The Sounds of Language

These lectures are based on the following text, by kind permission of the author, who was Professor of Phonetics at Reading until retirement:

THE SOUNDS OF LANGUAGE 1: THE SCIENCE OF SPEECH

Speaking to each other is one of the most interesting things that we human beings do. Each of us has a mind, a private world filled with thoughts, feelings and memories. We have many ways of communicating these in such a way that they enter the minds of other people. Sometimes we find it convenient to communicate by means of writing, and good writing can let us see things clearly from the writer's own perspective. For people who are for some reason unable to speak, it is also possible to communicate by sign language, or by using a pointer and a computer screen. Many art-forms work by conveying the thoughts and feelings of the artist – music, for example, can tell us a great deal about the inner feelings of a composer, even one who has been dead for centuries.

A quite different form of communication is one that we share with many other animals: the use of gestures and facial expressions. We make extensive use of these, and can describe in great detail how people do so: we can talk about someone “waving his hand dismissively”, or “giving someone an appealing look”, or “turning away in mock anger”. But although there any many different ways of communicating, when it comes to telling other people what we want to tell them, what we use most is speech, and this is something which is only available to human beings.

The speech chain
To describe the process of speaking in the simplest way, we need to look at three main events. To begin with, we produce sounds, using parts of our chest, throat and head. Then the sounds travel through the air in the form of vibrations. Finally, the sounds are received by the ear of the listener. We can show this in diagram form like this (taken from Denes & Pinson 1993: 5):

However, this is only part of the story. If we look at speech more carefully, we find we must also take into account the fact that the brain of the speaker is involved in controlling the production of speech, and the brain of the listener has to do the job of analysing the sounds that have been heard and converting them
into a meaningful message. You might say of someone, in a joking way, that they were speaking without first connecting their brain, or that what was said to them went “in one ear and out of the other”, but in reality the control by the brain is essential. Until recently, we knew little about what was going on in the brain when people are speaking, and this is why the science of phonetics has concentrated on the three central components of the Speech Chain, where observation of what is going on is relatively straightforward. However, our understanding of how the brain works in speech communication has grown enormously in recent years. One of the most significant advances in recent research has been the development of safe and accurate brain scanning techniques that can show us the activity of different parts of the brain when someone is speaking or listening to speech.

Phonetics – the study of speech production
Speech is a complicated process, and to study it requires a whole scientific subject – the science of phonetics. In this part of the module, we will look at the main areas of phonetics and see why they are important.

Many of the early lectures are taken up with considering the way in which speech sounds (often called segments) are made, and how phoneticians can classify them in a scientific way. This is based on the fundamental distinction between vowels and consonants.

Another fundamental aspect of the subject is the use of symbols. In phonetics, we must be able to use a particular symbol to represent a particular sound. This is quite similar to the principle of alphabetic writing: some writing systems give a very clear indication of the sounds (for example, the writing systems of Finnish and Italian represent almost perfectly the sequence of sounds required to say something in those languages); at the other extreme, the ideographic writing system used for Chinese and some other languages represents ideas, not sounds. Many symbols are common to Chinese and Japanese (“kanji”). They convey the same meaning, but they would be pronounced in completely different ways by a speaker of Mandarin Chinese, of Cantonese and of Japanese. The nearest equivalent for users of alphabetic writing is our number system: the numbers 1, 2, 3 mean the same thing to speakers of Russian, or French or of English, yet they would pronounce them in completely different ways.

One of the most important achievements of phonetics in the past century has been to arrive at a system of phonetic symbols that anyone can learn to use and that can be used to represent the sounds of any language. This is the International Phonetic Alphabet (IPA). Taking English as an example of a writing system that does not always give a reliable guide to pronunciation, we find that for various purposes (including teaching the pronunciation) it is helpful to use phonetic transcription instead of ordinary spelling. Many of the symbols we use are the same as the familiar alphabetic ones. The list given below shows the symbols used to represent one accent of English, and the symbols are given with “key words”, drawn from Wells’s Standard Lexical Sets (1982) in the case of vowels, which help you to see which sound is represented.

For each of the many different accents of English, a slightly different set of symbols might be needed. The “standard accent” of English described in these lectures is similar to one which for much of the twentieth century was known by the name “Received Pronunciation” (RP for short); some modern writers on phonetics now prefer to use the name “BBC pronunciation” or “General British”. This is not to claim that all speakers on the BBC or all British speakers
have the same accent – an increasing number of professional broadcasters now have a variety of accents – but it is still possible to identify a reasonably consistent pronunciation used by English-born announcers and newsreaders on channels such as Radio 3 and Radio 4, BBC television news and the World Service of the BBC. One advantage of this choice is that anyone with a radio can listen to BBC broadcasts as much as they like; tracking down the elusive RP speaker has always been a problem for researchers, since nobody could ever agree on exactly what they should be looking for.

VOWELS
Short: /ɪ/ kit; /e/ dress; /æ/ trap; /ʌ/ strut; /ʊ/ lot; /u/ foot
Long: /iː/ fleece; /aː/ palm; /ɜː/ bird; /ɔː/ thought; /uː/ goose
Diphthongs: /au/ goat; /au/ mouth; /ai/ price; /ei/ face; /ɔɪ/ choice;
/ɪə/ near; /eə/ square; /ʊə/ cure
Weak: /ə/ letter, comma; /i/ happy, she; /u/ thank you

CONSONANTS
/p/ pin; /t/ tin; /k/ kin; /b/ bin; /d/ din; /g/ give;
/f/ fin; /θ/ think; /s/ sing; /ʃ/ shin; /h/ hymn;
/v/ van; /ð/ this; /z/ zoo; /ʒ/ measure; /ʃ/ chin; /dʒ/ gin;
/m/ more; /n/ no; /ŋ/ thing; /l/ low; /r/ red; /w/ wet; /j/ yet

Using these symbols, we can write English in a way that tells you exactly which sounds are pronounced. Notice how different this looks from ordinary English spelling in the example sentence:

Spelling: She bought some chairs and a table
Transcription: ʃɪ ˈbɔːt ʃɑm ˈteɪz ən ə teɪbl

Phonetics has links with many other subjects: when we look at the parts of the body which are used in speaking, we make use of information from anatomy and physiology; the study of the transmission of speech sounds from speaker to hearer is related to acoustics, which is a branch of physics; and when we look at how we hear and understand spoken messages, we are in the territory of other subjects including audiology (the study of the hearing mechanism) and cognitive psychology (the study of how the brain receives and processes information).

Phonology – the study of sound patterns in languages
We should not forget that the whole science of phonetics is an essential part of the subject of linguistics. When we look at the subject from this point of view, we need to understand the basic principle of contrast in language: in phonetics in general, we can identify an enormous variety of different sounds that humans can make, but only a small number of these are used a particular language to make a difference between words. In French the words tout (which we write in phonetic symbols as [tu]) and tu [phonetically [ty]] are recognisably different because of their vowels. In English the word two is phonetically [tu:], which is between [u] and [y]; if we were to substitute the vowel [y], this would not result in an English listener hearing a different English word. This is because the set of distinctive sounds, known as phonemes, is different in French and English. Each language
has a different phonology from all the others, although individual speech sounds may be shared.

Linguistics studies all aspects of human language. In some cases (syntax, for example) it deals with complex and abstract concepts, and the mental processes involved can only be guessed at, not observed or measured. In other areas of linguistics, such as the study of social or regional variation in language, the data and its patterns may be easier to grasp. But, unless we are studying written language, it is only through analysing spoken language that we are able to collect the data we use in linguistic research. The great British scholar Henry Sweet described phonetics as “the indispensable foundation” for the study of language, and this view is as valid today as it was a hundred years ago.

What to read
Chapter 4 (“The Sound System of Language”) gives a simple introduction to phonetics and phonology, and looks ahead to a number of topics that are introduced in later lectures.

The book as a whole contains much useful reading to supplement several chapters in the present book, though you may find some of the details of the physics, anatomy and physiology of speech difficult. The examples are based on American English.

This book presents the Standard Lexical Sets for English vowels which have been adopted by the international socio-phonetic community. For example, we talk about “the FOOT/STRUT split” to indicate that speakers from the south of England tend to have a different vowel sounds in those words (/fʊt/ vs. /strʌt/), whereas speakers from the north of England do not (pronounced /fʊt/ and /strʊt/ with the same vowel phoneme), and “DRESS-raising” to indicate that a New Zealander’s DRESS vowel is produced with the tongue raised higher in the mouth than in many other varieties.
THE SOUNDS OF LANGUAGE 2: MAKING SPEECH SOUNDS

Speech and Breathing
In the languages of the world there is a huge and fascinating variety of different speech sounds, and we need to understand how these speech sounds are made. The most basic fact to remember is that all the sounds we use in speech are produced by moving air. Speech is sometimes described as “modified breathing”: the process of speaking begins, except for a relatively small number of cases, with the air inside the speaker’s chest escaping from the lungs, through the throat and the mouth, and out into the open air. If we produce this flow of air without impeding it in any way, the activity would just be what we call “breathing out”, or, if we do it loudly enough to make a sound, a “sigh” (which can convey many different meanings). But usually in speaking we use our articulators to modify the flow of air so that sounds are produced. It is a very interesting fact about speech that it has evolved by making use of parts of the body which already have some biological reason for being there (breathing, biting and chewing food, swallowing, getting attention, listening) and, as far as we know, there is nothing in the human body which exists exclusively for making or recognising speech sounds. We would still need our lungs, our tongues, our vocal folds and our ears, even if they were not required for speech.

There are, in fact, some speech sounds which are made by using something other than the lungs to make the air move. To English speakers, many of these sounds are familiar but “non-linguistic”; they are not used as phonemes of the language. For example, “click” sounds are found when we make the “annoyance” noise that used to be written as “tut-tut” (or nowadays more usually as “tsk-tsk”), and people still sometimes use a “gee-up” click to tell a horse to move on. But in a number of languages of southern Africa, such as Zulu and Xhosa, we find a wide variety of click sounds being used as consonant phonemes of the language. There are also sounds known as “ejectives” and “implosives” in many of the world’s languages, and to produce these you have use the larynx: the vocal folds are closed or nearly closed, and the larynx is moved upwards (in ejectives) to push air out, or downwards (in implosives) to suck air in. These sounds are found in many different languages, including some accents of English.

The passageway through which air passes from the larynx, past the lips and out into the air outside our bodies is called the vocal tract. Below this is the trachea, the “windpipe”, which is connected to the lungs and which passes from the chest into the neck. We do not have the ability to move or modify this.

The larynx
The larynx is very important. The vital part of the larynx is a pair of folds of muscular tissue called the vocal folds, and we can move these into a number of positions between wide open and tightly closed. We open them widely to allow a rapid escape of air. If they are slightly narrowed, so that the gap between them is only a few millimetres, the air makes a rushing noise that we associate with the sound at the beginning of the English word head. If we close them enough for them to be lightly touching each other, the air passing between them causes them to vibrate; this is called voicing, or phonation, and it can be varied in many ways, most importantly in pitch, which may be high or low according to how we adjust our vocal folds. Many speech sounds are voiced, while others are voiceless. If you want to practise detecting the difference, compare [s] and [z]: the consonant [s] is voiceless and the only sound it makes is the hissing sound as air escapes over the tongue. However, its voiced counterpart [z] not only has this hissing sound,
but also the buzzing of the vocal folds vibrating. You can hear this best if you put your fingers in your ears and produce a long, buzzing [z] - you can also detect the vibration if you press your fingers gently against your larynx as you produce the [z]. Vowels are almost always voiced, as are nasal consonants like [m] and [n]. Finally, if we close the vocal folds firmly, we prevent air from escaping. This is often called a glottal stop.

