Natural Language Processing & Applications

Phones and Phonemes

1 Phonemes

If we are to understand how speech might be generated or recognized by a computer, we need to study some of the underlying linguistic theory. The aim here is to UNDERSTAND the theory rather than memorize it. I've tried to reduce and simplify as much as possible without serious inaccuracy.

Speech consists of sequences of sounds. The use of an instrument (such as a speech spectrograph) shows that most of normal speech consists of continuous sounds, both within words and across word boundaries. Speakers of a language can easily dissect its continuous sounds into words. With more difficulty, they can split words into component sounds, or 'segments'. However, it is not always clear where to stop splitting. In the word *strip*, for example, should the sound represented by the letters *str* be treated as a unit, be split into the two sounds represented by *st* and *r*, or be split into the three sounds represented by *s*, *t* and *r*?

One approach to isolating component sounds is to look for 'distinctive unit sounds' or **phonemes**.[1] For example, three phonemes can be distinguished in the word *cat*, corresponding to the letters *c*, *a* and *t* (but of course English spelling is notoriously non-phonemic so correspondence of phonemes and letters should not be expected). How do we know that these three are 'distinctive unit sounds' or phonemes of the English language? NOT from the sounds themselves. A speech spectrograph will not show a neat division of the sound of the word *cat* into three parts. Rather we know these are phonemes because BOTH of the following are true:

- The three are 'unit' sounds. A different English word cannot be formed by replacing part of the *c* sound and part of the *a* sound by a different sound. The whole of a phoneme must be replaced to make a valid English word. Thus the *c* sound in *cat* is a 'unit' sound because it can be removed entirely to change *cat* into *at*, or replaced entirely by a *b* sound to change *cat* into *bat*.
- The three are 'distinctive' sounds. Changing a single phoneme in *cat* is sufficient to make a word which is recognizably different to a speaker of English. The words *bat*, *kit* and *cad* are each minimally different from the word *cat* but are recognizably different words to an English speaker.

In summary, a phoneme is defined as a 'distinctive unit sound' of a language: 'unit' because the whole of a phoneme must be substituted to make a different word; 'distinctive' because changing a single phoneme can generate a word which is recognizably different to a speaker of the language.

Note that 'phoneme' is a subjective concept, not an objective one. To test whether a particular sound operates as a phoneme in a language we cannot use instruments such as speech spectrographs. Rather we have to ask a speaker of the language whether removing that sound from a word and substituting another (preferably one already known to be a phoneme of the language) generates a new word (or what could be a new word if the new sound sequence doesn't exist in the language).

Since English spelling in particular is non-phonemic, we need some way of consistently representing phonemes. I will use IPA (International Phonetic Alphabet) symbols where appropriate. You are NOT expected to learn these; a table (see Appendix) will be given if and when required. By convention, phonemic representations of sounds are enclosed in slashes. Thus the English words discussed earlier, *cat*, *bat*, *kit* and *cad*, can be represented phonemically as /kæt/, /bæt/, /kɪt/ and /kæd/. Comparing 'minimal pairs' confirms that /k/, /b/, /æ/, /ɪ/, /t/ and /d/ are indeed English phonemes; e.g. /æ/ is a phoneme because in the word *cat* it can be substituted by /ɪ/ to make the word *kit*. (Note that these six might or might

[1] I've noticed that a common mistake in reproducing this definition in examinations is to replace *distinctive* by *distinct*. Don't! *Distinctive* here refers to the ability of a phoneme to make distinguish between words; *distinct* would just mean that the phonemes were different, which isn't the same.

not be phonemes in another language.)

It’s important to note what I have NOT said. I have not said that a phoneme corresponds to a specific sound. Indeed it does not. No two individuals pronounce the English /k/ phoneme in exactly the same way – for one thing their vocal tracts are of different shapes. Neither does an individual produce exactly the same sound on different occasions. More importantly, the pronunciation of a phoneme is affected by its neighbours in a word. For example there is a consistent difference between the pronunciation of the /k/ phoneme in *cat* and its pronunciation in *kit*. In normal speech phonemes ‘run together’. One consequence is that because /æ/ is pronounced further back in the throat than /ɪ/, any preceding /k/ will be as well. A phoneme of a language represents a CLUSTER of similar sounds which a speaker of that language does not regard as distinctively different from one another. I will return to this issue later.

