An overview of phonetics for language teachers

Joanna Przedlacka

Introduction

Hearing an unfamiliar language for the first time means perceiving a continuous stream of sounds, some of them resembling those of our native language. Others, previously unheard, appear to an unaccustomed ear exotic and hard to describe. This experience is far removed from the one of listening to and speaking our native language. Using our mother tongue we are able to segment the flow of speech into words, understand their meanings and respond accordingly. This multi-level linguistic processing operates below the level of conscious awareness. Doing so, we do phonetics. Its knowledge in the case of our native language is subconscious and taken for granted.

Phonetics is the science of speech. The goal of descriptive phonetics is to account for all the sounds of human languages. It provides a framework to describe speech at many levels. We can list the vowels and consonants of a language and present their properties. We can also provide an account of the changes in melody of the voice (intonation) throughout an utterance and how they may influence meaning. Any variety can be described to a good degree of detail. We can also characterize developing, disordered or learner speech. Phonetics deals with the physical properties of sounds, the possibilities of the human vocal tract and the representation of speech. Phonology is interested in how sounds function as a system, i.e. how they form patterns in languages and which physical properties of sounds contribute to the differences in meaning.

Pronunciation instruction, which finds its place in foreign and second language teaching, is an obvious practical application of phonetic science. Efficient communication in the target language depends on the ability to perceive and distinguish its sounds and making one’s own speech intelligible. Studying, classifying and describing speech enable us to learn another language, thus making the transition from perceiving a string of unfamiliar sounds to making sense of them. Phonetic knowledge is necessary to alter pronunciation, an area of L2 learning notoriously resistant to change. It is a precursor to phonetic awareness and, combined with practice, it allows us to produce sounds in the appropriate contexts and consequently communicate efficiently in a language not spoken natively.

In the process of helping learners achieve this goal, teachers tackle several tasks, such as selecting a pronunciation model, setting targets, diagnosing the students’ difficulties and
designing a syllabus. In their expertise portfolio, teachers need the knowledge of the systemic
and phonetic differences between the native and target languages, coupled with an auditory
ability to distinguish sounds. The training and maintenance of this skill, key for a phonetician
(Ladefoged & Johnson, 2015), is also indispensable for pronunciation instructors.

This chapter discusses key concepts in phonetics indispensable to teaching practice. It
begins by setting out the basics of speech production and looks at different levels of organ-
ization in a language, such as segment, syllable and suprasegmentals. It is shown how aspects
of speech can be represented. Theoretical concepts such as the linguistic status of sounds are
placed in the context of second language learning. Finally, the issue of pronunciation model
choice is discussed. Examples from English and other languages are used throughout.

An overview of speech production

As we breathe, the air travels from the lungs, moving upwards through the trachea (windpipe)
and into the oral and nasal cavities on its way out. It is also the building material for speech,
being modified at several locations along the vocal tract (Figure 3.4). Those changes
influence the final ‘shape’ of the sounds we hear.

Voicing

The first modification occurs at the larynx (voice box), a cartilaginous structure sitting atop
the windpipe. Its front wall protrudes more in males due to the sharper angle of the thy-
roid cartilage (Adam’s apple). The back wall of the larynx and its floor are formed by the
cricoïd cartilage. On its top sit two further cartilages, called aritenoids. The larynx is shown
in Figure 3.1.

The larynx houses the vocal folds (Figure 3.2), two flaps of soft, fleshy membranes,
attached horizontally at the front to the thyroid cartilage and at the back to the aritenoids.
The space between the vocal folds is the glottis. See the vocal folds in action at www.
linguistics.ucla.edu/faciliti/demos/vocalfolds/vocalfolds.htm.

The vocal folds are moved by muscles attached to the aritenoids and can adopt different
settings. They contribute to speech in a number of ways. In voiceless sounds the folds are
held apart, allowing the air to flow freely and producing no sound. In voiced sounds they are
close together and vibrating when impacted by the lung air. In English we distinguish between
pairs such as <pat–bat>, <seal– zeal>, <bicker–bigger>, but a third of the languages surveyed
for the World Atlas of Language Structures Online do not have the voiced–voiceless contrast
(Maddieson, 2013). The difference can be demonstrated by placing your palm on the larynx
when articulating sounds such as s–z and feeling the vibration in the latter, or by covering
your ears and feeling the sensation of the resonance. In English, vocal folds also contribute
to the production of two consonants /h/ and the glottal stop /ʔ/, as in <uh-uh> ‘no’ and
<uh-huh> ‘yes’. The vocal folds may vibrate at different frequencies and we perceive those
vibrations as pitch. Not only does this help distinguish between male, female and children’s
voices; pitch also has linguistic functions (see section on suprasegmentals).

Nasal and oral airflow

After the air has passed through the larynx, it may flow into both oral and nasal cavities.
This is true of the sounds called nasal, as at the end of <whim–win–wing>. However, in most
speech sounds the nasal cavity is cut off by the raised soft palate (the velum), the soft tissue
Figure 3.1
The front, side and back views of the larynx (top) and a side view of the larynx, showing the location of the vocal folds (bottom).

