers of the same dialect. Differences in articulatory settings between languages/accents account for many of the stereotypical differences between accents. For instance, Laver (1994: 411) explains how differences in the setting of the tongue body result in a ‘velarized voice’ typical of Liverpool and Birmingham accents. Differences in articulatory settings between speakers result in some speakers producing more nasalized or pharyngealized speech compared to others in their speech community.

**Aspiration**

Aspiration is a phonetic term for the audible breath that accompanies a sound’s, usually a plosive’s, articulation. Aspiration is notated by a superscript [ʰ] placed immediately after the consonant symbol. In almost all English accents voiceless plosives are aspirated when initial in stressed syllables, as in *pin, tin* and *kin* are
pronounced as \([p^h]\text{n}\), \([t^h]\text{n}\) and \([k^h]\text{n}\). Celtic-influenced accents tend to have more strongly aspirated syllable initial plosives than RP does. If \([l, \, s, \, w, \, j]\) immediately follow an aspirated phoneme they are devoiced and articulated as \([l, \, s, \, w, \, j]\). Readers may note the difference in voicing by slowly and carefully articulating the following pairs of words:

- Lay \([\text{lei}]\)       Play \([p^h]\text{lei}\)
- Ray \([\text{rei}]\)       Pray \([p^h]\text{rei}\)
- Ween \([\text{win}]\)    Queen \([k^h]\text{win}\)
- You \([\text{ju}]\)       Pew \([p^h]\text{ju}]\)

There are a number of exceptions to the ‘rule’ that initial voiceless plosives are aspirated. Wells (1982b: 370) reports that speakers from the Pennine Valleys north of Manchester may have little or no aspiration before a stressed vowel. Clark (2008: 169) reports that aspiration of syllable final plosives may be a feature of West-Midland accents. South African speakers, presumably because of interference from Afrikaans, tend not to aspirate voiceless plosives that precede stressed vowels (Wells 1982c: 618). Wells further reports (ibid.: 625) that it is a well-known feature of Anglo-Indian accents that voiceless plosives are unaspirated in all positions.

A simple test that readers can do to check if, and how strongly, they aspirate voiceless plosives is to slowly and carefully pronounce words with initial and final voiceless plosives such as \(\text{pan}, \, \text{nap}, \, \text{tan}, \, \text{gnat}, \, \text{can}, \, \text{knack}\) with their hand held a few inches in front of their mouth. If a plosive is aspirated readers will perceive a burst of air striking their hand. When I read the above words my pronunciation – and no doubt that of the majority of the readers – is \([p^h\text{æn}], \, [næp], \, [t^h\text{æn}], \, [næt], \, [k^h\text{æn}]\) and \([næk]\). Readers are likely to notice that the burst of air following \([p]\) is the strongest, and that following \([k]\) the weakest. This is because the closure for \([p]\) is at the lips, thus there is a far larger amount of compressed air waiting to be released once the lips are opened than there is when the stop closure is made by the back of the tongue against the velum.

Many languages have voiceless aspirated plosives either as phonemes or allophones, but a few languages such as Hindi and Igbo are reported to have voiced aspirated plosives. Ladefoged and Maddieson (1996: 69–70) note, however, that voiced aspirated plosives do not contain a period of voicelessness or audible breath. Consequently they argue that if these sounds are truly aspirated, aspiration must be redefined in terms of voice onset time ‘as period after the release of a stricture and before the start of regular voicing in which the vocal folds are markedly further apart than they are in modally voiced sounds’.

**Assimilation**

Assimilation is a process that can lead, in historical terms, to a sound change. Sounds assimilate when a feature of one sound changes to match those of another feature, which either precedes it or follows it. In connected speech where words are pronounced without audible gaps between them the pronunciation of one phoneme may alter how a nearby phoneme is pronounced. In English, assimilation is more likely to occur in rapid casual speech rather than in slow careful
<table>
<thead>
<tr>
<th>Nominative</th>
<th>Accusative</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[i porta]</td>
<td>[tim borta]</td>
<td>the door</td>
</tr>
<tr>
<td>[i trapeza]</td>
<td>[tin drapeza]</td>
<td>the bank</td>
</tr>
<tr>
<td>[i kori]</td>
<td>[ti] gor[i]</td>
<td>the girl</td>
</tr>
</tbody>
</table>

speech. This, however, is not the case in other languages where assimilation is not optional but rather obligatory. Table 2 (adapted from Lodge 2009: 149–50) provides examples of assimilation from Greek when a nominal feminine element switches from the nominative to the accusative case. In Greek the clitic /i/ signals a feminine singular noun in the nominative case while /tin/ signals a feminine singular noun in the accusative case.