The vocal tract above the larynx

Immediately above the larynx is a passageway called the pharynx. This carries air, but also food when we are eating. As we all occasionally find when we are trying to eat and speak at the same time, this passageway can get rather crowded. Although we are able to narrow it if we want to, and this ability is used in some languages, its role in speech is generally small. Above the pharynx, the vocal tract divides. One passageway goes up into the nasal cavity from which air escapes through the nose; however, this only happens if we allow it. We can close off the access to the nasal cavity by raising the velum (also known as the soft palate), or allow air to go into the nasal cavity by lowering it. The extreme end of the velum is a small piece of tissue called the uvula (you can see it by looking into the back of your mouth in a mirror - it is not a pretty sight), which plays a part in the pronunciation of some languages. Inside the mouth there are many parts which are used in speaking. The most obvious one is the tongue, which is so important in speaking that in many languages the word for “language” is also the word for “tongue”. It is amazingly mobile and flexible: consider, for example, how far you can push your tongue outside your mouth (some people can touch the tip of their nose with the tip of their tongue). This is done with muscles, yet muscles can only pull, not push. The tongue can move upwards and downwards, forwards and backwards. In producing vowels, the tongue does not make a great deal of contact with the palate (the upper surface of the mouth, sometimes called the “roof” of the mouth); in many consonants, there is extensive contact. The lower jaw can also move upwards and downwards, and to a small extent forwards and
backwards too. The teeth can be important in speaking, though we can’t move them.

The outer end of the vocal tract is formed by the lips which, like the tongue, are very flexible and manoeuvrable. They can be moved towards each together and firmly closed, or can be moved further apart. They can be pushed forwards and rounded, or pulled back and widened as in a smile.

**Describing speech production**

Now that the various parts of the vocal tract have been introduced, let us look in detail at an example of how a word is produced. We will take the word *sand*. You should read the description and see if it agrees with how you feel you say it. You will certainly find it strange to think of such a simple bit of speech requiring so many actions and such careful coordination. In speaking normally, we never have to think about what we are doing. It is very different for people recovering from a stroke, when with the help of a speech therapist they have to re-learn how to make even simple speech sounds.

You would probably take a breath before starting to speak, since there has to be enough air in your lungs to produce speech. The velum is then raised so that air cannot escape through the nose. The first sound, [s], is made with the vocal folds apart and not vibrating. The air passes up through the larynx and the pharynx and into the mouth. Here it meets an obstacle: the front part of the tongue has been raised so that it is touching the roof of the mouth just behind the upper front teeth. This obstruction does not block the escape of air completely, but it forces the air to pass through a narrow gap and hit the teeth, making a hissing noise. The next sound is [æ], and for this your vocal folds must be vibrating. So to move from [s] to [æ], you must bring the vocal folds together until they are in contact with each other, or nearly so. At the same time the tongue must be lowered so that it is no longer obstructing the flow of air. To produce an [æ] sound the tongue must be quite low in the mouth, and the jaw is usually lowered. The velum is still raised to prevent the escape of air, but as you approach the [n] sound which follows, the velum must be lowered, and this begins to happen towards the end of the [æ]. To make the [n], the tongue and jaw must be raised into almost the same position it was in for making the initial [s] sound, but this time the closure between the tongue and the roof of the mouth is complete. The vocal folds continue to vibrate, but now, with the velum lowered and the escape through the mouth blocked, all the flow of air passes through the nasal cavity and escapes through the nostrils. The final sound is [d]. The move from [n] to [d] is a very simple one: remember that during the [n], the air is escaping through the nasal cavity and no escape of air is possible through the mouth. You now raise your velum (most people are not aware of doing this when they speak), and the escape of air through the nasal cavity is stopped. The flow of air stops and the vocal folds soon stop vibrating. This raising of the velum results in the [d] sound. Once this has been done, it is necessary to return to the position for normal breathing (you would not live very long if you stayed making the [d] sound and did not start the flow of air again). The vocal folds are moved apart, and both the tongue and the velum are lowered, so that air can flow freely into and out of the lungs.

This description sounds very complicated; it may help in understanding the process to read through the diagram in Figure 2.2. We have looked at just four speech sounds: [s], [n] and [d] are consonant sounds, while [æ] is a vowel. Although most people have heard of vowels and consonants, they usually find it difficult to say what the difference is. In phonetics, the essential difference is that
consonants make some obstruction to the flow of air, while vowels make relatively little obstruction. This means that if you produce alternate consonants and vowels (as in the word potato), your vocal tract is changing from closed to open, open to closed alternately. However, although the consonants in potato are clearly different from vowels, there are quite a lot of cases where the difference is not so easy to see.

<table>
<thead>
<tr>
<th>Lips</th>
<th>CLOSED</th>
<th>OPEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongue tip</td>
<td>RAISED</td>
<td>LOWERED</td>
</tr>
<tr>
<td>Velum</td>
<td>CLOSED</td>
<td>OPEN</td>
</tr>
<tr>
<td>Voicing</td>
<td>ON vv</td>
<td>OFF —</td>
</tr>
</tbody>
</table>

Figure 2.2. Diagram of articulator movements for the word sand

**What to read**

Chapters 2, 3 and 4 of this introductory book give a brief outline of such matters as the difference between vowels and consonants, the production of the airflow in speech, voicing, and some major classes of speech sound. There is a CD ROM with this book, and we will be using it in compulsory module LS2EP English Phonology in Year 2.

This is a fairly recent text book, and comes with a CD ROM. Sections A4, A5 and A6 give information on speech production and vowel and consonant possibilities.

O’Connor’s book is a clear and thorough introduction to phonetics. Chapter 2 describes in a comprehensive way the articulation of speech sounds.

Chapters 2 and 3 of this book give a more advanced and up-to-date treatment of the subject than O’Connor’s. Some of this material is also relevant to the topic of Chapter 3 of this book. Chapter 6 gives a thorough and detailed account of the anatomy and physiology of speech production.
THE SOUNDS OF LANGUAGE 3: CLASSIFYING SPEECH SOUNDS

Our articulators enable us to make an infinite number of different speech sounds, but in our scientific study we need some way of bringing order into the apparent chaos of all these different sounds. We have already seen how the most fundamental division we make (and have made for thousands of years) is between vowels and consonants. Within each of these basic categories, we can make finer and finer distinctions. It is interesting to compare our science with another such as botany. The world is full of plants of amazing variety, and each of them must be put into a scientific category; to do this requires a taxonomy of plants, a framework for classification. If a new plant is discovered, a botanist will try to decide whether it belongs within an existing family group, or whether a new class must be created. We do the same with the great variety of speech sounds we find being used in the world’s languages.

For each language we examine, we are able to identify a number of phonemes (introduced in Lecture 1) which function in that particular language as distinctive - they work to distinguish meanings of different words in the language. This is done by indicating a linguistic contrast. E.g., the difference between the phonemes /p/ and /b/ in English causes us to hear the difference between the words pin and bin; the difference between /s/ and /æ/ makes the difference between bit and bat. These word pairs are called minimal pairs as the only difference is in the target sound, i.e., it is a minimal difference. In the theory of the phoneme, the phoneme is abstract (like a letter of the alphabet), and what you hear is the realization of the phoneme – its physical form. Phonemes can have several different physical forms, or allophones.

When we have decided what category a sound belongs in, we can represent it with a symbol. Some symbols have already been introduced, but there are many more. It is a long-established convention that when one wishes to distinguish between symbols for the phonemes of a particular language and IPA symbols for allophones, or for sounds in general, one should put the symbols between different brackets. Phoneme symbols should be between slant brackets: /ɛ/, /s/, and so on; phonetic symbols should be between square brackets: [ø], [Ʉ]. As an example, we will look at the word ostrich. We can write this relatively simply in phonemic transcription as /ɒstrɪtʃ/; however, we might want to be more precise about the exact pronunciation. The /tʃ/ is usually a voiceless sound in this context and can be represented phonetically as [ɹ̥]. Most English speakers have rounded lips for [ʃ], which is therefore transcribed phonetically as [ʃ]; they also tend to put a glottal stop [ʔ] in front of [ʃ], though this sound is not a phoneme of English. A phonetic transcription of ostrich, then could look like this: [ɒs̩tɹ̥ɪtʃ] (think about it – are your lips rounded for the [st] sounds?).

When all the classification possibilities available to phoneticians have been listed, they can be put together in the form of a chart, the best known of which is the IPA (International Phonetic Association) Chart, included in this Handbook.

Vowels
We have already looked briefly at vowels in the previous lecture. It has been claimed that the most basic vowels in languages [i] (similar to the English word key) and [a] (similar to the vowel in half) – they are found in the great majority of the world’s languages, and they are the vowels that babies learn first. They are also different in one very important way: in [i], the tongue is close to the palate, while in [a], the mouth is open, with the tongue low in the mouth. You can see
the difference if you look in a mirror while saying these two sounds, one after the other. We therefore class [i] as a close vowel and [a] as an open vowel.

We can find another very basic vowel: its symbol is [u]. Although the English vowel in who is a reasonable example, we can find clearer cases in some other languages; the vowel in French vous, German du or Spanish tu. The [u] vowel differs from [i] in two important ways. The one which is easiest to observe is that the lips are rounded in [u], but unrounded or spread (as for a smile) in [i]. More difficult to observe is that the back of the tongue is raised for [u] while in [i] it is the front of the tongue which is raised. As a result, we say that [i] is a front vowel, while [u] is back.

By looking at these three vowels, we have seen the most important features used for classifying vowels: a vowel may be close or open; front or back; rounded or unrounded. We can now look at where other vowels fit in this scheme – there are many different vowels in the world’s languages, and we need to be able to put them in their place. On the open/close scale, we place two intermediate levels: mid-close and mid-open. In between front and back, we put central. At each point on the outside of the diagram we place two symbols: the left-hand item of each pair is unrounded, while the one on the right is rounded. There is a well-known diagram used to represent the vowel “space”, known as the Cardinal Vowel diagram. This can be seen in Figure 3.1.

![Cardinal Vowel diagram](image)

Figure 3.1. Cardinal Vowel diagram

The vowels that we place on this diagram are Cardinal Vowels, and these are not the vowels of any particular language. Indeed, there is some doubt about whether one of these vowels, [a], has ever been observed as a phoneme in any language in the world, but it is nevertheless a vowel which we are capable of making, so we give it a symbol and a place on the diagram (in fact, the nearest sound to this rare beast that you are likely to hear is the sound of someone yawning loudly). Given this way of classifying vowels, we can give any vowel a phonetic label by describing its frontness, openness and rounding: thus, using cardinal vowels as examples, we can say that [i] is a front close unrounded vowel, while [u] is a back close rounded vowel. The vowel [e] is a front mid-close unrounded vowel, while [ə] is a back mid-open rounded vowel.

There are many other ways in which vowels differ from each other, though the features we have looked at in this section are the most important ones. Some languages have nasalized vowels in addition to normal ones; in these, some of the air-flow is allowed to escape through the nose. French is a well-known example – the vowels in fin, bon, dans, brun are nasalized. Among other
European languages, Portuguese and Polish also have nasalized vowels. There are also differences in vowel length, and some languages contrast long and short vowels. It has been claimed (but not accepted by everyone) that the Estonian language distinguishes short, medium and long vowel and consonant phonemes.