2 Production of Phonemes

Remembering that a phoneme represents a cluster of sounds treated in some sense as equivalent by speakers of a given language, some 40-odd phonemes can be distinguished in most dialects of English. Although all the sounds corresponding to a phoneme may not be produced in exactly the same way, for each phoneme we can describe the ‘typical’ way in which it is produced.

The sounds corresponding to all English phonemes are powered by lung air being pushed out. A sound is then produced in two ways:

- By vibrating the vocal ‘cords’: two muscular folds of skin low down in the throat which can be made to vibrate. The frequency of the vibration can be changed (within limits).
- By altering the positions of components of the throat and mouth between the vocal cords and the exit of air. These alterations may merely modify the note produced by the vocal cords (by changing the size of the cavity) or may themselves produce a noise (for example by causing air friction).

Vowels When lung air passes over the vibrating vocal cords and then passes freely out of the mouth, the sounds are called **vowels**. Thus vowels can be continued until you run out of breath. The positions of the lips and tongue alter the size and shape of the resonating cavity to produce different sounds.

Vowels can be classified along a number of independent directions, including:

- The height of the tongue (i.e. the size of the smallest opening).
- The part of the tongue (front to back) causing the smallest opening.
- The degree of lip rounding (open to rounded).

Some examples in ‘Standard English English’ (SEE):
/i/ is a high, front, unrounded vowel, as in *b**ee**t* /bit/ or *n**ea**t* /nit/.[2]
/ɑ/ is a low, back, unrounded vowel, as in *b**ar*** and *b**a**th*.
/u/ is a high, back, rounded vowel, as in *sp**oo**n*.

	Front	*Mid*	*Back*
High	/i/ *b**ea**t*		/u/ *b**oo**t*
	/ɪ/ *b**i**t*		/ʊ/ *p**u**t*
Mid		/ə/ ***a**bout*,	
	/ɛ/ *b**e**t*	***Ber**t*, *sof**a***	/ɔ/ *b**ough**t*
Low	/æ/ *b**a**t*	/ʌ/[3] *b**u**t*	/ɒ/ *p**o**t*
		/ɑ/ *b**ar***	

In English, back vowels (other than the very lowest) are automatically rounded, front and

[2] A more precise representation of this phoneme is /iː/, where the ː shows length. I will generally ignore length distinctions; the theory presented here is intended to be the minimum necessary.

[3] Although it is traditional to use /ʌ/ for this English phoneme, strictly this is not the correct IPA symbol.

mid vowels are not, so that a classification needs only two dimensions, as in the table below. Note that 'Standard English English' (SEE) pronunciation is intended.

In addition to these 'pure' vowels, English makes considerable use of **diphthongs**: sequences of two vowels 'run together' to form a SINGLE phoneme. A diphthong may include vowels not normally found alone. Examples of SEE diphthongs are given in the following table.

/eɪ/ *baby, wait, day*	/oʊ/ *bone, soap, no*	/ɪə/ *ear, cheer* [4]
/aɪ/ *kite, cry*	/aʊ/ *cow, out*	/ɛə/ *air, share*
/ɔɪ/ *coin, toy*		/ʊə/ *tour*