The Greek examples illustrate that there is more than one kind of assimilation. In the /porta/ example, the place of articulation of the final nasal in /tin/ is influenced by the bilabial place of articulation of the initial consonant /p/ in porta, and that nasal assimilates to the bilabial place of articulation. Thus /n/ is realized as [m]. Simultaneously, the voicing produced during the articulation of the nasal /n/ is maintained and the word initial voiceless phoneme /p/ is realized as a voiced sound [b]. In the trapeza example, the final nasal of /tin/ and the initial consonant in /trapeza/ are both alveolar but, as in the previous examples, the voicing of the nasal is maintained. The initial phoneme of the nominal element /t/ is realized as [d]. In the final example kori, as in the earlier /porta/ example, the place of articulation of the nasal assimilates to that of the initial consonant and /n/ is realized as [ŋ]. As in the other two examples, the voicing of the nasal is maintained and /k/ is realized as [g].

The examples set out in Table 2 indicate that there are two main types of assimilation: regressive (also known as anticipatory) and progressive (also known as perseverative). In the Greek examples, above, the place assimilation is regressive while the voicing assimilation is progressive. Regressive assimilation means that a phoneme that comes later in time influences a phoneme which comes earlier in time. For instance, if an English speaker wishes to say the phrase ten guns in fast casual speech, it is likely that the /n/ in ten will assimilate to [ŋ] in anticipation of the following velar stop /g/.

Progressive assimilation, which is not as common as regressive assimilation, is the opposite case and refers to a situation where the phoneme that is earlier in time influences the pronunciation of a phoneme that is later in time. English examples of progressive assimilation may occur in fast and casual pronunciations of words such as happen and quicken. Assuming that the nasal is syllabic the words may be realized as ['hep predictable or ['kwik predictable. The place of articulation of the plosive is preserved and continues during the articulation of the nasal. In fast and casual articulations, the sequence of words in the is often realized as [tŋ na]. The influence of the alveolar nasal perseveres and changes the dental fricative /ð/ into a dental nasal [ŋ]. At the same time, the dental articulation of the fricative influences the place of articulation of the nasal and both are articulated as dentals, which is an example of regressive assimilation.
A further type of assimilation is coalescent assimilation, where each of the adjacent sounds reciprocally influence one another’s pronunciation. In English the sequence /t/, /d/, /s/ and /z/ plus /j/ may coalesce to /tʃ/, /dʒ/, /ʃ/ and /ʒ/. Within word boundaries this coalescence has led historically to the presence of /tʃ/, /dʒ/, /ʃ/ and /ʒ/ medially in words such as nurture, grandeur, mission and pleasure. Some examples of possible coalescences across word boundaries are:

| They hit you /ðeɪ hɪt ˈjuː/ | becomes [ðeɪˈhɪtˌfu] |
| I need you /aɪ ˈniːd ˈjuː/ | becomes [aɪˈniːdʒu] |
| I miss you /aɪ ˈmɪs ˈjuː/ | becomes [aɪˈmɪʃu] |
| I lose you /aɪ ˈluːz ˈjuː/ | becomes [aɪˈluːʒu] |

In connected speech, assimilation may change a phoneme’s place of articulation, voicing or (more rarely) manner of articulation. A typical example of regressive assimilation leading to a change in voicing is where a voiced obstruent is followed by a voiceless consonant, for instance the /b/ in [æbˈætslʊtli] is devoiced. Voiceless fricatives found in a fully voiced environment tend to be voiced. This happens in three phonetic environments: first between two vowels in words such as Asia realized as [ˈeɪʒə]; second where the fricative is immediately preceded by a voiced consonant and immediately followed by a vowel, as with transit realized as [ˈtraŋzɪt]; third where the voiceless fricative immediately follows a vowel and is itself immediately followed by a voiced consonant, as with Muslim realized as [ˈmʊzlɪm].