Figure 3.2 The larynx and the vocal folds
at the back of the mouth. The velum acts like a valve, directing the airflow only through the oral cavity. Sounds produced in this way are called oral. The velum position is responsible for the differences between pairs of otherwise similar sounds such as the final consonants of <bid–bin>, shown in Figure 3.3.

Consonants: place of articulation

As the air progresses through the vocal tract, it may be further altered in a number of ways. Most consonant sounds involve an obstruction to the airflow through a contact of the parts of the vocal tract called articulators. Articulators are of two types: active, which move in speech production (tongue tip, tongue body, lower lip) and passive, which are stationary (teeth, hard palate, the wall of the pharynx). Consonants are produced along the entire length of the vocal tract, i.e. articulators may approach anywhere along its length. The area of the narrowest constriction is called the place of articulation (Figure 3.5), usually named after the passive articulator.

Consonants: manner of articulation

At each place of articulation, different degrees of constriction are described as differences in the manner of articulation (Figure 3.6). If the articulators touch, a complete closure is formed. Sounds formed in this manner are called stops, because the air is stopped from flowing freely. They can be oral, such as the initial sounds of <pie–buy>, or nasal, such as the one in <my>. In oral stops the velum is up, directing all air through the mouth and leading to an increase in the air pressure. When the articulators move rapidly apart, the compressed air is explosively released. This auditory effect is called plosion and oral stops are also called plosives. In nasal stops the lowered velum allows the air to escape through the nasal cavity.

With the articulators close together (narrow approximation), the air squeezing through the gap produces audible turbulence, producing fricatives, such as the initial sounds of...
<fat, thin, van, thick and hid>. The fricative /h/ is unusual because there is no narrow approximation. The breath friction results from the air passing through the vocal tract. A subset of fricatives where the air directed at the teeth produces a high-pitched loud noise is called sibilants, such as the initial sounds of <sin, zero, shin, Zsa Zsa>. In English, sibilants behave differently to other fricatives, i.e. the nouns ending in sibilants form plurals and the third-person present tense with the presence of a vowel in the suffix: compare <kiss, quiche> with <laugh, myth>.

When the air flows unhindered (the articulators in wide approximation), the resulting sounds are approximants, such as the initial sounds of <wed, led, read and yet>. The only source of sound is the vocal fold vibration amplified by the shape of the resonating cavity. Manner of articulation is not confined to vertical distance. In our four approximants other gestures contribute to their quality. Thus, /w/ has two simultaneous strictures (velar and bilabial); during /l/, the only lateral approximant in English, the air escapes over the sides of the vocal tract, not the midline and the post-alveolar stricture for /ɹ/ is accompanied by lip rounding.

Sounds involving articulator contact can also differ in the speed of the closure release, as in the initial consonants of the English [t] in <tongue> versus German [ts] in <Zunge> (stop versus affricate). Another difference is the duration of the contact (stop versus tap). For example, the middle consonant in the carefully spoken word <British> will have a stop. However, in fast speech the contact is shorter, producing a tap. The variant choice is related not only to speech tempo but also to the variety of English. American and Canadian accents will usually have a tap, while in British English both stop and tap are possible.

Manner of articulation can also involve gesture repetition. This difference is absent from English, but is responsible for meaning contrasts in other languages, exemplified by the Spanish.
<table>
<thead>
<tr>
<th>Articulators</th>
<th>Place of articulation</th>
<th>Examples (from MRP unless otherwise stated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive: upper incisors (front teeth)</td>
<td>LABIODENTAL</td>
<td>[f v] as in fife, viva</td>
</tr>
<tr>
<td>Active: lower lip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive: upper lip</td>
<td>LINGUOLABIAL</td>
<td>Tangoa [t] (Extremely rare sound)</td>
</tr>
<tr>
<td>Active: tongue tip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive: upper incisors</td>
<td>DENTAL</td>
<td>[θ ð] as in think, the</td>
</tr>
<tr>
<td>Active: tongue tip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive: alveolar ridge</td>
<td>ALVEOLAR</td>
<td>[n t d s z l] as in nine, tot, dad, sauce, zoos, loil</td>
</tr>
<tr>
<td>Active: tongue tip and/or blade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive: rear of the alveolar ridge</td>
<td>POSTALVEOLAR</td>
<td>[j] as in red roses</td>
</tr>
<tr>
<td>Active: tongue tip</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive: hard palate</td>
<td>RETROFLEX</td>
<td>[t̚] (GAm r-sound)</td>
</tr>
<tr>
<td>Active: tongue tip</td>
<td>(Note This rather impressionistic term is used instead of the literal place names such as: apico-palatal, sub-apico-palatal, etc.)</td>
<td></td>
</tr>
<tr>
<td>Passive: hard palate + rear of alveolar ridge</td>
<td>PALATOALVEOLAR (Sub-category of Postalveolar)</td>
<td>[ʃ ʒ ɹ ʒ] as in sheepish, pleasure, church, judge</td>
</tr>
<tr>
<td>Active: front tip/blade of tongue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive: front of hard palate + rear of alveolar ridge</td>
<td>ALVEOLOPALATAL (These sounds are represented under 'Other Symbols' on the IPA chart.)</td>
<td>Polish [c z], Serbian [ć c], Japanese [ʦ ʣ]</td>
</tr>
<tr>
<td>Active: front of tongue</td>
<td>PALATAL</td>
<td>[j] as in yes</td>
</tr>
<tr>
<td>Passive: hard palate</td>
<td>VELAR</td>
<td>[ɡ k ɡ] as in sing, kick, gag</td>
</tr>
<tr>
<td>Active: front of tongue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive: soft palate (velum)</td>
<td>UVULAR</td>
<td>French [x], German [ʁ]</td>
</tr>
<tr>
<td>Active: back of tongue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive: uvula (= extreme end of soft palate)</td>
<td>PHARYNGEAL</td>
<td>Arabic [ʕ] (the ayn)</td>
</tr>
<tr>
<td>Active: back of tongue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive: rear wall of pharynx</td>
<td>EPIGLOTTAL</td>
<td>Aquil []: (Extremely rare sounds)</td>
</tr>
<tr>
<td>Active: root of tongue</td>
<td>GLOTTAL</td>
<td>[h ʔ] as in ho’</td>
</tr>
<tr>
<td>Passive: arytenoid cartilages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active: epiglottis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The vocal folds</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.5 Places of articulation. English examples refer to received pronunciation. Examples of places of articulation not employed in English are from other languages.