**Consonants**

As we saw in Lecture 2, all true consonants are the result of obstructing the flow of air through the vocal tract. When we classify consonants, we look at the following characteristics:

(i) Whether the sound is **voiced** or **voiceless**.
(ii) The **place** or places where the obstruction is made.
(iii) The **manner** of the obstruction.

We also take into account the **airstream** used to make the consonant. In English, almost all sounds are produced on a **pulmonic egressive** airstream mechanism, i.e., using air which is expelled from the lungs.

In the descriptions that follow, the examples are from RP/BBC English unless otherwise specified.

(i) **Voicing** is sometimes seen as a binary (yes/no) matter – a sound is either voiced or it isn’t. It is in fact rather more complex than this – some sounds are voiced for only part of their time. For example, in English, the phonemes /b/, /d/ and /ɡ/ (these are consonants of the type called plosives) often occur at the beginning of a word; although they are classed as voiced, the voicing usually begins only just before the end of the consonant. When the same consonants occur at the end of a word, we find the reverse: /b/, /d/ and /ɡ/ have some voicing at the beginning, but then this dies away. Often the “voiced” English fricatives /v/, /z/ and /ʒ/ behave in much the same way. Compared with these, other sounds such as /m/, /n/ and /l/ usually display full voicing, as do /b/, /d/ and /ɡ/ in many other languages.

(ii) **Place of articulation.** We have already seen in the preceding lecture how consonants are made. We need to look in more detail at the vocal tract in order to identify the places. Starting from the outer end of the vocal tract, we have the lips, which give us the **bilabial** place of articulation. Behind these are the front teeth; if the lower lip touches the upper front teeth, we have a **labiodental** place, while if the tongue touches the teeth, the place is **dental**. Behind the upper front teeth is the **alveolar ridge**, and if the tongue is in contact with this, the place is alveolar. The tongue can make contact with the upper surface of the mouth a little further back than the alveolar region, giving a **post-alveolar** place, while looking further back in the vocal tract brings us to the **palatal** area. Looking now at the back part of the mouth, we can see the velum. When the back of the tongue is in contact with this, the place is **velar**. If the contact is further back, with contact against the extreme lower end of the velum, the place is **uvular**.

    Moving downwards towards the larynx, we can see the pharynx, and if we constrict this we get a **pharyngeal** place. A constriction between the vocal folds, inside the larynx, has a **glottal** place of articulation, since the space between the vocal folds is known as the glottis.
In addition to these places in the mouth, there is another characteristic of some sounds which is traditionally classed as a place of articulation: this is retroflex. In a retroflex sound, the front part of the tongue is curled backwards so that if it makes contact with the upper surface of the mouth, it is the tip of the tongue, or even the underside, which makes contact. Consonants of this sort are commonly found in languages of the Indian sub-continent. “Retroflex” is not really a place, but a shape of the tongue, and does not really belong in any of the normal categories of sound classification. Retroflexion is found also in vowels — it is often heard in the speech of Americans in vowels where there is an ‘r’ in the spelling, such as car, more, bird.

(iii) **Manner of articulation:** here we have to describe the type of obstruction to the airflow that is formed. This can range from a complete closure of the vocal tract, which prevents any air from escaping, to an articulation which in most ways is identical to that of a vowel. A plosive is a consonant which stops air from escaping. A closure is made at some point in the vocal tract, and air is compressed behind this. There is a brief period of complete, or almost complete, silence, and then the compressed air is released. When this air is released, there is a very short explosive noise, called plosion. This may be followed by a [h]-like sound known as aspiration.

A nasal consonant involves a complete closure in the oral cavity, but air is allowed to escape through the nose, since the velum is lowered for the production of this type of consonant.

A fricative requires a constriction which allows the air to escape with difficulty, so that a hissing sound is produced. An affricate is a consonant which starts as a plosive, but instead of ending with plosion, ends with a fricative made in the same place.

There is a class of sounds which are very brief: tap and flap. To make a tap, the tongue is briefly flicked up against the roof of the mouth, interrupting the flow of air for a short time. A flap is similar, but the tongue is curled back and then flicked forward against the alveolar ridge.

An unusual and difficult sound is the trill. Various articulators are mobile enough to vibrate if the air-flow is suitable. The tongue-tip and the uvula are suitable for trilling and are used for this in a number of languages. If you trill your lips, you produce what is sometimes called a “raspberry” or, in America, a “Bronx cheer”.

Finally, there is a class of sounds which are collectively called approximants. One of these is the lateral: in this type, the centre of the tongue is in close contact with the roof of the mouth, but the sides of the tongue are lowered so that air can escape along the sides of the tongue. A post-alveolar approximant is a rather vague concept, but the term is normally used to refer to the ‘r’ sound of the English of America and England, where the tongue is slightly curled backwards but does not make contact with the upper surface of the mouth.

(iv) **Airstream mechanism:** finally, we should (if we are being as precise as possible) also specify the airstream mechanism of a consonant: it may be pulmonic, made by the movement of air out of or (much more rarely) into the lungs, glottalic, made by moving air inwards or outwards by lowering or raising the larynx in the throat, or velaric, made by making a velar closure (as for [k] or [ɡ]) and sliding the tongue backwards or forwards to
move air inwards or outwards. When the air is moved outwards, we call it **egressive**, while inward movement is called **ingressive**.

Having established these principal ways of classifying consonants, we can make up labels which define any given consonant. We will assume for now that we are looking at egressive pulmonic consonants (made with air expelled from the lungs).

\[
\begin{align*}
[s] & \text{ is a voiceless alveolar fricative} \\
[g] & \text{ is a voiced velar plosive} \\
[l] & \text{ is a voiced alveolar lateral approximant} \\
[tʰ] & \text{ is a voiceless aspirated alveolar plosive.}
\end{align*}
\]

These labels are important in phonetics for specifying a sound in an unambiguous way, but it can take a long time to learn how to use them properly.

**What to read**

Generally a very useful book, written as a course. See chapters 5 and 6 for information on the organs of speech and airstream mechanisms.

This is the official account of the I.P.A.’s categorisation of speech sounds and use of phonetic symbols, and is a very valuable book for students of phonetics.

This account contains a large amount of information about the enormous range of sounds used contrastively in the world’s languages; in their field work over many years, the writers have themselves often been the first to describe them in detail.

In Parts II, III and IV of this very substantial book (pages 95-335), Laver sets out a detailed and comprehensive account of how the sounds of speech can be unambiguously categorised, with a wealth of illustrations from different languages.

The IPA Chart, [https://www.internationalphoneticassociation.org/content/ipa-chart](https://www.internationalphoneticassociation.org/content/ipa-chart), and sections of it are available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2015 International Phonetic Association.
# THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC) © 2005 IPA

<table>
<thead>
<tr>
<th>Plosive</th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Retracted</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal</td>
<td>p</td>
<td>m</td>
<td>t</td>
<td>d</td>
<td>[t]</td>
<td>d</td>
<td>c</td>
<td>f</td>
<td>s[z]</td>
<td>[s]</td>
<td>[z]</td>
</tr>
<tr>
<td>Trill</td>
<td>b</td>
<td>n</td>
<td>n</td>
<td>η</td>
<td>η</td>
<td>η</td>
<td>η</td>
<td>η</td>
<td>η</td>
<td>η</td>
<td>η</td>
</tr>
<tr>
<td>Tap or Flap</td>
<td>v</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>φ β f v</td>
<td>θ z j f z</td>
<td>s [s]</td>
<td>c j x y</td>
<td>ξ</td>
<td>ξ</td>
<td>ξ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral Fricative</td>
<td>i h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>u</td>
<td>i</td>
<td></td>
<td>j</td>
<td>μ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral Approximant</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

CONSONANTS (NON-PULMONIC)

<table>
<thead>
<tr>
<th>Clicks</th>
<th>Voiced implosives</th>
<th>Ejectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilabial</td>
<td>b</td>
<td>Bilabial</td>
</tr>
<tr>
<td>Dental</td>
<td>p</td>
<td>Dental/velar</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>Dental/velar</td>
</tr>
<tr>
<td>lateral</td>
<td>g</td>
<td>Uvular</td>
</tr>
</tbody>
</table>

OTHER SYMBOLS

M Voiceless labial-velar fricative
W Voiced labial-velar approximant
Y Voiced labial-velar approximant
H Voiceless epiglottal fricative
F Voiceless epiglottal fricative
? Epiglottal plosive

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

VOWELS

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

SUPRASEGMENTALS

<table>
<thead>
<tr>
<th>Primary stress</th>
<th>Secondary stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td>Half-long</td>
</tr>
<tr>
<td>Extra-short</td>
<td>Major (intonation) group</td>
</tr>
<tr>
<td>Linking (abscence of a break)</td>
<td></td>
</tr>
</tbody>
</table>

TONES AND WORD ACCENTS

<table>
<thead>
<tr>
<th>Level</th>
<th>Extra high</th>
<th>High</th>
<th>Mid</th>
<th>Low</th>
<th>Extra low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contour</td>
<td>Rising</td>
<td>Falling</td>
<td>High rising</td>
<td>Low rising</td>
<td>Rising-falling</td>
</tr>
<tr>
<td>Downstep</td>
<td>Global rise</td>
<td>Uptop</td>
<td>Global fall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We have seen in Lecture 2 how we are able to control the pitch of our voice. In this lecture we look at one of the most important functions of this pitch control. We will begin with a simple example. In the following examples, the words are given in phonetic transcription and each has a diagram of the pitch that it is produced with the upper line representing the highest pitch of the speaker, and the lower line the lowest pitch. The words are from Kono (Sierra Leone).

\[
\text{kɔɔ “to mature”} \quad \text{kɔɔ “rice”}
\]

Note that both words contain exactly the same segments, but they are different in pitch. In some cases it is the pitch level that distinguishes a word, while in others it is a pitch movement. We refer to these characteristics of pitch as tone. Each word in the above example has a distinctive tone. One of the interesting facts about languages is that some use tone in this distinctive way, while others (including most European languages) do not. Languages which use tone distinctively are called tone languages. It is probably true to say that the majority of the world’s population speaks a tone language, so it is surprising how little importance is given to this aspect of speech in books on phonetics. Perhaps this shows that the subject is dominated by speakers of European languages. We find tone languages in South-East Asia (e.g., Chinese, Thai, Vietnamese), in much of southern Africa (e.g., Hausa, Yoruba, Zulu) and in indigenous languages of America (e.g., Mixteco, Zapotec, Navajo).

**Lexical and grammatical use of tone**

Tone works in different ways in different languages. The easiest to explain is what we can call lexical tone, where the meaning of a word (as it would appear in a dictionary) is determined by its tone. The example from Kono, given above, is typical.

In some languages, tone may function as a way of showing different aspects of grammar. In Kono, the following sentences differ only in the use of high and low tones (shown by the tone-marks placed before each syllable in the transcription):

\[
\text{_a¯a ˙do ˙MA ˙ko} \quad \text{“Wash his shirt”}
\]

\[
\text{_a˘a ˘do ‘MA ˘ko} \quad \text{“He has washed a shirt”}
\]

**Tone Levels and Tone Contours**

In the phonetic study of tone, we are not just concerned with the meaning of tones, but with the physical nature of their sounds. In some tone languages, the most important thing about the tones is the difference between tone levels: whether the pitch is high or low. Some languages distinguish only high and low, while others may have as many as four different levels. Each speaker will have their own individual pitch level for high and low, and for the tones in between, so the important thing is the difference between the tones rather than the exact pitch of the tone. In other languages, however, the important difference is between the shapes of the tones rather than the levels – tones can have rising pitch, falling pitch, falling-rising or rising-falling as well as level. It has been claimed for a long time that there is a fundamental difference between these two
types of tone language (they are sometimes called register and contour tone languages), but there are many cases where it is difficult to decide which type a language belongs to, so the distinction does not seem to be very useful.