In this section, the assimilation described is continuous or contact assimilation but the influence of a phoneme may extend beyond the immediately adjacent sound. For instance, Crystal (2003: 38) reports that a possible articulation of the expression turn up trumps is [tʃ:mpɬaʊmpz] where the nasal phoneme in turn has assimilated to the place of articulation of later consonants. While non-continuous or distance assimilation is rare in English it is common in many other languages where it is known as harmony.

**Auditory phonetics**

Auditory phonetics is the subfield of phonetics that studies how people perceive speech sounds and how they interpret their meaning. It complements acoustic and auditory phonetics. Traditionally auditory phonetics has been ignored in many treatments of phonetics, which have gone no further than (1) describing how a speech sound is made and (2) detailing the physical properties of the speech sound. A solid and approachable introduction to auditory phonetics is Ball and Rahilly (1999: chs 10 and 11). Today, as a result of numerous psychoacoustic studies, much is known about how people perceive sound, including speech, but far less is known about how people interpret meaning from the speech signal. Accordingly, the following paragraphs describe how people hear sound, including speech, prior to sketching some of the little that is known about how people listen to speech.

Air set in motion by the vocal organs oscillates. As speech is directed at hearers the vibrating air reaches the hearer’s outer ear, the visible part of the ear, the pinna or auricle. The pinna itself is largely irrelevant to the ability to hear. It serves merely
to channel the vibrating body of air into the ear canal. The ear canal is a tube that connects the pinna with the middle ear. It is approximately 25 mm long with a diameter of around 7 mm, although the shape and size vary between individuals. The ear canal functions as a resonator. It amplifies the sound waves that travel through it. It is particularly responsive to sounds between 3,000 Hz and 4,000 Hz, although it can amplify sounds between 500 Hz and 12,000 Hz.

After passing through the ear canal the sound wave enters the middle ear and strikes the air-filled eardrum, or tympanic membrane, which acts as a seal between the outer and middle ear. The eardrum contains three connected bones collectively known as the ossicles. The bones individually are the malleus (hammer), the incus (anvil) and the stapes (stirrup). Sound entering the middle ear pressurizes the bones and causes them to vibrate. The vibration of the bones of the middle ear converts the lower-pressure eardrum vibrations into higher pressure sounds that can be transmitted to the inner ear. The eardrum, as well as amplifying sounds, regulates the volume and protects the inner ear from damage.

Sound is transmitted to the inner ear through the oval window that is connected to the stapes. By the time the sound exits the middle ear it has been amplified by a factor of between 20 and 30 times. The inner ear contains the cochlea, which is the organ that converts sound waves into neural impulses. The cochlea is a coiled structure rather like a snail’s shell. It is filled with a watery liquid and divided by two membranes, the vestibular membrane and the basilar membrane which divide the cochlea into two chambers, the scala vestibule and the scala tympani. It contains the organ of Corti, which is a structure deep inside the inner ear consisting of hair cells resting on the basilar membrane.

Sound pressure variations at the oval window cause the fluid in the upper chamber, the scala vestibule, to vibrate. This in turn triggers vibration in the fluid inside the scala tympani, the organ of Corti and the basilar membrane. The nature and location of the vibration along the basilar membrane is dependent on the frequency of the sound waves, with higher frequencies causing vibrations near the oval window and lower frequencies causing vibrations near the organ of Corti. The vibration of the basilar membrane results in movement in the hair cells in the organ of Corti. It is this movement that converts the vibration into the electrical impulses which are transmitted to the auditory centre of the brain as neural impulses.

The organ of Corti contains thousands of hair cells. As people age they experience a loss in the number of hair cells. This results in them being unable to hear higher frequency sounds. However, it is worth remembering that humans, unlike some other animals, can only hear sounds up to around 20,000 Hz and that speech sounds are relatively low frequency sounds.

Little is as yet known about how the brain categorizes sounds, though it is believed that there is a feedback loop from the brain to the organ of Corti, which helps hearers monitor their hearing. Experiments have demonstrated that humans categorize speech digitally. For instance, rather than hearing multiple potential vowels between cardinal vowels 1 and 2, hearers tend to perceive two categorical vowels and experimenters have succeeded in identifying the point at which one vowel switches into the other.
Some of the acoustic cues that allow listeners to decode the speech signal are known. Speakers of English distinguish word initial voiced and voiceless stops in syllables such as [pæ:] and [bɔ:] by the amount of time it takes voicing to commence on the vowel – see voice onset time. They use perceived vowel length to distinguish whether a final stop consonant is voiced or voiceless (e.g. /æt/ and /æd/) – see fortis and lenis. It is believed that speakers use formant2 transitions to perceive the place of articulation of a consonant. An F2 value that is lower than would be expected had the vowel been produced in isolation signals a bilabial consonant, while an F2 value that is higher than had the vowel been produced in isolation signals a velar consonant.