‘cero’ ‘zero’ versus ‘cerro’ ‘hill’. The middle consonant in ‘cero’ is a tap, a quick contact of the tongue tip with the tooth ridge. If this contact is repeated in quick succession, this produces ‘cerro’ and the middle consonant is a trill. Both sounds also appear in some varieties of Scottish English (although the latter is now rare), where they do not contribute to meaning distinctions. We will learn more about this in the section on the linguistic status of sounds.

**Vowels**

Vowels are produced with the air flowing relatively freely, in contrast to consonants, where the airflow is usually impeded in some way (e.g. blocked in stops or turbulent in fricatives). Vowels are produced in the small area between the hard palate and the velum (Figure 3.7), using the body of the tongue, which lies directly below. Moving the tongue, lips and jaw changes the shape of the vocal tract. Those differences contribute to each vowel sounding different, known as vowel quality.

Vowel quality depends on three main factors: how high the tongue is raised towards the roof of the mouth, how far forward or back within the palato-velar area the highest point of the tongue is, and the lip shape. Some vowels have one quality as the exaggerated vowel in ‘cheese’ uttered posing for pictures, others change as a result of the tongue and lip movement. They are called diphthongs, as exemplified by both words in the phrase ‘nice boy’. In English their starting quality is much longer and slowly decreases in loudness.

In some languages vowels also differ in length. For example English has short and long vowels, the latter indicated by the [ː] diacritic. Vowel duration can be affected by their context in connected speech. Vowels can also be oral (airflow through the mouth) or nasal (airflow through mouth and nose), exemplified by the French pair ‘sait’ ‘knows’ versus ‘saint’ ‘saint’. Not all languages have nasal vowels.

L2 users of English often find it difficult to distinguish and produce its vowels. One reason for this is that inner circle Englishes, most frequently chosen as models, have rather complex
systems with respect to both vowel number and type. For example, American, southern hemisphere and most British accents (except Scottish English) have a contrast between the high vowels of \(<pool–pull>\), which differ both in quality and length. However, many languages have only one vowel in that region. A complicating factor is that the exposure to one pronunciation model in the classroom does not prepare the students for a real-life experience of regional accent differences. Ladefoged and Johnson (2015, p. 97) state that many Californian and Midwestern US speakers do not distinguish between the vowels of \(<cot–caught>\), while others make contrasts absent from General American (GA).

Another issue is that learners often fail to use reduced vowels, such as schwa, the initial sound in \(<above>\) which has a central quality and is short. Spoken inner circle Englishes are characterized by the occurrence of schwa in most unstressed syllables. L2 speakers have a tendency to use strong vowels in its place. This may be due to an influence of spelling, with learners believing that the letters \(<i, e, o, a, u>\) correspond to full vowels, or simply to a lack of awareness that reduced vowels exist.

**Representing speech: the IPA chart and transcription**

The International Phonetic Alphabet (IPA) is a notational system used to represent sounds of the world’s languages, based largely on the Roman alphabet, with additional symbols. This is necessary, as the number of alphabet letters is far smaller than a variety of sounds in human speech. Any utterance can be represented by a sequence of symbols. This way of representing speech is called transcription and is used to convey information about the sounds that have been spoken. There are different kinds of transcription, depending on the degree and kind of detail needed.