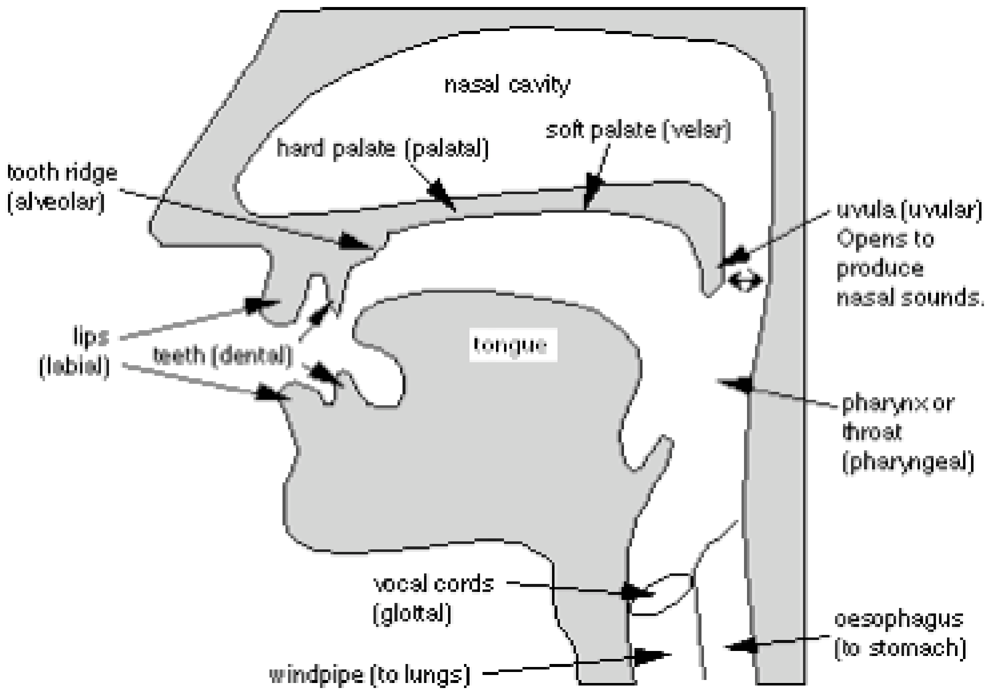
In principle, vowels are infinitely variable as the position of the tongue and lips can be varied continuously. (This makes learning to make the correct vowel sounds in a foreign language very difficult.) Languages tend to use different sets of more-or-less distinct vowels. English dialects vary greatly in the vowel phonemes used. In particular, American English differs considerably from English English. This is relevant because the English speech synthesis software currently available is often based on Standard American English (SAE). The main difference is that in SAE the back vowels /ɔ/, /ɒ/ and /ɑ/ are usually replaced by /a/, so that the vowels in *taught*, *tot* and *tart* are all pronounced as /a/ – the first sound in the SEE diphthong /aɪ/. (However, /ɔ/ is retained before /r/). The pure vowels /e/ and /o/ may be substituted for the SEE diphthongs /eɪ/ and /oʊ/. Further, most Americans at least partially pronounce *r* sounds which SEE speakers omit. Thus *ear* in SAE is closer to /ɪr/ rather than SEE /ɪə/.

Stops By contrast with vowels, some sounds are made by completely stopping and then releasing the flow of air out of the mouth. These sounds are called stops (or plosives). In SEE there are three stop positions, corresponding to the initial phonemes in *pale*, *tale* and *kale*. The sound is stopped respectively by the lips (bilabial), by the front of the tongue and the ridge behind the top teeth (alveolar), and by the back of the tongue and the soft palate (velar).

The phonemes at the start of *bale*, *dale* and *gale* involve exactly the same stop positions as the three above. One significant difference is that in these three phonemes the vocal cords vibrate during the production of the sound.

Thus /p/, /t/ and /k/ are voiceless stops; /b/, /d/ and /g/ voiced stops.

Nasals If air is allowed to flow out of the nose while being stopped in the mouth, the result is nasal stops or nasals. English has three such phonemes, corresponding to same three stopping positions as ordinary stops. These are the three phonemes at the end of *rum* /m/, *run* /n/ and *rung* /ŋ/. Nasals are (almost always) voiced.



[4] The diphthongs in this column can also be treated as sequences of two separate phonemes.