Tones and Context
Most speakers who do not speak a tone language would find it relatively easy to study tone if it were not for the fact that tones tend to vary in ways that are very difficult to predict according to the context in which they occur. Even if one can successfully recognise the individual tones of a tone language in words spoken in isolation, the tonal characteristics of those words may become virtually unrecognisable when they occur in connected speech. Around two thousand years ago, Sanskrit grammarians in India made detailed studies of the effects of context on speech sounds, and gave the term sandhi to these effects. In tonal sandhi, a tone is modified by the tones of neighbouring syllables. An example is found in Mandarin Chinese, in which the tones are numbered as follows:

(1) high level (e.g. ma mother)
(2) high rising (e.g. ma hemp)
(3) falling-rising (e.g. ma horse)
(4) high falling (e.g. ma scold)

However, when a syllable carrying tone (3) is followed another tone (3) syllable, the first syllable’s tone changes to tone (2). When a tone (2) syllable is preceded by a tone (1) or tone (2), and is followed by a stressed syllable, it becomes a tone (1). This does not make life any easier for someone learning Chinese.

In Vietnamese, whose writing system is based on the Roman alphabet, vowels are represented with additional marks which indicate one of the language’s six tones. Two of the tones, known as “broken” tones, are (in the Northern dialect) accompanied by glottalization (rather like a rapid glottal stop in the middle of the syllable). A Vietnamese dictionary has to list the different words with their tones. The marks, and the tones they represent, are as follows:

(1) Mid-level (no mark). Example: a
(2) Low falling (grave accent). Example: à
(3) High rising (acute accent). Example: á
(4) Low, rising after dip (circle). Example: à
(5) High “broken” (tilde). Example: ã
(6) Low “broken” (subscript dot). Example: ạ

Tones and pitch-accents
Nobody has ever claimed that English is a tone language. However, pitch and pitch movements play an important role in English and similar languages in marking out certain syllables as distinctive and important. If you say the word important on its own, you will probably notice that on the middle syllable the pitch of your voice steps up from the pitch of the first syllable, then glides down to the lower pitch of the final syllable. This distinctive pitch movement is sometimes known as accent, and the middle syllable of important can be said to be accented. It is not always easy to distinguish this function of pitch from the tonal contrasts that we have been looking at earlier in this lecture. In a number of languages which are not normally thought of as tone languages it is possible to find pairs of words which really seem to be distinguished from each other by pitch characteristics. Japanese is one such language, with pairs such as hashi with pitch accent on the first syllable meaning chopsticks, and hashi with pitch accent on the second syllable meaning bridge or edge.
In a tone language, pitch variation is being used on practically all the syllables or words of the language, but in the cases just mentioned, only a limited number of words are distinguished by pitch. To avoid having to class these languages as tone languages, we say that they are *pitch-accent* languages, and that certain words are distinguished by pitch-accents.

**What to read**


In tone languages, it is usually relatively easy to see the function of the different tones. However, in languages which do not use tone in this way, it is harder to explain what we are doing when we make use of changes in pitch, loudness and other suprasegmental features of speech. There are several such features: what they all have in common is that we usually see them as properties of pieces of speech which will be at least as long as one syllable, and may extend for many words. So, for example, I say no loudly, it is most likely that both the /n/ sound and the /əʊ/ will be loud. If I say hurry quickly, then all the phonemes of that word will be said quickly. The most important suprasegmental features of speech are pitch, loudness, tempo (that is, speed) and voice quality, but these are by no means the only ones. The study of these features is often referred to as the study of prosody. Two such features form the basis for especially important functions, stress and intonation.

Stress and accent
In any language you listen to, you will notice that some syllables sound stronger and more noticeable than others. In English, for example, the middle syllable of the word tomato is clearly stronger than the first and last syllables. We say that the middle syllable is stressed. In some languages the position of the strongest stress may affect the meaning of a word. The following Spanish words are shown with the stressed syllables underlined, and their meanings are given:

(i) término terminus
(ii) termino I terminate
(iii) terminó I terminated

English has some pairs of semantically-related words whose grammatical category is reflected in their stress pattern, such as:

(i) import (noun)
(ii) import (verb)

There are other pairs of words where the difference in stress signals functions in other ways: in the pair subject / subject, the two words differ in grammatical category (noun / verb) but seem in some contexts to be unrelated semantically; the pair recall / recall, however, may have the same grammatical category (both may be nouns) and be semantically related. In some other languages, it is possible to hear the difference between stressed and unstressed syllables, but the stress usually falls in the same position in a word of more than one syllable. In French, it is usual for stress to fall on the final syllable of the word, while in Polish it is usually on the penultimate syllable (the syllable before last), with a few exceptions such as uniwersytet (university). In languages such as these, we cannot say that stress is able to determine the meaning of a word. We may guess that stress performs a different function: it helps us to divide the continuous flow of speech into separate words. For example, if I am listening to someone speaking French, I know that when I hear a stressed syllable, that is the last syllable of a word. One of the great unsolved mysteries of speech perception is how we manage to divide continuous speech up into separate words in languages like English (in which stress gives us relatively little help).

A number of factors cause a syllable to be made prominent so that it is heard as stressed. In English, stressed syllables are usually louder and longer than unstressed syllables, and have distinctive pitch (as in the example of the word important given in the previous lecture). We can detect different levels of stress in
words of several syllables. Try saying to yourself the four-syllable English word *understanding*, the strongest stress should be heard on the third syllable, but the second and fourth syllables are much weaker than the first syllable. Usually, only the third syllable has a noticeably distinct pitch.

Distinctive pitch, such as we find on the third syllable of *understanding*, is given special importance in the study of stress in English, and the term **accent**, which was introduced in the previous lecture, is used to refer to it. Consider now how you would say the phrase *understanding English*: you will probably find that there is no longer a noticeable pitch movement on the third syllable of *understanding*, but there is one on the first syllable of *English*. The *stand* syllable is still quite prominent, but it isn’t accented. The word **stressed** can be retained for syllables (such as *un* and *stand* in this example) which are made prominent by other features such as length or loudness. The same distinction can be made in most other languages in which stress and accent play linguistically important roles.

**Intonation**
Intonation has always been a difficult thing to define. According to traditional descriptions, intonation is “the melody of speech”, and is to be analysed in terms of variations in pitch. We have seen, in the lecture on tone, how changes in pitch can change meaning, but in the case of intonation the way meaning is changed is not so clear. If we look at a typical example, we would expect a falling pitch pattern on a statement like this:

```
You’re from London
```

```
— — \ —
```

but a rising pitch pattern if the same words are used as a question:

```
You’re from London?
```

```
— — / —
```

Intonation can, then, indicate different types of utterance, such as statements and questions. Other examples of meaning being changed by differences in intonation are often quoted: the difference between

```
She won’t go out with anyone
```

```
— — — — \ / —
```

and

```
She won’t go out with anyone
```

```
— — — — / _/
```
is that the first one (with a falling pitch movement starting on any) says that she will go out with nobody, while the second (with a falling-rising pitch movement) says that she is careful about who she goes out with.

Intonation is said to indicate the attitudes and emotions of the speaker, so that a sentence like *I think it's time to go now* can be said in a happy way, a sad way, an angry way, and so on. While this is certainly true, there is clearly more than just pitch variation involved in conveying such things, so the definition of intonation becomes considerably more complicated. It is clear that when we are expressing emotions, we also use different voice qualities, different speaking rates, facial expressions, gestures and so on. Intonation also gives the listener a lot of information about what is being said. In English (of South-East England), we make a lot of use of a falling-rising intonation pattern: for example, it is often used for polite requests:

<table>
<thead>
<tr>
<th>Can you lend me some money?</th>
</tr>
</thead>
<tbody>
<tr>
<td>—- —- —- \ /</td>
</tr>
</tbody>
</table>

and for expressing reservation:

<table>
<thead>
<tr>
<th>It might be good enough</th>
</tr>
</thead>
<tbody>
<tr>
<td>— \ — — — —</td>
</tr>
</tbody>
</table>

We can use intonation to signal that we are giving a list:

<table>
<thead>
<tr>
<th>We need crisps, drinks, fruit, sweets and cheese</th>
</tr>
</thead>
<tbody>
<tr>
<td>—— —— —— —— —— \ /</td>
</tr>
</tbody>
</table>

It is difficult to pin down the functions of intonation. A further problem lies in trying to work out a satisfactory way of transcribing it. In the examples given above, a “wiggly-line” representation of the pitch movement is given, but although this helps to explain the examples, it does not indicate which bits of intonation are significant. The problem lies in the fact that intonation is claimed to be meaningful and contrastive in a similar way to the tones of a tone language. In the latter case, we know that we can represent the tones with a set of marks which correspond to the contrastive tones of the language. But it is much more difficult to decide how to divide intonation up into contrastive units, and many different ways have been proposed.

**Rhythm**

There are many parallels between speech and music, and one thing that is always found in music is rhythm. In music, the rhythm is usually produced by making certain notes in a sequence stand out from others by being louder or longer or higher. We should not make the mistake of thinking that musical rhythm is just an unvarying repetition of beats at equal intervals. This may be true of commercial pop music (as can be heard coming out of someone’s headphones, or through the wall from the room next door), but throughout the world in traditional folk music and other serious musical forms we can find some amazingly complex rhythms which are still immediately recognisable as regular. In speech, we find that syllables take the place of musical notes or beats, and in
many languages the stressed syllables determine the rhythm. If you were asked to clap your hands in time with the sentence

This is the 'first 'time I've 'ever 'eaten a 'chocolate 'caterpillar

you would be most likely to clap at the points marked with the stress mark /ˈ/. It is often claimed that English speakers try to keep an equal time between the stressed syllables, so the time between claps of your hands would be quite regular; rhythm of this type is called stress-timed. In fact this is only found in a style of speech (slow, emphatic) where the rhythm is strong, and in ordinary conversational speech it is much harder to make a convincing case for this “isochronous” rhythm (where the time intervals between stressed syllables are equal); as with music, we should not expect rhythm to be simple. Other languages have different rhythms (as you can easily hear by listening to them). To the ears of English speakers, Italian and Swedish have a very different rhythm from English. Spanish, French and Chinese sound syllable-timed to English listeners – it sounds as though all the syllables are of equal length, and the dominant role of stressed syllables in making up the rhythm is much less noticeable. But these judgements are very subjective, and finding scientific evidence about what makes us hear languages as rhythmically different is proving to be very difficult. What does seem to be clear is that rhythm is useful to us in communicating: it helps us to find our way through the confusing stream of continuous speech, enabling us to divide speech into words or other units, to signal changes between topic or speaker, and to spot which items in the message are the most important.

What to read

Cruttenden, A. (1997). Intonation (2nd ed.). Cambridge: Cambridge University Press. An excellent introduction to the study of intonation, this book also gives good coverage of other suprasegmental features such as stress, accent and rhythm. It is mainly based on English.