While useful, the IPA is a somewhat simplified way of representing spoken language, involving several underlying principles. Primarily, it operates on what is termed segments, that is, an assumption that the flow of speech ‘can be represented partly as a sequence of discrete sounds or “segments”’ (Handbook of the International Phonetic Association, 1999, p. 3). They are divided into two categories: vowels and consonants. Each segment is assigned a unique phonetic symbol and referred to by mostly three-part labels describing its articulatory configurations. For consonants, these are voice, place and manner. For example, \(/p/\) in \(<pin>\) is a voiceless bilabial stop, while \(/j/\) in \(<yet>\) a voiced palatal approximant. For vowels, there is also a three-way label, representing its articulatory properties, such as tongue height (high–mid–low), frontness–backness (front–central–back) and lip shape (rounded–neutral–spread). For example, our \(<cheese>\) vowel is high, front and spread. Thus, the division into segments serves to illustrate information, which may be linguistically relevant, but the IPA notation is also used to represent linguistically relevant features of speech above and below a segment. Aspects of speech, such as voice quality, tempo or emotion are not represented by the IPA. A full IPA chart is attached as an appendix to this chapter.

We have said that the IPA uses segments as the basis for sound classification. Indeed, the concept of linear phonetic units that are relatively unchanging (Laver 1994, pp. 112–113) is a convenient way of thinking about sounds and representing their key properties. Additionally, segments have a certain psychological reality for language users brought up with writing systems. However, bear in mind that each sound, even if represented by a single symbol, in fact encompasses a number of dynamic vocal tract configurations. This happens because its production involves a number of articulators, all moving independently. For example the production of the voiceless bilabial plosive \(/p/\) involves the contact of the lips, which causes a blockage in the vocal tract, the build-up of air pressure and finally the rapid parting of the
lips and an explosive release. While the mouth is adopting this setting, the vocal folds are apart and the velum blocks the passage to the nasal cavity. This is what would happen if /p/ were pronounced in isolation, which is rarely the case. We need to remember that speech is produced by constant transitions between sounds. It is therefore convenient to think of the IPA labels and symbols as a way of capturing one of the stages of a sound production, in the same way that a photo captures a snapshot of activity. A photograph is informative of what occurred and captures the spirit of the moment but does not tell the whole story.

Another common misconception regards the transcription system. The type of transcription found in dictionaries, textbooks and language teaching materials is broad or phonemic. Its purpose is to provide a framework for studying a language, indicating the type and number of sounds it has. A dictionary entry clarifies what sounds occur in a word, in what order and which syllable is stressed. For example, the transcription of <boots> /buːts/ shows that the word is composed of four sounds. It starts with a b-type sound, has an u-type vowel in the middle and ends with two voiceless consonants. However, this representation does not provide us with information about sound properties, i.e. it does not tell us when or if the vocal folds start vibrating during the first consonant, whether the /u/ has a back or a central quality and whether the /t/ is dental, alveolar or pre-glottalized.

Therefore, it contains little information how a word would be uttered by a native speaker. This needs to be taught from audio materials and by a teacher who is aware of those details and able to demonstrate them. It is of course possible to include this detailed information in transcription, through the use of diacritics and symbol choice – the word <boots> in narrow transcription might be represented as [bʊːts]. This type of representation is called narrow or phonetic and is often used when collecting field data or working in a speech therapy clinic. However, most of the time it would be impractical to include this level of detail in language instruction materials. In other words, teaching resources employ the IPA to indicate phonemic representation, which does not show phonetic features and is not sufficient to teach pronunciation.

The linguistic status of sounds: contrastiveness, phonemes and allophones

This section looks at two kinds of phonetic differences and their role in communication. In the second section we saw that some physical features of sounds contribute to distinguishing between words, e.g. voicing <seal–zeal>, nasality <bid–bin> or tongue height <pool–pull>. Let us now examine another type of physical difference. For example, comparing the words <late> and <plate> in Southern British English we can hear that the /l/ sound in <late> is voiced, while the one in <plate> sounds voiceless and has slight friction. We can represent it as [ɬ]. We can find more examples of l-type sounds in the same variety of English, e.g. in <hell> and <health>. We can observe that the tongue tip has a different position, touching the alveolar ridge in <hell> and the teeth in <health> because of the adjacent dental fricative. We refer to the place of articulation of the latter as dental and represent it as [ɹ]. Now we can make some more observations. Comparing the /l/ in <late> with the one in <hell>, we also note that they sound different. This is because in the latter the back of the tongue moves towards the velum, resulting in a more constricted vocal tract. We hear this quality as a ‘dark’ resonance and represent it by a tilde written across the mid-section of the symbol [ɬ]. Many speakers do not have a dark [ɬ] but a back vowel of an [o] quality in those words.

So we see that the /l/ sounds [ɬ][ɬ][ɬ][ɬ] in <late, plate, hell, health> are produced differently, yet a native speaker of Southern British English intuitively recognizes them as ‘the
same’, despite their varied physical characteristics. This is because of their linguistic status. They do not change meaning. They are members of one phoneme, which is a linguistic unit of contrast. In other words, a phoneme is how native speakers conceptualize which physical differences between sounds are key to signalling differences in meaning and which are not. The different variants of a phoneme are called allophones. We can represent the status of sounds by the use of brackets. A symbol in slant brackets indicates a phoneme /l/ and square brackets indicate its variants (allophones) [l][l][l][l].