	Bilabial: Upper & lower lips	**Labio-dental**: Upper teeth & lower lip	**Dental**: upper teeth & tongue tip	**Alveolar**: tooth ridge & tongue tip	**Palatal**: hard palate & middle of tongue	**Velar**: soft palate & back of tongue	**Glottal**: at the very base of the throat
Stop (Voiceless/ Voiced)	/p/ ***p***_ale_ /b/ ***b***_ale_			/t/ ***t***_ale_ /d/ ***d***_ale_		/k/ ***k***_ale_ /g/ ***g***_ale_	
Nasal (Voiced)	/m/ _ru_***m***			/n/ _ru_***n***		/ŋ/ _ru_***ng***	
Fricative (Voiceless/ Voiced)		/f/ ***f***_at_ /v/ ***v***_at_	/θ/ _hea_***th*** /ð/ _hea_***th***_en_	/s/ ***s***_ip_ /z/ ***z***_ip_	/ʃ/ _me_***sh*** /ʒ/ _mea_***s***_ure_		/h/ ***h***_ot_
Affricative (Voiceless/ Voiced)				/tʃ/ ***ch***_ur_***ch*** /dʒ/ ***j***_u_***dg***_e_			

Fricatives If air is not completely stopped from flowing out of the mouth, but made to pass through a narrow passage, a ‘friction’ sound or fricative is produced (i.e. a more-or-less ‘hissing’ sound). In SEE there are four positions where the narrowing can be made. These correspond to the voiceless terminal phonemes in _li_***f***_e_ /f/, _brea_***th*** /θ/, _hi_***ss*** /s/ and _me_***sh*** /ʃ/. If the vocal cords vibrate as well, we have the corresponding voiced phonemes in _li_***v***_e_ /v/, _brea_***th***_e_ /ð/, _hi_***s*** /z/ and _mea_***s***_ure_ /ʒ/.

The /h/ sound in *hot* is also a kind of fricative, although no narrowing is involved. Instead air friction is produced low down in the throat. /h/ can be classified as a voiceless glottal fricative.

Affricatives The phonemes that begin and end *church* and *judge* are voiceless and voiced affricatives respectively, composed of a very fast combination of a stop and a fricative; thus *ch* = *t* + *sh* /tʃ/; compare *why choose?* with *white shoes*, both spoken rapidly. The *j* in *judge* can be written phonetically as /dʒ/. Note that affricatives are SINGLE phonemes.

Stops, nasals, fricatives and affricatives can be arranged in the neat table above.[5]

Approximants In these phonemes the tongue partly closes the airway, but not enough to cause a fricative. The phonemes that begin *lap* /l/ and *rap* /r/[6] are sometimes called **liquids**. Both are produced in the alveolar/palatal region and are normally voiced. (They differ in whether air passes by the side or over the centre of the tongue.)

Two further approximants are the initial phonemes in *woo* /w/ and *you* /j/. These are sometimes called **glides**: the tongue and lips move during the production of the sound. Both are normally voiced. Liquids and glides (especially the latter) have some similarities with vowels.

The phonemes described above are NOT an exhaustive set for all dialects of English. For example, the ‘Cockney’ pronunciation of words like *bottle* includes a ‘glottal stop’ /ʔ/.

Remember that a phoneme in the preceding tables really represents a set of closely similar sounds. For example a /t/ can be made with the tip of the tongue in different positions on the tooth ridge (say *tea trip* quickly and notice the position of your tongue during the /t/ sounds).

As noted earlier, different languages use different sets of phonemes. Filled cells in the table above may be missing: for example, French has no dental fricatives, making it difficult for French-speakers to pronounce ***th***_ink_ or ***th***_at_. (They may produce something like *sink zat* instead of *think that*. Why?) Empty cells may be filled. Many languages have dental stops (e.g. Spanish, the Indic languages); Spanish – at least in the dialect spoken around Madrid – has bilabial fricatives; German and some dialects of English (Scottish, Liverpool) have velar

5 Note that this table is simplified. E.g. ‘palatal’ can be subdivided into ‘post-alveolar’, etc.

6 This symbol should really be /ɹ/; strictly speaking /r/ is the IPA symbol for a Scottish ‘trilled’ *r*.

fricatives as in ***Bach*** or *lo**ch***; German has the labial affricative /pf/. Other languages have phonemes which won't fit into the table. For example, some southern African languages use 'click' sounds in speech, similar to the sound sometimes used to tell horses to 'gee up'; these sounds are not powered by air expelled from the lungs.