Ladd, D. R. (1996). Intonational phonology. Cambridge: Cambridge University Press. Although difficult in places, Chapter 1 of this important book is a very good survey of some of the key questions in contemporary thinking on intonation. Ladd attempts a definition of intonation, a justification of its phonological status and an evaluation of laboratory studies. He excludes “lexical” features (properties of words such as word-stress patterns), and proposes that intonation instead conveys “post-lexical” (i.e. non-lexical) pragmatic meanings in a linguistically structured way. Section 1.2 of this chapter looks at one particular theory of intonation (the IPO approach), and it is not necessary to read this to understand the principal points of Ladd’s treatment.
We have seen in previous lectures how we can describe speech sounds, and classify them according to established frameworks. We must not forget, however, that we produce speech for a particular purpose – to communicate. We can only do this because speech sounds are used by speakers of a particular language according to its rules, and speech perception can only work by our being able to recognise these sounds. Each language has only a limited set (or “system”) of sounds that its speakers use. Each sound in the system must be distinctive: we must be able to show that the identity of a word can change if we substitute one phoneme for another. In addition, these sounds can only be combined in particular ways.

It is interesting to look at how speech develops in children. In the child’s earliest months of life, it makes a meaningless babble of noises that have little or no resemblance to the sounds of the language it is going to learn. But in a year or two from birth, most of the strange sounds the baby makes will disappear from its speech for ever (unless it decides to study phonetics later in life), and the child will have learned the sounds and the patterns of sounds that are typical of the language. What is it that the child has learned? The short answer is to say that it has learned the phonology of its native language. The phonology of a language is part of its linguistic structure, which explains why phonology of the sort described here is sometimes called “linguistic phonetics”. The theoretical study of phonology involves far more than the rather basic observations about sound systems that are made in this lecture; to give a proper account of phonology would need a whole module at least.

**Systems of sounds**

Space does not allow a detailed treatment of the different systems of contrasting sounds found in different languages, but the following brief survey shows some of the variety found in vowel and consonant systems.

**(1) Vowels**

Many interesting observations have been made about the vowel systems of different languages; many are the subject of argument and competing interpretations. One question concerns the number of contrasting vowels that a language may have. Many of the world’s languages have only three vowel phonemes, and these are almost always /i/, /a/ and /u/. Many more have five, and these are most often the three above, plus /e/ and /o/. It seems that there are more odd-numbered vowel systems than even-numbered in the world. The lowest number is disputable, but there have been claims of languages with only two vowel phonemes, and a somewhat mythical one-vowel language. When we find a language with a large number of vowels, we usually find it possible to sub-divide the vowels into categories. British English (RP/BBC accent), for example, is claimed to have twenty vowels, usually divided into the following groups:
With a little ingenuity, this large number can be reduced to a much smaller system, for example by treating long vowels and diphthongs as being combinations of two phonemes. A six-vowel analysis of the English vowel system is possible.

(2) Consonants
All languages have consonants, but their number and variety is very different from language to language. As with vowels, we tend to look for groups and patterns rather than overall lists. Some languages manage with very few consonants (less than a dozen, in some cases). Very few of the many Australian aboriginal languages have fricative consonants, yet almost all other languages have at least some. English has six plosive consonant phonemes (/p t k b d ɡ/), but many languages of India have more: Hindi, for example, has sixteen. They can be arranged like this:

<table>
<thead>
<tr>
<th>Plosive type</th>
<th>Place of articulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bilabial</td>
</tr>
<tr>
<td>Voiceless unaspirated</td>
<td>p</td>
</tr>
<tr>
<td>Voiceless aspirated</td>
<td>ĥ</td>
</tr>
<tr>
<td>Voiced unaspirated</td>
<td>b</td>
</tr>
<tr>
<td>Voiced aspirated</td>
<td>b̂</td>
</tr>
</tbody>
</table>

Table 6.1. Plosives in Hindi

English speakers find it difficult to learn to manage such a complicated set of contrasting plosives, but Hindi speakers apparently manage with no trouble. Another Indian language, Malayalam, contrasts bilabial, dental, alveolar, retroflex and velar places of articulation for plosives and nasals. The use of voicing and aspiration for making phonemic distinctions varies greatly among different languages: a few languages do not have a voiced/voiceless distinction at all. Korean does have voiced and voiceless consonants, but among its plosives there are no voiced plosives such as /b/, /d/ and /ɡ/. There are voiceless plosives with (i) no aspiration, (ii) weak aspiration and (iii) strong aspiration. We might represent this in table form as follows:
<table>
<thead>
<tr>
<th>Korean Plosives</th>
<th>Place of articulation</th>
<th>Unaspirated</th>
<th>Weakly aspirated</th>
<th>Strongly aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilabial</td>
<td>p</td>
<td>pʰ</td>
<td>ph</td>
<td></td>
</tr>
<tr>
<td>Alveolar</td>
<td>t</td>
<td>tʰ</td>
<td>th</td>
<td></td>
</tr>
<tr>
<td>Velar</td>
<td>k</td>
<td>kʰ</td>
<td>kh</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2. Plosives in Korean

The great variety of the sound systems of the world’s languages is one of the most fascinating aspects of phonetics and phonology.

What to read


Cruttenden, A. (2014). *Gimson’s pronunciation of English* (8th ed.). London: Arnold. Chapter 5 (“Sounds in Language”) is useful reading, and in addition provides a brief survey of the phonology of suprasegmentals. Other editions of this text are also suitable.

Syllable structure and phonotactics
Something else that varies from language to language is the way in which the sounds can be combined together. When this is studied, it is usually done in the context of syllable structure: if you can describe the form which syllables may take in a particular language, you are able to describe the possible combinations of phonemes. In many of the world’s languages, the description is very simple: every consonant must be followed by a vowel, giving a syllable form that we can label CV (Consonant + Vowel), and the only other possible syllable is V (a vowel on its own). Most Japanese syllables are of this form. More complex syllable patterns might allow, for example, up to two consonants at the beginning of a syllable and up to one at the end: we find this in Spanish – the word tren (train) is an example of a maximum syllable.

When one studies syllable structure, one is looking at the make-up of syllables in terms of consonants and vowels, i.e., how many consonants or vowels a syllable in a certain language can have. English syllables may have three consonants at the beginning (in words like strong /strɒŋ/, screw /skrʊ:/), and up to four at the end (in words like sixths /sɪksθs/; texts /teksts/). It is very important to note that the choice of which consonants and vowels can occur in syllables is not free: each language has quite firm restrictions on what is and what is not a real syllable of the language, and learning these “rules” is part of learning the language. The rules concerning which consonants and vowels can occur are called phonotactic constraints. It isn’t enough to say that Spanish syllables may have one consonant at the end: that consonant can only be one of a small set of consonants. English syllables may end with no consonant (as in me /mi:/), but this statement does not tell you everything you need to know, since most short vowels cannot occur in a syllable with no final consonant. You could not have an English word such as /piː/, /pe/, /pæ/, /pɒ/, /pʌ/ or /pʊ/, but it is perfectly possible to have “open” single syllable words like this if the vowel is long: pea /piː:/, par /paː:/, pore /pɔː:/, pooh /puː:/.

The study of such patterns of sounds is part of the subject of phonology, and forms a very important part of the study of the sound systems of different languages.

Phonetic vs. phonological inventories
Two languages can have the same phonetic sound inventory, but differ in their phonological inventory. For example, the two synthetic languages which follow both have [p t k b d ɡ] as consonant sounds in their phonetic inventory, but demonstrate different patterns in how they are used.

Language 1: Pobo

| [pu]  | tree   | [bu]  | fly   |
| [tu]  | Six    | [du]  | sky   |
| [ku]  | Ear    | [gu]  | fingernail |
| [pubu] | ancestor | [pupu] | green |
| [putu] | glass  | [pudu] | lake  |

The phonological inventory is the same as the phonetic inventory, as [p t k b d ɡ] all contrast.
Language 2: Paga

<table>
<thead>
<tr>
<th></th>
<th>Key</th>
<th>Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>[pu]</td>
<td>Fat</td>
<td>[bu]</td>
</tr>
<tr>
<td>[tu]</td>
<td>blue</td>
<td>[du]</td>
</tr>
<tr>
<td>[ku]</td>
<td></td>
<td>[gu]</td>
</tr>
<tr>
<td>[pubu]</td>
<td>beer</td>
<td>[pugu]</td>
</tr>
<tr>
<td>[pudu]</td>
<td>earache</td>
<td></td>
</tr>
</tbody>
</table>

Additional information: *[pupu]; *[puku]; *[putu] (* = ‘not legal’).

The phonological inventory is [p t k]. In word initial position, [b d ɡ] are in free variation (i.e., either the voiced or voiceless variant can occur) with their homorganic counterparts, and are also positional variants, or allophones, intervocalically (i.e., between vowel sounds), where one can only have a voiced sound.

When looking for sound patterns in languages, we are interested in meaning and contrast. Looking at the data above, we can see that changing, e.g., [p] for [b] in Pobo results in a change of meaning (from tree to fly), but doing the same for Paga does not. Therefore, we can say voicing in Paga is non-contrastive. Also, voiceless consonants [p t k] cannot occur between vowels in Paga – the voiced variant always occurs there. Pobo allows both voiced and voiceless sounds in intervocalic position.

What to read

Chapter 4, in particular sections 4.7 and 4.8, look at patterns of sounds in languages like the ones for Pobo and Paga above.

Chapter 8 presents an analysis of the syllable structure of English.
We normally speak by producing a continuous, connected stream of sounds, except when we pause. In most languages we can find occasional cases where a speaker makes a single segment in isolation with no sound preceding or following it (in English, for example, we can say *ah* /ə:/ if we make an interesting discovery, or *sh* /ʃ/ if we want to call for quiet), but such cases are rare. The usual situation is for segments to fit closely together with each other, particularly in rapid speech.

We have seen that phonetics tends to look on speech as a sequence of segments. However, to imagine these segments as discrete and independent of each other would be quite wrong. In every language we find that segments have a strong effect on other segments which are close to them. The process by which a segment is modified by its neighbours is called **assimilation**, and the description of assimilation has been a part of phonetic description for a long time. As we will see later, much recent phonetic research in this area refers to **coarticulation** instead, and we will have to discuss whether there is any significant difference between these terms. Another phenomenon of connected speech is **elision**, the process by which sounds that would be pronounced in slow, careful speech seem to disappear.

**Assimilation**

Let us look at some examples of assimilation. In French, a word-final voiceless consonant will often become voiced if followed by a voiced segment. For example, the word *avec* on its own is pronounced /avek/, but when it is followed by a word beginning with a voiced consonant such as /v/ in *vous* /vu/, we usually hear /aveg/. So the phrase *avec vous* is often pronounced /aveg vu/. In English, we also find assimilations of voice, but it is more common to find them in the form of loss of voice, or **devoicing**. If the word *have* occurs in final position, its final consonant /v/ will usually have some voicing, but when that /v/ is followed by a voiceless consonant it normally becomes completely voiceless; thus *I have to* is likely to have the pronunciation /aɪ hæf tu/.

Assimilation, then, is concerned with one sound becoming phonetically similar to an adjacent sound. The examples given so far are of anticipation, where a sound is influenced by the sound which follows it; another term frequently used for this type is **regressive** assimilation. We also find cases where the assimilation can be called **progressive**: here, not surprisingly, the process is for a sound to take on characteristics from a sound which precedes it. In general, this effect is less frequently found, though it is difficult to explain why this should be so. Historically, it must have been effective in English in order to produce the different pronunciations of the -s ending: the plural of *cat* /kæt/ is *cats* /kæts/ with a final /s/; the plural of *dog* /dɒɡ/ is *dogs* /dɒɡz/ with /z/. The voicing of the suffix is conditioned by the voicing of the preceding final consonant.