Native speakers’ intuitions are not the only factor in deciding on the linguistic status of sounds. Grouping them as allophones of the same phoneme or belonging to different phonemes, we primarily take into account their contrastiveness. Consider again the /l/ sound in <late>. As an experiment, we could substitute [l] for other sounds from the /l/ family: [l], [L] or [L]. As a result, the word sounds different, as if uttered by a speaker of another variety of English or an L2 user, but it is still recognized as the same lexical item.

The way we group sounds into families is therefore more dependent on their role in communication, rather than their physical similarities. This intuitive knowledge forms part of our language competence and tends to be taken for granted. On the one hand, unless linguistically trained, we tend not to pay attention to those physical differences that cause no change in meaning. On the other, we assume that what is contrastive in our mother tongue is also contrastive in other languages. An example is the alveolar lateral approximant [l] and post-alveolar approximant [ɾ], the difference being contrastive in English, as in <late–rate> or <play–pray>. Japanese also has [I] and [ɾ], yet they are variants of one phoneme, with [I] in all positions and [ɾ] occasionally present (Okada, 1999). Thus, like a native English speaker, a Japanese native speaker has experience of using both sounds and is capable of pronouncing both. The reason they might be unable to hear or produce the difference the English <late> and <rate> difference in a controlled manner is then not a result of their unfamiliarity with those sounds but their status in their L1. Fraser (2001, p. 20) skilfully illustrates this well-known difficulty using an example of ‘two Australian friends, Alison and Bronwyn, travelling in Japan. They found themselves with new names: Arison and Blonwyn!’ She advises that teaching the articulation of the sounds that the students can already make perfectly well is unhelpful as the difficulty lies in conceptualizing them differently than in their L1, i.e. in learning to discriminate and organize them in the mind. Thus the teaching of underlying concepts needs to precede production and perception work. There is evidence that high-variability identification training improves perceptual learning (Bradlow, Pisoni, Akahane-Yamada & Tohkura 1997). For classroom practice, this means designing exercises that expose learners to a large amount of tokens (preferably in a variety of accents) containing the problematic sounds and asking students to identify them.

Syllables and phonotactics

Speech is organized on many levels. So far we have looked at the segmental level but sounds are also grouped into syllables. In a very basic sense, a syllable is the smallest unit of speech because every utterance contains at least one. It is well recognized that ‘there is no agreed phonetic definition of a syllable’ (Ladefoged & Johnson, 2015, p. 253). Simultaneously, despite problems in formally defining the concept, it appears to be an important unit of structure. There are several reasons for this. One of them is native speakers’ intuitions. If asked to divide a word into syllables, any native (adult or child) user of a language will be able to do so without linguistic training. Second, the construct of the syllable constitutes an important
domain for phonological regularities and someallophones are conditioned by their place within a syllable. A syllable is divided into two main parts. The universally adopted convention terms the first part an onset and the second a rhyme, in turn composed of a nucleus and a coda (Gussenhoven & Jacobs, 2011). The basic syllable template is shown in Figure 3.8.

There are some universal properties of syllables. It must contain one loud or prominent part, which is almost always a vowel sound. There may optionally be consonants preceding or following the vowel. Thus, consonants are usually found at the syllable edges and with a vowel nucleus in their core. The most common syllable type across languages is CV (consonant–vowel). The exact syllable composition is language-specific.

Phonotactics is the study of the permitted syllable structures, i.e. combinations of sounds within a language. It specifies which consonants can form clusters, how many members a cluster may have and where in the syllable they are allowed to occur. For example, languages differ with respect to the type of clusters in the onset. Some allow only simple ones or none at all.

Let us look at English and Polish. Both allow a *pt- cluster, as in English <opt> or <stopped> and Polish <deptak> ‘promenade’ and <ptak> ‘bird’. An English speaker learning Polish will have no difficulty pronouncing the former word but the latter will usually be problematic. They will either reduce the cluster, saying /tak/ or split it with a vowel, producing /pətak/, thus creating two simpler syllables. The reason for the adoption of those strategies lies in the phonotactic difference between the languages. The *pt- in <ptak> occurs in a syllable-initial position, disallowed in English. English has no syllables (or indeed words) beginning with a *pt-, which is what causes the production difficulty. Pronouncing <deptak> the NES can split the cluster between the two syllables and therefore it is less of an issue when trying to pronounce it. Both languages allow this cluster in syllable-final position, hence an English speaker should have no problems producing Polish words such as <szept> ‘whisper’.

Another example is the English word <sport>, which may be pronounced [espɔ:t] by Spanish-speaking learners. This will contribute to a perception of a foreign accent but is likely comprehensible. However, inserting vowels to break up a cluster may also result in complete misunderstandings. Thus, when uttered by a Japanese speaker, <sport> may become [suptɔ:t], and be interpreted as <support>. Coda consonants and clusters are also difficult for many L2 users of English, since many languages permit few or no coda consonants while English permits up to four in this position, as in <strengths>. In contrast, Cantonese allows only a single consonant in the coda and this can only be a stop or a nasal.
Suprasegmentals

In previous sections we looked at segments and how they are organized into syllables. Many features of speech extend beyond the segment, spreading over a syllable, a word or a longer utterance. They are called suprasegmental or prosodic and, like segments, they contribute to the meaning and organization of the spoken discourse. Suprasegmental aspects of language can be analysed, systematically described and taught to non-native users, although in practice some are more difficult than others. For example, according to Pennington and Ellis (2000), nucleus placement appears to be learnable but pitch movement and the intonation of question tags are not.