However, far less is known about how hearers become listeners and decode meaning from the speech signal. Two models of listening, active and passive, are proposed in the literature (see Crystal 1997: 148). Both of the proposed models are problematic in one way or another. So it may well be that both models of listening are available to listeners to utilize when appropriate. Active models of listening claim that listeners use their pre-existing knowledge of the language to decode novel strings of speech. There are two main types of active models. The first, ‘the motor theory’ of speech perception (see Liberman 1996) proposes that listeners identify sounds by internally shadowing the articulatory movements of the speaker. This model is able to explain the rather puzzling McGurk effect, which suggests that speech perception is multimodal. The effect can be produced by video recording a speaker repeatedly producing an open syllable such as [gagagaga], removing the original soundtrack and dubbing the speaker’s production of a different string of open syllables, such as [dadadada]. Hearers who watch the video recording will perceive the string of sound as [gagagaga], though if they close their eyes they will hear it as [dadadada]. However, it singularly fails to explain how individuals with pathological speaking disorders are able to successfully listen to speech.

The second influential active listening approach is ‘analysis by synthesis’ (e.g. Stevens and Halle 1967), which proposes that humans perceive speech by implicitly synthesizing speech by using their internal distinctive feature matrix to match the auditory input. Perception begins when the listener analyses the speech signal into a series of auditory patterns, which are then represented in terms of distinctive features. The resulting distinctive feature matrix, assuming that there has been no significant contextual interference, is assigned to a phoneme. In cases where there has been significant contextual interference, the listener hypothesizes a representation of the distinctive feature matrix and compares it with various representations generated by the listener’s own internal phonology. All proposed models of active listening – readers interested in an accessible description of various active listening models are advised to read Yeni-Komshian (1998) – are able to account for attested experimental effects such as the listener’s ability to fill in missing phonemes or words in a contextualized speech signal. However, they are unable to explain misidentifications such as /dɔn/ being perceived as /dʒɔn/ – see below.

Passive models of listening propose that listeners hear a message and match the incoming signals to a set of templates stored in the brain. If a match is located the
sound is recognized, but if (as in the case of a foreign language) no internal template is available the listener is unable to decode the sound signal. Ball and Rahilly (1999: 187) note that it is by no means clear what type of internal templates listeners would need to enable them to decode the speech signal. Listeners in response to the question, *Was it John or Mike?*, who hear the answer string it was /dɔn/ are almost certain to interpret /dɔn/ as /dɔn/. To account for these facts, Crystal (1997: 148) proposes that a combination of active and passive theories is required to explain the process of speech perception.

**Autosegmental phonology**

Autosegmental phonology is an influential *generative* model of phonology that was developed by John Goldsmith in his 1976 PhD dissertation. The most accessible theoretical introduction to the model is Goldsmith (1999); Goldsmith (1990) is a comprehensive textbook on the model. While Goldsmith overtly situates autosegmental phonology within the generative tradition some scholars, notably Lass (1984: 269fn.10.2.3), have noted the strong resemblance between autosegmental phonology and earlier models, especially *firthian prosodic phonology*. Despite this resemblance, autosegmental phonologists share many assumptions with generative phonologists, notably the existence of an underlying lexical representational level, which generates forms that represent the input to a series of linearly ordered rules, the operation of which generates surface forms.

The model was developed through the study of African *tone* languages and designed to provide a more phonetically grounded and less abstract phonological representation of language. The main insight of the model is that the phonological representation of a language comprises several autonomous *tiers* or levels with each tier consisting of a linear arrangement of elements. Tiers are linked by *association lines*, which formally represent the relation between the elements notated on the different tiers.