Assimilations are traditionally classified into three main types, though as we shall see this classification is not completely adequate.

1. One type is assimilation of voice (we have seen examples of this taken from French and English); this may take the form of a voiced segment becoming voiceless as a consequence of being adjacent to a voiceless segment; alternatively, a voiceless segment may become voiced.
2. Another type is assimilation of place: this refers to changes in the place of articulation of a segment (usually a consonant). A well-known case is that of English word-final alveolar consonants such as /t/, /d/, and /n/; if a word ending in one of these consonants is followed by a word whose initial consonant has a different place of articulation, the word-final alveolar consonant is likely to
change so that it has the same place of articulation. Thus the word that /ðæt/ may be followed by boy /bɔɪ/, which begins with a bilabial sound, and be realised as /ðæp/ (thus that boy /ðæp bɔɪ/), or it may be followed by girl /ɡɜːl/, which begins with a velar sound, and be realised as /ðæk/ (thus that girl /ðæk ɡɜːl/).

(3) A third type is assimilation of manner: here one sound changes the manner of its articulation to become similar in manner to a neighbouring sound. Clear examples of this type are not easy to find; generally, they involve a change from a “stronger” consonant (one making a more substantial obstruction to the flow of air) to a “weaker” one, and are typical of rapid speech. An English example could be a rapid pronunciation of Get some of that soap, where instead of the expected /ges sʌm ʃæt sawp/ the speaker says /ges sʌm ʃæs sawp/, with /s/ replacing /ʃ/ in two words.

An important question arises at this point, which concerns the role of the phoneme in assimilation processes. Much of the earlier writing on assimilation has suggested that assimilatory changes generally involve a change from one phoneme to another; for example, the example I have to is expressed as showing a change from /ʃ/ to /tʃ/: that girl is supposed to show final /tʃ/ changing to /k/ in /ðæk ɡɜːl/. Does this mean that all assimilations involve phonemic change of this sort? The answer must be no; we can observe many cases in which there is a clear assimilation that does not involve phonemic change. An easy process to observe is the position of the lips. In a vowel such as English /iː/ (as in see), the lips are spread, as for a smile. In a vowel such as English /s/ (as in saw), the lips are rounded and pushed forward. This spreading and rounding of the lips is quite a slow process, and it often happens that preceding and following sounds are also affected by it, even when they belong to a different word. Thus, the /s/ at the end of this will tend to have spread lips in the phrase this evening (where it precedes /iː/) and rounded lips in the phrase this autumn (where it precedes /s/). The effect is even more noticeable within a word: for example, the two /s/ sounds in see-saw, which precede /iː/ and /sː/ respectively, usually have very different lip-shapes. You can easily observe this effect in a mirror. The difference between rounded and non-rounded /s/ is not phonemic in English.

Can we always find an articulatory explanation for assimilation? These explanations seem to assume that we are basically lazy, and do as little work as possible – this is sometimes called the “principle of least effort”, and it does seem to explain a lot of human activity (or lack of it) in a very simple way. A good example is nasalization, particularly of vowels, and to understand this process we need to look at the activity of the soft palate or velum. When we produce a nasal consonant such as [m] or [n], the soft palate must be lowered to allow air to escape through the nasal cavity; however, for most vowels the velum is raised, preventing the escape of air by this route. In the English sentence I know /aɪ nɔʊ/ we would expect that if each segment were produced independently of its neighbours the soft palate would first rise for /aɪ/, then be lowered for /n/, then raised again for /nɔ/. But speech research has shown that the soft palate moves slowly and begins to make its movement some time before the completion of that movement is needed – in other words, we can see anticipation in its activity. As a result, the diphthong preceding [n] will be nasalized. We can see a more extreme example in a word like morning /mɔːnɪŋ/ where all the vowels are next to nasal consonants, and the soft palate is often left in the lowered position for the whole word, producing nasalisation of each of the vowels. In some languages, the difference between nasalized and non-nasalized vowels is phonemic, but this is not the case in English.
Coarticulation
The more deeply we look into the complexity of assimilatory processes, the more we need to move away from simple ideas like phoneme change and a single influencing neighbouring segment. This subject is of the most profound importance for understanding how speech is produced. If we want to follow recent experimental and theoretical work in this area that might help us to understand these processes, we must move on to the area of study known as coarticulation.

The name and the concept of coarticulation have been around since the 1930’s, but it remains difficult to explain or define what coarticulation is. It has been demonstrated by many experimental studies that coarticulation has effects which extend much further than just from one segment to another, so coarticulation studies have to assume a more widely spread effect of segments on each other. It is a basic principle that coarticulation is something that can be explained in physical terms, and is not arbitrary. Coarticulatory effects often extend further than just from one sound to its neighbour. For example, in the word screws /skruːz/; lip-rounding is often found extending over the whole word [s“k“uːz“]; it is actually required for the pronunciation of /uː/ and, for most English speakers, for the /t/ too, but it seems that the command to round the lips is sent to the articulators in time for the initial /s/ to be rounded, and this command will remain in effect after the end of the /uː/ so as to produce lip-rounding in the final /z/. This is not just an English characteristic: similar sound-sequences in French have been observed to behave in the same way. The French word structural contains two rounded /y/ vowels, and the lip-rounding may, again, be seen on the initial /s/ and through the word up to the beginning of the a vowel. We have already seen how the vowels in the English word morning /mɔːnɪŋ/ will tend to be nasalized as a result of the lowering of the soft palate for the nasal consonants. All languages appear to exhibit some degree of coarticulatory nasalisation of vowels adjacent to nasal consonants.

There is a lot that we still do not understand about the changes that take place when we change from slow, careful speech to rapid, conversational speech. So much research is being carried out on this subject at the present time, however, that our knowledge is growing rapidly.

What to read
Chapter 7 of this book covers a number of topics in this area, including consonant articulations with double (or “secondary”) constrictions, coarticulation and the representation of continuous speech by “parametric” diagrams.

The authors, on pages 81-90, explain how the facts of coarticulation and the effects of context affect our way of looking at phonemic analysis.
When we speak to each other, the sounds we make have to travel from the mouth of the speaker to the ear of the listener. This is true whether we are speaking face to face, or by telephone over thousands of miles. What is important for us in our study of speech is that this acoustic signal is completely observable: we can capture everything that the listener hears in the form of a recording, and then measure whichever aspect of the signal that we want to know about. There is an interesting observation to make here: for each of the phonetic classes of sound that we have identified, we can find corresponding acoustic patterns. However, if we had started by studying the types of acoustic pattern without knowing anything about how they were made by a human speaker, we would probably have set up a quite different way of classifying them. We will begin by setting out a classification of acoustic patterns, and then see how this fits with the traditional phonetic classification of speech sounds.

**Acoustic waveforms**

All audible sound is the result of variations in air pressure that produce vibration. In vibration, the pressure in a particular place (for example, inside the ear) becomes alternately higher and lower. This is usually described in terms of wave motion, using diagrams like Fig. 9.1 that suggest up-and-down movement, though sound waves do not really move up and down like waves on the sea. They are more like the shock waves that travel outwards from an explosion. We can show the pattern of a particular sort of vibration by displaying its waveform. If the vibration happens rapidly, we say it has a high **frequency**, and if it happens less rapidly, we say it has a lower frequency. If the vibration is regular, repeating its pattern over and over, we call the sound **periodic**, while a pattern of vibration which is not regular is called **aperiodic**. If the sound contains a large amount of energy, we say that it has high **amplitude**. Fig. 9.1 shows the waveform for the word *see*; the first part, /s/, is aperiodic, having an irregular, rather messy pattern, while the vowel /i:/ is periodic, and we can see a more regular pattern.
It is a fundamental principle in acoustic analysis that any waveform, however complex it might be, can be broken down into simple waveforms of different frequencies. The operation of doing this is called **spectral analysis**, and in some ways is rather like breaking down white light into the rainbow pattern of colours that make up its spectrum. In carrying out the acoustic analysis of speech sounds, we can discover much more by looking at the result of a spectral analysis than by looking at the original waveform that was captured by the microphone. This type of picture is called a **spectrogram**. At one time there was a fashion for calling such pictures “voice-prints”, but this led to some very dubious claims being made about identifying people by their voices for legal purposes, and the name is not now used except (sometimes) by gullible journalists. Here is the spectrogram of **see**.

![Spectrogram of see](image)

**Figure 9.2. Spectrogram of see**

In a spectrogram, the vertical axis of the picture represents the frequency scale: the lowest frequencies are shown at the bottom. From left to right is the time axis, with the beginning of the sound analysed shown at the left. The degree of blackness of the markings shows the amplitude at different frequencies in the signal at a particular point in time. You can see that in /s/ the energy is spread widely on the frequency range, but it is greater towards the higher frequencies and almost non-existent at the lowest frequencies. In /i:/, the energy is concentrated in three or four narrow bands (**formants**) in the lower part of the spectrum. Some spectrographic displays now show levels of energy with different colours instead, but although these look pretty and are nice to pin on your wall, most people find they are harder to interpret than the grey-scale spectrograms that have been around since the 1940’s.

There is a general theory of how the acoustic signal is produced by the human vocal tract, based on the principle that we have some way of producing sound (**source**), and for most sounds also a way of modifying that sound (**filter**). This source-filter theory is widely accepted as a fundamental concept in speech acoustics. To take vowels as an example, the source for a vowel is the vibration of the vocal folds; as the vibrating flow of air passes through the vocal tract, the vocal tract acts as a filter, weakening the energy at some frequencies while at other frequencies the energy remains relatively strong. The shape of the vocal tract (which depends on factors like the tongue-shape, the position of the lips and the position of the velum) determines the characteristics of this filter so that a particular vowel is produced; if you change the shape of the vocal tract, you change the resulting vowel.
Here is a spectrogram of the English phrase *she bought some chairs and a table* /ʃi boʊt som tʃeəz ənə tɛbl/.  

![Spectrogram Image]

Figure 9.3. Spectrogram of *she bought some chairs and a table*

**What to read**

Chapter 9 introduces fundamental concepts of acoustics and their application to the study of speech.

In Chapter 8, the characteristics of the main classes of speech sounds are explained. The chapter also covers the interpretation of spectrograms and the relationships between acoustic and traditional phonetic classification of sounds.
Technology is changing all our lives, and one aspect of technology is fast becoming a leading influence in this revolution. One of the most fundamental attributes of the human intellect is the ability to communicate by speaking, and we have at last achieved the goal of developing machines that are themselves capable of spoken communication. To many ordinary people, speech technology seems an idea out of science fiction, yet the work of scientists around the world involved in this area of technology in recent decades has resulted in products with real commercial and industrial potential at relatively low cost.

For most speech scientists, speech technology comprises two principal areas: automatic speech recognition, and speech synthesis. In principle, any area where technology is involved in the process of spoken communication should be regarded as an example of Speech Technology. Recognition and synthesis are, to a large extent, complementary fields, and might be thought of as two sides of the same coin. There are, however, major differences. We look first at the basics of speech recognition.