Stress

The term stress may be used as a theoretical construct or a phonetic property of a syllable. A stressed syllable is more prominent than the adjacent syllables. In terms of its phonetic manifestation, a stressed syllable may be louder and longer and have a more peripheral vowel quality. Most stressed syllables involve some combination of those features. Stress is a relative property and it does not make sense to describe an isolated syllable as stressed or unstressed. Stress may be contrastive as in English noun–verb pairs, such as <an invite> and <to invite>. Both have the same segment sequence, represented as /ɪnvæt/, but the difference lies in stress placement; on the first syllable for the noun /ˈɪnvaɪt/ and second for the verb /ɪnˈvaɪt/. In many dictionaries, the location of stress is marked by a raised diacritic before the syllable.

We distinguish two kind of stress, lexical and rhythmic. The former is a property of an individual word, i.e. the potential of a syllable to become more prominent, as in the example above, where either syllable can attract stress depending on the word’s lexical category. This is called the citation form, the way the word is pronounced in isolation. In connected speech not all potential stresses are realized. If either of our <invite> words were part of a longer utterance the syllable stressed in isolation would likely lose prominence, as in:

I missed the party because rather than tell me in good time, they sent me a LATE invite.

Rhythmic stress is therefore the degree of syllable prominence within a longer stretch of speech and depends on the structure and intended meaning of the utterance. Stressed syllables act as rhythmic beats in speech. English belongs to a group of languages with variable lexical stress, but a description of rules is beyond the scope of this chapter. Languages such as Czech, Polish, French or Welsh have fixed lexical stress. Thus stress in Polish always falls on the penultimate syllable of the word and in French on the last one.

(Pitch) accent

In addition to increased length, loudness and strong vowel quality, some stressed syllables also involve a pitch movement. Syllables belonging to this subset are called accents (also pitch accents). Accented syllables are pitch-prominent, i.e. they differ in pitch from the preceding syllable. They can jump up or down in pitch or initiate a pitch movement. Their prominence marks important turning points in intonational tunes.
Intonation

In the flow of utterances our voices continually go up and down. This is true even of the speakers we call monotonous. Those changes are perceived as rises and falls in pitch. The articulatory feature responsible for pitch variation is the rate of vibration of the vocal folds, called the fundamental frequency of speech. It refers to the number of vibratory cycles per second and is measured in hertz. For example, if the vocal folds vibrate a hundred times per second, the fundamental frequency is 100 Hz. The faster the vibrations the higher the perceived pitch. Male voices in conversational speech have on average a lower fundamental (50–250 Hz) than female ones (120–480 Hz), which in turn are lower-pitched than children’s voices (Laver 1994, p. 451). Every speaker has a pitch range, i.e. a scale of pitches they are able to produce. Within their range they also perform pitch movements. When we talk about high or low pitch we do not refer to an absolute frequency value but describe where a given syllable is placed within a speaker’s range in relation to other syllables.

Languages use pitch variations to signal meaning differences. For example, <John> uttered with a falling pitch is a statement, but spoken with a rising pitch a question. The same is true of longer utterances: <He’s out> spoken on a fall is a comment on the state of affairs, while if a pitch rises listeners interpret this as a question. Thus, like other aspects of speech, the direction of pitch change may be contrastive, i.e. indicate the grammatical category difference between a declarative and an interrogative. Not every pitch change is linguistically significant.

The parameters of intonation

A common misconception has it that intonation is hard if not impossible to pin down and consequently unteachable. This myth is partly fuelled by another incorrect belief that while segments (vowels and consonants) have clear boundaries, the continuous fluctuations of the pitch are somewhat intangible and cannot be captured within any systematic framework. However, intonation is used in a number of ways to clarify meaning in context. It can in fact be divided into components and described in a systematic manner. This section deals with the framework used to describe intonation within the British School, as applied in Wells (2006). The main variables are the ‘three Ts’: tonality, tonicity and tone. Those key aspects of intonation play different roles in signalling meaning and each of them makes an independent contribution.

Tonality

One of the functions of intonation is to divide speech into smaller, manageable chunks of information for the listeners to process. Speakers make choices about tonality, i.e. how an utterance can be divided into meaningful, coherent units. These are called intonation phrases (IPs) (also thought groups) and are indicated by ‘|’. In most cases, speakers are free to make choices about the number of IPs they use. An IP may be a complete sentence, a phrase or even a single word. A story can be told with a different number of IPs:

on Saturday | I woke up and didn’t know what to do | so I rang my best mate | to see if she wanted to come out | but she was like | oh I don’t know what to wear | and said help me choose my outfit | in the end by the time we got into town we were so hungry | that we skipped the shops | and headed for a pizza |
or

on Saturday | I woke up | and didn’t know what to do | so I rang my best mate | to see | if she wanted to come out | but she was like | oh | I don’t know what to wear | and said help me choose my outfit | in the end | by the time we got into town | we were so hungry | that we skipped the shops | and headed for a pizza |

Division into IPs can also be used to express differences in meaning. The following sentence, uttered as one IP:

I didn’t invite him because of Jessica

means that Jessica was not the reason I invited him. But, divided into two IPs, as in:

I didn’t invite him | because of Jessica

means that Jessica was the reason I did not invite him.