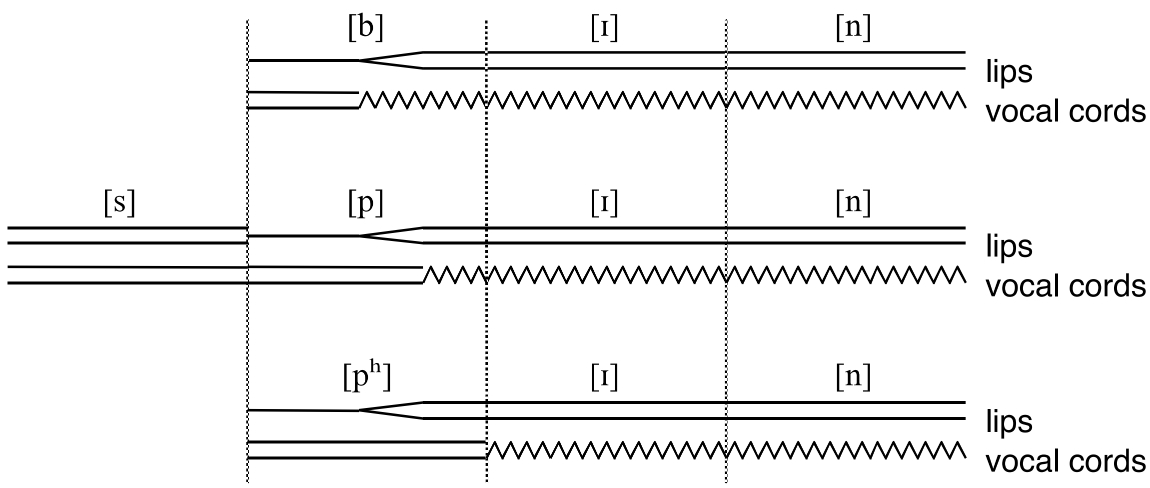
3 Phones

I pointed out earlier that the sounds corresponding to the letter *t* in the English words *tea* and *trip* are not in fact quite the same. The position of the tongue is slightly different, which causes a difference in sound detectable by an instrument such as a speech spectrograph. Thus the English phoneme /t/ corresponds to at least two different **phones**. By convention, phones are written using phonetic symbols enclosed in square brackets. There aren't any standard IPA symbols for the *t* phones in *tea* and *trip*, so I will use [t_v] for [t] before a vowel phone and [t_r] for [t] before [r]. Each of the set of phones which correspond to a single phoneme is called an **allophone** of that phoneme. Thus [t_v] and [t_r] are allophones of the English /t/ phoneme. The words *Dee* and *drip* show that /d/ similarly has at least two allophones, [d_v] and [d_r] say.

A phone can be defined as a 'unit sound' of a language. It is a 'unit' sound because the whole of the phone must be substituted to make a different word. [t_v] is a unit sound in English, and hence a phone, because the whole of it must be replaced by [d_v] to change *tea* into *Dee*. [t_r] is a phone because it must be replaced entirely by [d_r] to change *trip* into *drip*. However, [t_v] and [t_r] are not DISTINCTIVE unit sounds (and hence are not phonemes) because there are no English words in which the only difference is that [t_v] is replaced by [t_r]. If you try saying *tea* with the [t_r] allophone, you just get a slightly odd pronunciation of *tea*, not a new word.