A. Applications of speech recognition
The most frequently quoted application for speech recognition is in office dictation systems. It is believed that there will be major economic benefits when a fully reliable system is on the market. Up until recently, users had to choose between systems which recognised a small vocabulary (one or two thousand words) reliably and with reasonably natural speaking style, or a large vocabulary (tens of thousands of words) in an unnatural speaking style in which words were separated by pauses. Now, office dictation systems which are capable of taking dictation from many speakers using a large (though not unlimited) vocabulary and more or less natural connected speech are becoming available. Such a system receives the spoken input, and produces a letter or report with proper formatting and spelling. It must be remembered that achieving correct spelling is not easy in English, and the difficulty of converting spelling to sound and sound to spelling is one of the problems that receives most effort in English-speaking countries – a problem that could be avoided if English spelling were reformed. In this context we should note that most people can speak more rapidly than they can type, so a speech-input system is likely to speed up work in some areas.

An important application area for speech technology, and one with a value that everyone can see, is in helping the disabled. There are many people who are physically unable to operate a keyboard but have the power of speech. To be able to control their environment by spoken commands (open the door, switch on the heating, operate an alarm) is be a big help to such people, and voice-operated devices can provide this.

B. Techniques in speech recognition
It is a striking fact that the most lasting developments of speech technology have been the result of partnership between specialists in computer science and electronic engineering on the one hand and specialists in speech science and linguistics on the other. Attempts to solve the many problems of speech recognition simply by advanced engineering have resulted in systems that work satisfactorily within the laboratory for an ideal speaker, but have been unable to survive exposure to the enormous variability of speech in the real world. The input of speech science has been of different types in different applications, but phonetic expertise is always an essential component of a successful system.

In earlier times, it seemed obvious that we humans, having learned what are the principal characteristics of the speech sounds that must be recognised, must then instruct the computer on how to perform the same task: such systems were known as
knowledge-based systems. But more recently, we have been able to work with computer systems that are able to learn for themselves what characteristics distinguish the various units of speech; these systems are capable of learning by themselves. In the case of knowledge-based systems, the relevant input of the speech scientist was to provide the engineer with the best possible descriptions of the data. But in self-teaching systems, the input is completely different.

If we look at how a human child learns to understand speech, we can see that the process is one of repeated exposure to the data, day after day, with regular feedback on whether understanding has taken place correctly. There is no sudden transition in the child’s learning which is equivalent to the moment when a complex computer program begins to perform correctly. The process is one of providing the computer with very large bodies of carefully prepared training data, so that it will become familiar with each particular unit (such as a phoneme or syllable) that it must learn, in every context in which it may occur, spoken by a large and representative set of speakers. If the data is badly prepared, the learning will never be successful. There has, as a result of this, been an enormous growth in the development of speech databases to be used for training (and later on for testing) recognition systems. These databases comprise carefully-controlled recordings of many speakers (sometimes recorded under a variety of conditions), and expert transcriptions of the data made in such a way that the computer can link each symbol in the transcription with a particular part of the sound recording. Thus any present-day attempt to develop an effective speech recognition system must have a suitable speech database, and if such a database does not exist, it must be created.

C. Speech synthesis applications
In talking about speech recognition, we noted that we can speak more rapidly than we can type. One disadvantage to speech synthesis as a way of providing information is that, in general, we can read a screen more rapidly than we can listen to a voice. Receiving information from a synthesiser can be frustratingly slow, so we need to look carefully to find applications where the advantages of speech output compensate for this. Clearly we should look at cases where the user’s eyes are not available. In-car information is one example which is developing rapidly: as cars become stuck in congestion more and more often, there is a growing market for systems which advise on the least congested route; indeed, many cars now come with satellite navigation as standard equipment.

Speech synthesis can also help the disabled. One of the most attractive applications is that of reading machines for the blind. A printed page is scanned, the text is converted into phonetic symbolic form and speech is synthesised. This requires a synthesis-by-rule program, and the improvement to synthesis-by-rule is probably the most important activities in this field.

Of course, speech synthesis can also help those disabled who are unable to speak. One of Britain’s greatest scientists, Professor Stephen Hawking, is only able to speak by means of a “pointer” keyboard, operated by twitching his cheek, and speech synthesiser. However, the voice you hear is an American one and, as we know, Professor Hawking is not American. Recent developments are allowing scientists to devise bespoke voices which better represent a person’s identity.

Sadly, it must be admitted that the application of speech synthesis which is most likely to make money is that of talking toys.

D. Synthesis techniques
The self-teaching processes described under speech recognition above work also for synthesis – our work on constructing speech databases has value in this field also.
There are many applications where it has been found that a completely artificial synthetic voice is not necessarily the best solution. As signal processing techniques develop, it can be more practical to manipulate “real” speech signals to generate new messages without the old problem of noticeable discontinuities in the signal where pieces of speech are joined together.

Finally, it is important to remember that while high-quality synthesis of small sections of speech is important in the speech research context, it is synthesis-by-rule which represents the commercial future of this field, and at present there is still a long way to go before the goal of truly natural-sounding synthetic speech from synthesis-by-rule is achieved. Synthesis-by-rule takes as its input written text and produces as its output connected speech.

E. Education
There is growing interest in speech technology as a way of providing additional teaching for advanced-level language learners who need practice in using the spoken language. Computer systems have been developed which give learners tasks, evaluate their spoken performance and diagnose errors, one such example being Cool Speech, an application for iPad. It is not realistic to think of these as replacing teachers, but rather as providing support for students who require additional practice outside the classroom.

F. Forensic speaker comparison
Another use for speech technology is to back up claims made by phoneticians undertaking forensic speaker comparison for the police and legal professions. Phoneticians traditionally have used their ears to decide what sounds are being produced in a stream of speech. In forensic speaker comparison, however, instrumental information – such as whether vowel formants from a spectrographic display are similar between recordings of a known speaker and someone who could be the same speaker – is essential in being able to support judgements in court.

What to read
Autumn Term 2016

The Sounds of Language

Seminars
THE SOUNDS OF LANGUAGE SEMINAR 1:
ENGLISH TRANSCRIPTION SYSTEM

English Consonants

| p | pie | f | foot, photo | m | might |
| t | pie | θ | think | n | night |
| tj | chime | s | go, cinema | η | sing, think |
| k | key, cat | Ъ | ship, nation | l | like |
| b | by | v | yan | r | red |
| d | dye | ō | this | w | will, quick |
| dʒ | just, gentle | z | zebra, rige | j | yet |
| g | go | ʒ | vision, rouge | h | high |

English Vowels

**Stressed short vowels**

- i: bead
- e: bed
- æ: bad
- ʌ: bud (some accents do not have this sound)
- o: hod
- u: hood

**Stressed long vowels**

- iː: bead
- aː: bard
- ɔː: caught
- uː: boot
- ɜː: bird

**Stressed diphthongs**

- ei: bait
- ai: bite
- ɔi: boy

A38
Stressed diphthongs
(closing, backing)

Stressed diphthongs
(centring)

Examples of triphthongs

Unstressed ‘reduced’ vowels

Stress is marked by a short vertical line immediately before the beginning of the stressed syllable of the word.
Thus, the example words in the section above will be transcribed as follows:

ɪˈmɛnsˈmjuːzɪk ɪˈmɛn ˈpɒkɪt ˈmænɪdʒ
ˈhæpi ˈriːkt ˈɛdjʊkɪt ˈɪnfləns ˈθæŋk ju
pəˈswɜːrd ˈspɔːkan əˈteɪn ˈsɔːfə ˈlʌndən
prəˈtekt ˈmeɪnəs ˈmaʊntɪn
### SOME PITFALLS IN THE TRANSCRIPTION OF CONSONANTS

<table>
<thead>
<tr>
<th>Spelling</th>
<th>Transcription</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>k</td>
<td>word-finally, before consonants, and before a, o, u before i, e</td>
</tr>
<tr>
<td>s</td>
<td></td>
<td>in official, herbaceous</td>
</tr>
<tr>
<td>j</td>
<td></td>
<td>as in machine</td>
</tr>
<tr>
<td>ch</td>
<td>j</td>
<td>as in mechanical</td>
</tr>
<tr>
<td>k</td>
<td></td>
<td>as in mechanical</td>
</tr>
<tr>
<td>ed</td>
<td>t</td>
<td>as in asked</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td>as in filled</td>
</tr>
<tr>
<td>rd</td>
<td></td>
<td>as in waited</td>
</tr>
<tr>
<td>g</td>
<td>ʒ</td>
<td>as in rouge word-finally, except –ng; before consonants, before a, o, u; sometimes before i, e, as in give, get sometimes before i, e, as in gist, gem</td>
</tr>
<tr>
<td>dʒ</td>
<td></td>
<td>as in rouge word-finally, except –ng; before consonants, before a, o, u; sometimes before i, e, as in give, get sometimes before i, e, as in gist, gem</td>
</tr>
<tr>
<td>j</td>
<td>dʒ</td>
<td>as in judge</td>
</tr>
<tr>
<td>n</td>
<td>η</td>
<td>before k and g</td>
</tr>
<tr>
<td>ng</td>
<td>η</td>
<td>as in ring</td>
</tr>
<tr>
<td>ndʒ</td>
<td></td>
<td>as in orange in other contexts (finger)</td>
</tr>
<tr>
<td>ηɡ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qu</td>
<td>k</td>
<td>in some loan words (quiche) elsewhere</td>
</tr>
<tr>
<td>kw</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>word-initially, or when doubled as in dogs, or horses, and often word-finally, e.g. does</td>
</tr>
<tr>
<td>z</td>
<td></td>
<td>as in dogs, or horses, and often word-finally, e.g. does</td>
</tr>
<tr>
<td>j</td>
<td></td>
<td>in sugar</td>
</tr>
</tbody>
</table>
TRANSCRIPTION PRACTICE

Exercise

(i) Identify the words which are written down in IPA English phoneme symbols:

\[ \text{kæt} \quad \text{kwɪk} \quad \text{wʊz} \quad \text{ʃæl} \quad \text{θeft} \quad \text{brɪŋ} \quad \text{ðɪs} \quad \text{dʒæk} \quad \text{jæk} \quad \text{fɪks} \]

(ii) Write down the sequence of phonemes which corresponds to the following words. The first one has been done for you:

- thick /θɪk/
- wig bang gnat yell six quite chop push this jump

(iii) Write, using phonetic symbols, how you would say your own name and address:
(Remember: No numbers, No capital letters, No punctuation)

(iv) Transcribe the following place-names:

(a) Leicestershire  (b) Coleraine
(c) Dunfermline  (d) Huddersfield
(e) Stevenage  (f) Halesowen
(g) Penzance  (h) Wilmslow
(i) Guildford  (j) Barnstaple
A. Word/ lexical stress and schwa
Stress is basically where words or syllables stand out more than surrounding ones, giving extra prominence. In English, stressed syllables are usually louder, longer and higher in pitch than unstressed ones. Differences in stress can signal a difference in meaning.

There are two types of stress in English: lexical stress and sentence stress. Let’s start with lexical stress. For example, in the word *mother*, which has two syllables, the first syllable is stressed, as it stands out more than the second syllable. In *philanthropy*, a four syllable word, the second syllable is stressed. And in *topological*, a five syllable word, the first and third syllables are stressed, with the third being most prominent.

The vowel /ə/, or “schwa”, is the most common in English speech. It is a weak vowel, never appearing in stressed syllables in RP / BBC English. A word may have more than one syllable, one or more of which contains a schwa. E.g., in the four syllable word *philanthropy* the second is stressed, and the third is unstressed and contains schwa. Single syllable function (or grammar) words – pronouns, conjunctions, determiners, auxiliary verbs, etc. – commonly use this vowel, too, although not all of them do.