To illustrate using an English example based on Goldsmith (1999: 137–8), the phonological representation of /tɛn/ is:

\[
\begin{array}{cccc}
+ \text{Coronal} & + \text{Vocalic} & + \text{Coronal} \\
- \text{Nasal} & + \text{Dorsal} & + \text{Nasal} \\
+ \text{Consonantal} & - \text{High} & + \text{Consonantal} \\
- \text{Voice} & - \text{Low} & + \text{Voice} \\
- \text{Continuant} & - \text{Back} & - \text{Continuant} \\
+ \text{Anterior} & + \text{Anterior} & \\
/\text{t}/ & /\text{ɛ}/ & /\text{n}/ \\
\end{array}
\]

Yet a speaker is likely to pronounce *ten* as an utterance that can be notated as following: [tʰɛn]. This utterance is produced by the following articulatory sequence, or in Goldsmith’s terminology *orchestration* (see Table 3).

The aspiration that in segmental models is linked to [t] where it is classified as [+heightened subglottal pressure] continues into the articulation of the vowel. Nasality that is classified as belonging to the [+Nasal] /n/ in fact commences during the articulation of the vowel. To account for these facts, autosegmental
Table 3  A score for the orchestration of ten assuming a falling tone

<table>
<thead>
<tr>
<th>Lips:</th>
<th>Spread ......................................................................................................................</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongue:</td>
<td>Touch the palate ....... High and Front ........ Touch the palate...............................</td>
</tr>
<tr>
<td>Velum:</td>
<td>Raise ........................................ Lower ................................................................</td>
</tr>
<tr>
<td>Vocal folds:</td>
<td>Apart ......... Vibrate faster ........ Vibrate slower ...........................................</td>
</tr>
<tr>
<td>Segments:</td>
<td>/t/ /e/ /n/</td>
</tr>
</tbody>
</table>

phonology proposes a representation where certain features are autonomous and linked simultaneously with more than one feature, for example:

```
+ Coronal  + Vocalic  + Coronal
– Nasal    – High
+ Consonantal  + Dorsal  + Consonantal
– Continuant  – Low
  + Anterior  – Back
```

In the autosegmental representation of ten the nasality and voicing features are represented on their own autonomous tiers where they are free to associate with more than one segment.

The model was initially developed to account for five problems found in the study of African tone languages. The first of which contour tones is where a falling, rising or compound tone movement is attached to a word final vowel. Convincing phonetic evidence exists – see tone – demonstrating that lexical tones in African tone languages can be usefully analysed as a series of High or Low pitch targets. Thus, where the lexical tone on a word final vowel falls or rises, a segmental description runs into trouble, as with the word àkálà (from Goldsmith 1999: 143):

```
L H H L Tonal tier
```

It is impossible to ascribe the feature [+ high pitch – low pitch] and [– low pitch + high pitch] to the final vowel but the contour tone on the final vowel is easily explainable if tone is represented on an autonomous tier. A similar argument would be made by intonation scholars whose work is grounded in autosegmental phonology, for instance B’s utterance with a rising tone in the following example:

A: Who wants this?
B: ↗ ME
In autosegmental notation *me* would be represented as:

```
/m  i/
```

Goldsmith labelled his second reason *stability*. This refers to the situation where a vowel is elided but the accompanying tone is not deleted. If the tone were a feature of the vowel there would be no reason why it and not some other feature of the vowel segment should be preserved. His third argument, *melody levels*, is similar to Firth’s notion of prosodies. Goldsmith notes that different types of features attach to different segments in different lexical contexts. For instance, English voiceless stops in stressed syllable onset initial position are aspirated while voiceless stops in coda position are normally not aspirated. If the feature [+Delayed Release] is part of the phonological matrix which forms [p] it would be equally likely to occur in final or middle position as in initial position. Yet, this is demonstrably not the case. A description that places aspiration on an autonomous tier, where it is free to associate with the relevant feature, is more theoretically accurate.