The problem with the concept of a phone (and hence of an allophone) is that its boundaries are blurred. Although it's hard for English speakers to hear the difference between [t_v] and [t_r], it's fairly clear that a slightly different tongue position is involved. How about the /t/ sounds in *tea* and *tart*? A speech spectrograph may show a difference, but here the tongue position differs only very slightly (if at all). Are these different allophones?[7]

A clearer example of English allophones occurs in the words ***pin*** and ***spin***. If you are a native English speaker and you hold your hand in front of your mouth as you say *pin* and *spin* you will notice that the *p* in *pin* is accompanied by a short burst of air, i.e. is **aspirated**. The aspirated sound can be written as [p^h], the unaspirated as [p]. The words *bin*, *spin* and *pin* involve the closely similar English phones [b], [p] and [p^h]. They differ in the time difference between the release of lip closure and the start of vocal cord vibration (voicing). The somewhat idealized diagram below shows that in [b], voicing begins at lip opening; in [p], voicing begins very soon after lip opening; in [p^h], voicing is delayed (hence a significant puff of air escapes via the open vocal cords and lips).



7 The more neutral term 'segment' is used by many authors to avoid the need to make essentially theoretical distinctions in contexts where this issue isn't relevant.

Three important points can be illustrated by these three phones. Firstly, PERCEIVED differences between phones do not depend on actual differences. A sound spectrograph shows clearly the differences between [b] and [p] and between [p] and [p^{h}]. Objectively these three phones are easily recognizable. Yet English speakers normally notice only the first difference. Why? The answer is that the phones [p] and [p^{h}] are not DISTINCTIVE in English since they are allophones of the same /p/ phoneme. There are no two words whose only difference is that [p] is replaced by [p^{h}], so English speakers don't need to learn to distinguish them. To an English speaker [p] and [p^{h}] represent the 'same sound', even though they are actually different.

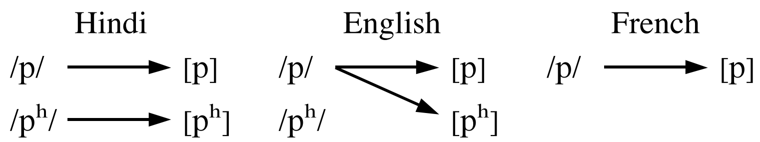
Put another way, the discreteness of phonemes is a property of the listener, not the sound. By generating speech sounds artificially, it is possible to vary the time between lip opening and the beginning of vocal cord vibration. It is found that if this time is less than 30 milliseconds, English speakers hear /b/; more than this and they hear /p/. Only very close to the boundary time do they sometimes report hearing /b/, sometimes /p/. (Neural networks offer a good model of how continuous patterns of variation can be converted into discrete decisions.)

Secondly, phoneme boundaries vary between languages. Both German and French have distinct /b/ and /p/ phonemes (i.e. in both languages there are words whose only difference is the replacement of one of these phonemes by the other). However, the precise boundary between the allophones of /b/ and those of /p/ is different. In the case of the /b/ phoneme, some German speakers tolerate a longer interval between lip opening and vocal cord vibration beginning than do most French speakers. Thus some German speakers producing what to them is /b/ appear to French speakers to be producing /p/. One illustration of this is that when the French novelist Balzac wanted to show that the speaker was German he wrote, for example, "Eh pien" instead of "Eh bien." It cannot be assumed that either vowel or consonant phones are the same in languages that appear to have the same phonemes. This complicates multilingual speech recognition. Infants presumably learn to make language-specific distinctions.

Thirdly, phonemes are specific to languages.

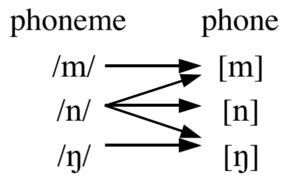
- In the Indic languages, aspirated and unaspirated stops represent different phonemes (although the difference in aspiration is greater than that in the English words *pin* and *spin*). In Hindi for example, /p/ and /p^{h}/ are different phonemes written using different characters (प and फ respectively). A Hindi speaker would be expected both to notice the difference between [p^{h}ɪn] and [pɪn] and also to be able to produce both sound sequences.
- In English, there is only one /p/ phoneme, but it has two allophones, [p^{h}] and [p], used in different positions within words. Thus an English speaker will find it difficult to hear any difference between [p^{h}ɪn] and [pɪn], but will consistently produce [p^{h}ɪn] for *pin* and [spɪn] for *spin.*
- The phone [p^{h}] does not occur in French, i.e. is not an allophone of the French phoneme /p/. Hence native French speakers are likely to produce [pɪn] for [p^{h}ɪn] when speaking English, just as native English speakers are likely to produce [p^{h}əti] for [pəti] (*petit)* when speaking French.