Exercises
(i) To work out where the lexical stress is on a word, try to put stress on each syllable in the following set of words until you can decide which sounds right. Where is the stressed syllable? Is there more than one? We mark a stressed syllable with a small raised vertical line to the left of the syllable, e.g. 'teacher edu'cation.

<table>
<thead>
<tr>
<th>camera</th>
<th>conference</th>
<th>identity</th>
<th>massive</th>
<th>crafty</th>
</tr>
</thead>
<tbody>
<tr>
<td>dispute</td>
<td>balloon</td>
<td>clinical</td>
<td>linguistics</td>
<td>filing</td>
</tr>
<tr>
<td>employee</td>
<td>information</td>
<td>clinician</td>
<td>lecturer</td>
<td>Japanese</td>
</tr>
</tbody>
</table>

(ii) The following words all contain /ə/ in one or more syllables. Identify which syllables contain /ə/, and which contain a different vowel. Find the other vowels in the list we used last time. Where is the stressed syllable in each word?

| about | banana | photograph | photography | motorboat |
B. Sentence stress
When we talk about sentence stress, we are interested in which syllables are prominent in a stream of speech, not in individual words. Unless the speaker is being particularly emphatic, not every word is stressed. Function words are not usually stressed, but can be for particular effect. For example, in the following sentences, *my* in (1) and *your* in (2) – both personal pronouns – are not likely to be stressed, but in (3) they are both stressed as they show contrast (underlining is used here as an orthographic convention).

(1) I like cycling to work on my bicycle.
(2) Can you give me a lift in your car this morning?
(3) No, I said *your* car, not *my* car.

Exercises

(i) Working in pairs or small groups, get a classmate to say one of these sentences. Where are the sentence stresses in his/her production? Mark them with a raised vertical line, or underline them.

a) He really didn’t want to eat the last cake but could see there was no choice.
b) That lovely picture on my cabinet was blown into the waste bin by the wind yesterday.
c) I had hoped to see Paul at the party but clearly he was much too busy.

(ii) Say these sentences, moving the stress around as you do. What is the linguistic effect of doing this?

a) I didn’t want to let her words make me feel annoyed.
b) The root of the problem is that I forgot to take enough money out of the cashpoint yesterday.
c) If you listen closely enough you may be able to hear the cat purring.
THE SOUNDS OF LANGUAGE SEMINAR 3:
ALLOPHONES AND PHONEMES IN ENGLISH AND OTHER LANGUAGES

A. English
The following data is from English. The phones [t], [tʰ] and [t̚] are allophones of the phoneme /t/. Examine the data and decide what the conditioning factors are (i.e., when each of these allophones is produced).

[tʰɛn] ten [pʰɒt] pot
[stept] step [stiːm] steam
[rat] rat [tʰɔk] talk
[kʰæt] cut [tʰɔust] toast

B. Korean
In the following Korean words, you will find the sounds [s] and [ʃ]. Determine whether the sounds [s] and [ʃ] are allophones of the same phoneme or separate phonemes. If the sounds are allophones of the same phoneme, give the environment in which they each occur.

2. [miʃin'] superstition 12. [tʃasal] suicide
5. [ʃilsu] mistake 15. [kæsil] hypothesis
7. [ʃafin'] self 17. [miso] smile
10. [kafij] thorn 20. [so] cow
C. **Spanish**
The data below shows allophones of the voiced obstruents: [b] and [β] (voiced bilabial fricative) are allophones of one phoneme, likewise [d̪] (voiced dental stop) and [ð] (voiced dental fricative), likewise [ɡ] and [ɣ] (voiced velar fricative).

a) What is the phonetic context in which each sound occurs? Do any patterns emerge?
b) Which allophone is the principal member of the phoneme in each case?
c) If there is time, consider the relationship between sounds [ɾ] and [ɾ].

<table>
<thead>
<tr>
<th>No.</th>
<th>Word</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>[aβa]a</td>
<td>Havana</td>
</tr>
<tr>
<td>2.</td>
<td>[bala]</td>
<td>ball</td>
</tr>
<tr>
<td>3.</td>
<td>[baya]</td>
<td>rope</td>
</tr>
<tr>
<td>4.</td>
<td>[boða]</td>
<td>wedding</td>
</tr>
<tr>
<td>5.</td>
<td>[buro]</td>
<td>burro</td>
</tr>
<tr>
<td>6.</td>
<td>[ɗamos]</td>
<td>we give</td>
</tr>
<tr>
<td>7.</td>
<td>[ɗios]</td>
<td>God</td>
</tr>
<tr>
<td>8.</td>
<td>[ɗeβer]</td>
<td>to owe</td>
</tr>
<tr>
<td>9.</td>
<td>[ɗon̪de]</td>
<td>where</td>
</tr>
<tr>
<td>10.</td>
<td>[ɗdurar]</td>
<td>to endure</td>
</tr>
<tr>
<td>11.</td>
<td>[ɡaɾar]</td>
<td>to earn</td>
</tr>
<tr>
<td>12.</td>
<td>[ɡaɾo]</td>
<td>cat</td>
</tr>
<tr>
<td>13.</td>
<td>[ɡola]</td>
<td>throat</td>
</tr>
<tr>
<td>14.</td>
<td>[ɡosar]</td>
<td>to enjoy</td>
</tr>
<tr>
<td>15.</td>
<td>[kasa]</td>
<td>house</td>
</tr>
<tr>
<td>16.</td>
<td>[kuβa]</td>
<td>Cuba</td>
</tr>
<tr>
<td>17.</td>
<td>[layo]</td>
<td>lake</td>
</tr>
<tr>
<td>18.</td>
<td>[ŋaða]</td>
<td>nothing</td>
</tr>
<tr>
<td>19.</td>
<td>[ŋuðo]</td>
<td>knot</td>
</tr>
<tr>
<td>20.</td>
<td>[pero]</td>
<td>but</td>
</tr>
<tr>
<td>21.</td>
<td>[pero]</td>
<td>dog</td>
</tr>
<tr>
<td>22.</td>
<td>[pipa]</td>
<td>pipe</td>
</tr>
<tr>
<td>23.</td>
<td>[poŋd̪eroso]</td>
<td>heavy</td>
</tr>
<tr>
<td>24.</td>
<td>[poŋgo]</td>
<td>I put</td>
</tr>
<tr>
<td>25.</td>
<td>[rumbo]</td>
<td>direction</td>
</tr>
<tr>
<td>26.</td>
<td>[siyaro]</td>
<td>cigar</td>
</tr>
<tr>
<td>27.</td>
<td>[țeŋgo]</td>
<td>I have</td>
</tr>
<tr>
<td>28.</td>
<td>[țoðo]</td>
<td>all</td>
</tr>
<tr>
<td>29.</td>
<td>[țaβako]</td>
<td>tobacco</td>
</tr>
<tr>
<td>30.</td>
<td>[uβa]</td>
<td>grape</td>
</tr>
</tbody>
</table>

1. The following spectrogram contains two words. One is *seed* and one is *seat*. Which is which? What is the duration (length) of the vowel in each case? What is the ratio between the two lengths?

2. The following spectrogram contains the words or phrases "mashes", "matches", "a shop" and "a chop". Can you identify which is which? Use phonetic symbols to explain your decision.
3. The diagram below contains the waveform (above) and pitch-track (below). Each of the three words in it is one of the group of Spanish words "término", "termino", "terminó" (see Handbook, Page 42). Try to identify which word is which.
The Assignments

The assignments are formally “set” in the 8th week of the appropriate term; by this we mean that at that stage you will have covered enough ground to make a start on them, and to discuss them with your seminar tutors. This does not prevent you from preparing them earlier.

Assignments should be submitted in hard copy as follows:

Where?

- Via the drop boxes at the Support Centre in the Edith Morley Building

By when?

- 12 noon on the first Wednesday of the following term – i.e.,
  - Assignment 1 (Sounds of Language) by 12 noon, Wednesday Week 1, Spring term;
  - Assignment 2 (Grammar and Meaning) by 12 noon, Wednesday Week 1, Summer term

Further instructions:

- Use ONE copy of the appropriate coversheet.
- Assignments should be word processed unless otherwise stated.
- We do not accept assignments by e-mail.

Assignment Cover Sheets

Coversheets for the assignments are available in your hard copy handbooks, on the departmental website and on Blackboard. You should fill in a coversheet and attach it to your assignments when you submit. Make sure you fill in your details in the box in the top right hand corner. The administrator will seal this down to ensure anonymous marking.

Do not submit any assignment without a cover sheet!
Assignment 1 (The Sounds of Language): Answer all questions below, and submit as one single assignment. Each section is worth 25%.

1. Transcription. Write on this sheet.
   a. Transcribe the following words and phrases using English phoneme symbols, based on an RP/BBC English accent.

   crutch

   we have to

   last time

   Gloucestershire

   it’s raining cats and dogs

   b. The following transcribed English words and phrases based on an RP/BBC English accent have between one and two errors each. Circle the error and transcribe the word or phrase out IN FULL using English phoneme symbols.

   cap

   mægɪk

   we laɪk thə sʌnʃain

   braʊm boots

   laɪs tɜːm wi stʌdɪd phənetɪks

RP/BBC English has one phoneme /t/ and one phoneme /d/, but these have distributional variants, or allophones, depending on where they occur in a word. Describe the patterns of distribution of the allophones of /t/, which in these data are [t], [tʰʷ] and [tʰ], and the allophones of /d/, which are [d], [d̥] and [d̥ʷ], based on the following data.

\[
\begin{array}{ll}
[tʰː] & \text{two} \\
[lædə] & \text{ladder} \\
[d̥wɔː] & \text{door} \\
[tiːn] & \text{train} \\
[tʰɪːm] & \text{team} \\
[θɛɪ] & \text{day} \\
[stɛi] & \text{stay} \\
[θɛɪt] & \text{date} \\
[loːd] & \text{lord} \\
[stiːm] & \text{steam} \\
\end{array}
\]

Give your answers in point form, e.g.:

- [tʰ] occurs …
- [tʰʷ] occurs …

3. Phonetic vs. phonological inventories. Word process on a separate sheet.

The following two languages have identical phonetic inventories, but different phonological inventories. Give the phonetic inventory the two languages share (consonants and vowels), and then the phonological inventory for each language. Indicate the different patterns which occur in each language. Give your answers in point form.

Language 1: Henmanese

\[
\begin{array}{ll}
[kap] & \text{yes} \\
[gap] & \text{no} \\
[kip] & \text{today} \\
[gip] & \text{tomorrow} \\
[kek] & \text{peace} \\
[pek] & \text{yellow} \\
\end{array}
\]

Language 2: Murryish

\[
\begin{array}{ll}
[kap] & \text{hello} \\
[pak] & \text{sandwich} \\
[kib] & \text{box} \\
[gib] & \text{box} \\
[peg] & \text{strawberry} \\
[bek] & \text{strawberry} \\
\end{array}
\]

[|kap|] and *[|pek|] are illegal in Henmanese.

NB Syllables such as *[kab] and *[pek] are illegal in Henmanese.
4. Reading spectrograms. Word process on a separate sheet.
The following spectrograms show a female speaker saying the phrases ‘a catch’, ‘a cat’, ‘adjust’, ‘a dust’. Decide which spectrograms match these phrases. Try to explain how you came to your conclusions.

Spectrogram 1

Spectrogram 2
Spectrogram 3

Spectrogram 4

Colour versions of these spectrograms are available on Blackboard.