The fourth justification for the model refers to *floating tones*. These are tones that are not known to be attached to any underlying vowel. While easy to explain in an autosegmental model, these are theoretically suspect in segmental models. According to segmental models the underlying vowel that the floating tone is posited to be attached to must surface in the pronunciation. As the vowel fails to surface, floating tones are an awkward anomaly for generative segmental descriptions. The final reason is *harmony*. Segmental models can notate phonological features such as vowel harmony. However, unlike autosegmental models they are unable to offer descriptions, which can generalize all instances of harmony in an economical manner.
Index

Accent, 1–4
   African American English, 63
   American, 1, 10, 32, 62, 63, 109, 114, 118
   Anglo-Indian English, 114
   Australian, 2, 33, 35, 154
   English, 2; Lancashire, 2, 51; Liverpool, 6,
      13, 22, 37, 61; London, 33, 36, 46, 63;
      Northumbria, 143; Tyneside, 35, 36,
      154, 156; West Midlands, 13, 35, 154,
      156
   Irish English, 3, 32, 33, 63, 85, 101, 114,
      116, 146, 154, 156
   Jamaican English, 122
   New Zealand, 10, 33, 35, 85, 109, 114, 154,
      157
   Prominence, 4, 70, 79, 91
   Scottish, 3, 36, 37, 63, 86, 116, 143, 146,
      154, 157
   South African English, 144
   Ulster, 3, 36, 101, 103, 154, 156
   Welsh English, 3, 4, 32, 36, 63, 88, 106,
      143, 144, 154
   West Country, 63, 116
Acoustic phonetics, 4, 64
   Acute, 5
   Advance Tongue Root, 4, 64
   Affricate, 6–7
   Affrication, 6, 22, 146
   Airstream mechanism, 7
   Allophone, 7–9, 51, 101, 103
   Alveolar, 9–10
   Amplitude, 4, 10, 78
   Aperiodic sounds, 5
   Aperture, 10–11
   Apical, 12, 41, 144
   Approximant, 11
   Archiphoneme, 11–12
   Articulation, 12–13, 39
      Manner, 3, 6, 9, 11, 12, 27, 28, 39
      Place, 13, 66, 67, 93, 103, 106
   Articulatory phonetics, 13
   Articulatory setting, 13
   Aspiration, 13–14, 83, 114, 149
Assimilation, 8, 15–16, 47, 67
   Coalescent, 16
   Progressive, 15
   Regressive, 15
   Auditory Phonetics, 16–19
   Autosegmental Phonology, 19–21, 90, 138
Bilabial, 22–3
   Breathy Voice, 24, 71
Cardinal Vowels, 17, 25–7, 34, 46, 151
Celtic influences, 22, 61, 103, 114
Clear<—L>, 7, 8
   Clicks, 27–28, 148
   Coalescence, 28, 104, 107
   Coarticulation, 29
   Compensatory Lengthening, 29
   Complimentary distribution, 7, 8, 107
   Consonant, 29, 37, 39, 40, 41, 43, 44, 101,
      102, 103, 105, 113, 115, 116, 118, 120,
      143, 144, 145, 146
   Constraint, 30, 95, 96, 97
Constriction, 31
   Contoid, 29
   Contrastive distribution, 7
   Creaky Voice, 31, 71
Dark<—L>, 7, 8, 32–3, 84, 148
   Deletion, 51, 67, 96, 112, 131
   Dental, 9, 33
      Stops, 4, 33, 114
   Devoicing, 47, 61, 83, 114
   Diphthong, 34–7, 153, 157, 162
   Distinctive Feature, 5, 18, 19, 37–48, 65, 95,
      109, 162
   Unary, 42
   Distribution, 48
   Dorsal, 12, 42
   Double Articulation, 48
   Duration, 48–9, 120
   Egressive, 7, 50, 71, 147
   Ejectives, 50–1, 71
   Elision, 51–2, 96, 112, 131
Epenthesis, 47, 52–3, 96
Epiglottal, 105, 106
Estuary English, 53–4

Feature geometry, 45
Firthian Prosodic Phonology, 19, 55–6
Flap, 56
Foot, 56, 57, 90, 110, 136, 137
Formant, 5, 18, 26, 57–9, 163
Fortis, 18, 59, 111
Fortition, 60–1, 120, 164
Free Variation, 61–62, 107, 125, 163
Frequency, 62
Fricative, 62–3
Fricionless Continuant, 64
Fundamental Frequency, 64, 148

General American, 4, 86, 101, 109, 114
Generative Phonology, 65–9
Glide, 69, 111
Glottal, 69–70, 114
Glottalic, 70–1
Glottis, 71
Grave, 71
Great Vowel Shift, 71, 73