Tonicity

Another function of intonation is highlighting the parts of an utterance the speaker considers the most important. Thus, after the division of an utterance into IPs, the next decision regards its focus. Phonetically, this is manifested as the placement of accented syllables within each IP and one of them becoming its nucleus or the tonic syllable. The nucleus is the last of the many pitch changes within an IP, but not the most prominent one as a widely held misconception has it. Nucleus placement (tonicity) reflects the focus of an utterance. There are two main types of focus: broad (neutral) and narrow.

In broad focus, when the whole utterance constitutes new information, the default position for nucleus is the last content word in the group, illustrated in (a). The nuclear syllable is underlined.

(a) I’ve finally finished my homework

If the utterance focus is narrow, this is reflected in the placement of the nucleus on the stressed syllable of the word that constitutes new information, as in (b–e). Here, the fact that homework is involved is already known to the interlocutors and so the focus is narrow.

(b) I’ve finally finished my homework
(c) I’ve finally finished my homework
(d) I’ve finally finished my homework
(e) I’ve finally finished my homework

Tone

We have seen that the nucleus is the last change of pitch in an IP. This pitch movement is called the nuclear tone. Tones are symbolized by the use of tonetic marks (shown below). English uses a number of tones and the most basic distinction (Wells, 2006) is based on the direction of pitch movement: falling and non-falling tones. The latter are further subdivided into rise, fall–rise and mid-level. Fall and rise tones are also divided into high and low, i.e.
high fall, low fall, high rise and low rise. This is based on what distance across the speaker’s pitch range the pitch movement covers.

<table>
<thead>
<tr>
<th>Falling</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High fall</td>
<td>Right</td>
<td></td>
</tr>
<tr>
<td>Low fall</td>
<td>Right</td>
<td></td>
</tr>
<tr>
<td>Rise–fall</td>
<td>Right</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-falling</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High rise</td>
<td>Right</td>
<td></td>
</tr>
<tr>
<td>Low rise</td>
<td>Right</td>
<td></td>
</tr>
<tr>
<td>Fall–rise</td>
<td>Right</td>
<td></td>
</tr>
<tr>
<td>Mid-level</td>
<td>Right</td>
<td></td>
</tr>
</tbody>
</table>

Broadly speaking, the choice of tone is related on the one hand to an utterance type (statement, question); on the other it helps organize discourse, signalling whether information is new or already known. Tone choice may have an attitudinal function, but its use is also motivated by other factors.

Regarding utterance types, Wells (2006, p. 15) points out that ‘a popular idea among language students is that statements are said with a fall, questions with a rise’. Indeed, statements more often than not have a fall, but questions differ. The basic principle is:

(a) falling tones: statements, commands and wh-questions:
   - What do you want to drink?
   - I’d like a glass of red.

(b) rising tones: yes/no questions:
   - Do you want a drink?

The choice of tone also helps organize discourse, i.e. it signals either continuity or completeness. Continuity is associated with the rising and level tones. Completeness (finality) is associated with falling tones. Consider the following example, where the rising tone on the final noun <vodka> indicates that the list is going to go on.

Do you want wine or beer or gin or vodka?

Compare this with the fall on the first syllable of <vodka> below, which suggests that the list of drinks is complete. The host does not have any other beverages.

Do you want wine or beer or gin or vodka?

Third, the choice of tone reinforces the status of information in an utterance, i.e. whether it is new or presupposed by the speakers. Falling tones indicate new information. A fall–rise suggests shared information as in the example below. Note that the order of the phrases does not contribute to the meaning, but the choice of tone does:

I’m going on holiday when I’ve finished all that marking
When are you going on holiday?

I’m going on holiday when I’ve finished all that marking
What are you going to do when you’ve finished all that marking?
Like other aspects of a language, intonation is subject to L1 interference. Thus, learners will transfer their native patterns to the language they are trying to master. However, lay perceptions of suprasegmentals and segments differ. We readily acknowledge that it takes time to learn L2 vowels and consonants both in terms of phonemic distinctions and the qualities. In contrast, the mastery of intonation is frequently taken for granted. Such views might spring from the fact that intonation has (among others) an attitudinal function and a common assumption is that its learning comes naturally.

Consequently, when an advanced learner uses intonation in a non-native manner, listeners frequently fail to make allowances for the discrepancies between their speech and the expectation. The NNS might be deemed impolite when no such intention existed. For example, the neutral tone in English declaratives is a high fall, while a low fall signals boredom or detachment. However, this would be the default choice in Polish for an unmarked statement. Imagine saying "I’m fine" with a low and a high fall in response to a greeting and think of the impression it gives the interlocutor.