We have the following phoneme:phone relationships in Hindi, English and French.



As a further example, /l/ and /r/ are different phonemes in English, each having a number of allophones. But in Korean and Japanese there is only one phoneme. Speakers of these languages do not normally notice the difference between sounds which English speakers divide into /l/ and /r/ phonemes. Thus a Japanese or Korean speaker may produce [gʌri] for the English word *gully.*

The discussion so far implies that the phoneme:phone relationship in a given language is 1:many. In fact it can be even more complex. Consider the prefix *in* (with its 'direction' meaning). The *in* in *input*, *intake* and *income* is clearly the 'same' element, so we would like to represent it by the same phonemes in each case, i.e. /ɪn/. In slow, careful speech the corresponding pronunciation is [ɪn]. Many speakers of Scottish English seem to be able to maintain this pronunciation in fast speech. But in fast SEE, these words will normally be pronounced [ɪmpʊt], [ɪnteɪk] and [ɪŋkʌm]. So in this dialect of English we have the relationship shown below, with the phone [m], for example, derived from more than one phoneme.



A question I asked but avoided answering earlier was how many allophones we need to distinguish for each phoneme. Because phones ‘blend’ in continuous speech, there is likely to be some variation in the sound corresponding to a phoneme in every different environment in which it occurs. For example, in *keen* and *card*, the /k/ phoneme is clearly affected by the following vowel, being pronounced further forward in the mouth in *keen* than in *card*. What about the /k/ phoneme in *keen* and *kin*? Do we need a separate allophone of /k/ for every possible following vowel sound?

In the context of speech synthesis or recognition, how far we divide up a phoneme into allophones is a purely pragmatic issue. In speech recognition, we may only need to recognise phonemes, ignoring all allophone distinctions. In speech synthesis, a large number of allophones may be needed for good quality, since although comprehensible speech can be synthesised using relatively few allophones, it will not sound very natural. For example, native English speakers will recognize both [p^hɪn] and [pɪn] as the word *pin*, and both [spɪn] and [sp^hɪn] as the word *spin*, even though they would normally only produce [p^hɪn] and [spɪn]. But they are likely to be aware that there is ‘something not quite right’ about the synthesised speech, although they may not be able to identify it. Currently, the best speech synthesisers use several hundred phones to cover all the different allophones of (American) English phonemes.

4 Phonological Rules

To summarize the previous sections, native speakers of a language hear phonemes but speak (allo)phones. An important issue, then, is how the interconversion of phonemes to phones takes place: how is it that native speakers of English hear [p^h] and [p] as the same but yet produce these phones quite distinctly to a native speaker of an Indic language? An answer which has the great merit of being programmable is that different **phonological rules** are being used.

Phonological rules describe the relationships between phonemes and phones. For example, the following is a possible (but incomplete) rule for English:[8]

> A voiceless stop at the beginning of a word is aspirated when followed by a vowel.

In NLP, the important question is whether we can make these rules sufficiently detailed so that they can be programmed. The answer to this question is yes. One general approach is:

- First, describe phonemes/phones by **sets of features**.
- Second, write **rules** which describe changes in features based on the left and right contexts.

For example, English consonants can be described using the set {Type, Position, Voicing, Aspiration}. [p^h] is then {stop, bilabial, voiceless, aspirated}.[9] The ‘best’ set of features is a pragmatic question for computer scientists (although of great theoretical interest to linguists). The table in Appendix 1 shows a minimal set (it omits aspiration). As always in this module, note that I have tried to simplify as much possible in order to show the principles without

---

[8] Incomplete because voiceless stops are also usually aspirated at the beginning of *stressed syllables*. Thus *upon* is pronounced [ə=p^hɒn], where = marks the syllable boundary.