Harmonic components, 5
Harmony, 21, 55, 56, 74–5
<h> dropping, 63, 71

Implosive, 76–7
Information, 78, 79, 126
Given/New, 79
Ingressive, 7, 27, 76, 77, 147
Initiator, 77
Intensity, 78
Interdental, 78
Intonation, 31, 64, 78–82, 114, 138, 139, 142
Tonality, 78, 79, 81, 98–9, 139, 163
Tone movement, 4, 20, 80, 139, 140
Tonicity, 4, 79
Intrusive <r>, 89

J or (yod) dropping, 53, 103, 111, 118
Juncture, 83–4

Labialization, 85
Labial-velar, 3, 85–6
Labiodental, 86
Laminal, 12, 41
Larynx, 5, 13, 69, 71, 76
Lateral, 32, 88, 113
Fricative, 63, 88
Lax, 88
Lenis, 18, 59, 88
Lenition, 114, 120, 164
Liaison, 88–9
Linking<->, 88
Liquid, 89, 111
Listening, 18

Markedness, 90, 96
McGurk effect, 18
Metrical Phonology, 57, 90–1
Mora, 91–2, 132

Nasal, 8, 60, 93–4, 113, 119, 163
Nasalization, 94, 153
Natural class, 46, 47
Natural Phonology, 30
Neutralisation, 12, 13, 94

Obligatory Contour Principle, 95
Obstruent, 6, 7, 62, 63, 95
Old English, 119, 121
Onset Maximalisation, 95, 96, 97, 129, 131
Optimality Theory, 31, 95–100
Oral, 100

Palatal, 101–3
Palatalization, 4, 65, 103, 114
Palato-Alveolar, 103–5
Parametric speech, 105
Paratones, 81
Periodic sounds, 4
Pharyngeal, 105–6
Pharyngealization, 106

Pharyngeatalization, 106
Phonation, 50, 77, 106–7
Phoneme, 6, 7, 9, 37, 93, 107–9, 127
Phonetics, 109
Phonetic similarity, 8, 46
Phonological word, 90, 110
Phonology, 109
Phonotactics, 68, 110–113
Pitch, 4, 81, 82, 113, 120
Plosive, 113–14
Praat, 4
Prosody, 114
Pulmonic, 7, 114
Received Pronunciation, 9, 70, 86, 115, 152, 154, 155
Retracted Tongue Root, 106
Retroflex, 115–16, 157
Rhoticity, 1, 57, 117, 153
Rhythm, 56, 91, 114, 125, 132, 135
Rule, 65, 66
Scottish vowel length rule, 155
Secondary articulation, 85, 94, 103, 107, 118, 145, 148
Semi vowel, 29, 118
Sibilant, 118–19
Sonorant, 119
Sonority, 98, 130, 131
Sound change, 51, 60, 68, 71, 72, 73, 119–20
Speech error, 127, 133
Speech Perception, 18, 19
Stop, 120
Stress, 90, 120–6
Clashes, 125, 126
Compound, 124, 125
Phrasal, 90, 125, 126
Primary, 120
Secondary, 164
Strong form, 120
Syllable, 111, 126–33
Closed, 128, 129, 158, 164
Coda, 97, 112, 128
Nucleus, 128, 129, 132
Onset, 97, 128, 164
Open, 111, 112, 128, 129
Rhyme, 128, 132
Tap, 56, 109, 114, 134
Target, 105, 134, 140, 141, 142
Tense, 59, 134–5
Timing, 132, 133, 138
Stress, 136, 137
Syllable, 136, 137
ToBI, 138–40
Tone (Lexcial), 19, 20, 21, 75, 138–42
   Contour, 139, 149
   Register, 140
Tongue, 27
Tonogenesis, 141, 142
Transcription, 142
Trill, 12, 143
T-Voicing, 10, 109, 114
Underlying Representations (Forms), 66, 67, 68
Underspecification, 45
Uvular, 144
Uvularization, 145
Velar, 9, 11, 12, 146–7
Velaric, 7, 27, 147–8
Velarization, 8, 116, 148
Vocal folds, 148
Vocal organs, 4, 77, 148
Vocal tract, 7, 148
Vocoid, 29, 32, 118
Voice, 149
Voiceless, 149
Voice onset time, 14, 114, 149–50
Voice quality, 24, 31, 150
Vowel, 33, 34, 35, 36, 37, 41, 45, 58, 150–8
   Checked, 122, 123
   Free, 122, 123
   Waveform, 5
Weak form, 51, 159–61