Model choice

Teaching a language necessitates selecting a spoken variety as a model. In the EFL context, this is standard English (the dialect used in media, education and other public life domains), possibly allowing for some regional grammatical, morphological or semantic traits, depending on where in the world the instruction is taking place. The accent of choice is an educated variety. In the Americas and Asia this is usually General American, a broadcast news network accent. In Europe this is primarily received pronunciation (RP), an accent commonly heard in the BBC news broadcasts. Artificially derived varieties such as lingua franca core have been proposed as more suitable alternatives when English is used for communication between non-native speakers.

An accent choice may be open to objections such as that very few people speak it (as is the case for RP) or that it is hard to delimit (for problems in defining GA, see Preston, 2008). However, we have to ‘teach something after all’ (Trudgill, 2001) and the extensively described RP and GA are a practical solution. Second, we need to remember that our choice serves essentially as a reference variety. Primarily, we teach a phonemic framework. In other words, the model serves to point out the type of distinctions the students might not have in their L1, e.g. the RP vowels /æ – ʌ – ɑ:/ contributing to the <cat–cut–cart> contrast. This is notoriously hard to grasp for beginners as many languages have one low vowel where English has three. Indeed, while teaching the framework, we also demonstrate and teach a lot of phonetic detail.

However, phonetic accuracy is usually neither the main focus of instruction nor required for successful communication. In the classroom we do not teach what is in lay terms called ‘an accent’ and should therefore not be afraid of passing on an undesirable one. A native-like pronunciation is never achieved by most L2 learners (owing to factors such as motivation, learning goals, age and individual ability). Indeed, a handful (mainly language professionals) accomplish this feat but it requires exceptional motivation and an intensive input (Bongaerts, van Summeren, Planken & Schils, 1997). It is therefore important to take into account the immediate and long-term needs of our learners and set realistic goals. A model is a vehicle to teach syllable structure, connected speech phenomena and aspects of intonation crucial for intelligibility. If we do find time to focus on phonetic detail, a bigger concern is staying current with empirical research that signals instructional priorities (for a summary see Derwing & Munro, 2015) and keeping up to date with recent developments in English speech.
Future directions

Familiarity with fundamental phonetic concepts will assist teaching professionals in better understanding of learners’ problems and facilitate the choice of the most effective teaching techniques and activities. In addition to journal articles there are further professional development opportunities such as conference attendance; two major events on both sides of the Atlantic are the annual Pronunciation in Second Language Learning and Teaching Conference and the biennial Phonetics Teaching and Learning Conference. The *Summer Course in English Phonetics*, held annually in London, has groups specifically aimed at EFL professionals. Teachers might also consider taking the IPA Examination for the Certificate of Proficiency in the Phonetics of English, a formal non-degree certification of phonetic competence recognized by employers. IATEFL has a Pronunciation Special Interest Group, which has its own journal and a range of events. There are numerous online phonetics resources and blogs, compiled on the IPA website. The links are provided below:

https://speechlab.utah.edu/PSSLT2017.php  
www.ucl.ac.uk/pals/study/cpd/cpd-courses/ptlc/  
www.ucl.ac.uk/pals/study/cpd/cpd-courses/scep  
www.internationalphoneticassociation.org/content/ipa-exam  
http://pronsig.iatefl.org/index.html  
www.internationalphoneticassociation.org/content/links

Further reading


References


55


Appendix to Chapter 3

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2015)

<table>
<thead>
<tr>
<th>CONSONANTS (PULMONIC)</th>
<th>© 2015 IPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilabial</td>
<td>Labiodental</td>
</tr>
<tr>
<td>Passive</td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td></td>
</tr>
<tr>
<td>Tap or Flap</td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td></td>
</tr>
</tbody>
</table>

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

<table>
<thead>
<tr>
<th>CONSONANTS (NON-PULMONIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clicks</td>
</tr>
<tr>
<td>O Bilabial</td>
</tr>
<tr>
<td>Dorsal</td>
</tr>
<tr>
<td>(Postalveolar)</td>
</tr>
<tr>
<td>Palaoralveolar</td>
</tr>
<tr>
<td>Alveolar lateral</td>
</tr>
</tbody>
</table>

OTHER SYMBOLS
- Voiceless labio-velar fricative
- Voiced labio-velar approximant
- Voiced labio-palatal fricative
- Voiced labio-velar approximant
- Voiceless epiglottal fricative
- Voiceless epiglottal fricative
- Epiglottal plosive

DIACRITICS Some diacritics may be placed above a symbol with a descender, e.g. ą.
- Voicesss
- Voiced
- Aspirated
- More rounded
- Less rounded
- Advanced
- Retracted
- Centralized
- Mid-centralized
- Syllabic
- Non-syllabic
- Rhoticity

VOWELS
- Close
- Close-mid
- Open-mid
- Open

SUPRASEGMENTALS
- Primary stress
- Secondary stress
- Long
- Half-long
- Extra-short
- Minor (foot) group
- Major (intonation) group
- Syllabic break
- Linking (absence of a break)

TONES AND WORD ACCENTS
- Level
- Contour
- Extra high
- Extra low
- High
- Low
- Rising
- Falling
- Downstep
- Upstep

57