[9] Throughout I’m using the standard convention, derived from Prolog, that an initial capital letter marks a variable, an initial lower-case letter marks a constant. This is a ‘set’ in that elements may be omitted; however it’s usual to keep the same ordering,

excessive detail. The sets of features normally used by linguists are considerably more complicated than those discussed here.

It appears to be an empirical fact about English (but not necessarily other languages) that at most one element before and one after a given phoneme determines which of its allophones is chosen. So the rule "a voiceless stop at the beginning of a word is aspirated when followed by a vowel" can be written as:

word-boundary voiceless-stop vowel → word-boundary aspirated-voiceless-stop vowel

Using the feature set notation makes such rules even more explicit. The rule above can be written as:

word-boundary {stop, voiceless} {vowel} →
word-boundary {stop, voiceless, aspirated} {vowel}

The general format of a rule is:

left-context input right-context → left-context output right-context

In order to avoid repeating the unchanging left and right contexts, slightly different notations are often used. One such is:

input → output : left-context _ right-context

Thus the rule that a voiceless stop at the beginning of a word is aspirated when followed by a vowel can be written as:

{stop, voiceless} → {stop, voiceless, aspirated} : word-boundary _ {vowel}

In this notation, the information after the ':' gives the context, the information before gives the change required by the rule.[10] Either of the left or right contexts can be omitted if irrelevant.

Note that I have written the rule as it would be used in the production of speech, i.e. converting phonemes to phones. Phonological rules can, in principle, be used in either direction. In recognition, the rule could be used 'backwards', to convert (allo)phones to phonemes.

5 Case Study: Nasal Assimilation

The pronunciation of *income* (discussed earlier) illustrates an important type of phonological rule found in many languages. There is a tendency for neighbouring phonemes to influence one another in such a way that their phonetic representations become more similar; this presumably makes the word easier to pronounce.[11] A rule which covers *income* and *uncle* in SEE is:

/n/ when followed by [k] becomes [ŋ]

In other words, /n/ in the context '_ [k]' becomes [ŋ]. Or in notation I've adopted above:

/n/ → [ŋ] : _ [k]

Replacing the IPA symbols by their feature sets (ignoring aspiration) gives:

{nasal, alveolar, voiced} → {nasal, velar, voiced} : _ {stop, velar, voiceless}

While it seems clear that it is useful to regard *income* as /ɪnkʌm/ rather than /ɪŋkʌm/, in order to maintain the *in* component, it is less clear whether words like *uncle* or *ankle* should be treated as containing /n/ or /ŋ/. However if /n/ is chosen, the rule will generate the correct phones, replacing /nk/ by [ŋk].

Whenever we have specified a rule, it is useful to ask whether it can be generalized. If we omit voicing from the rule above, we have:

{nasal, alveolar} → {nasal, velar} : _ {stop, velar}

---

10 A '/' is often used in place of my ':', but can be confused with the '/' used to mark a phoneme.

11 Another general term for this phenomenon is 'co-articulation'. I've restricted my examples to cases where the effect is so strong that an allophone of a different phoneme is involved.

(In words, an alveolar nasal followed by a velar stop becomes a velar nasal, regardless of voicing.)

Is the revised rule correct? We need to test it with a voiced velar stop (i.e. /g/) rather a voiceless velar stop (i.e. /k/). The word *anger* could be represented phonemically as /ængə/, but is pronounced in SEE as [æŋgə], which suggests that the revised rule may indeed be correct.

However, the rule as formulated so far does not show what seems to be the underlying reason for the change. This is that the nasal /n/ is being assimilated to the following stop so that it has the same position of articulation. An even more general rule is:

> A nasal in the context '_ stop' becomes articulated at the same position as the stop.

Or in the more formal notation adopted here:

> {nasal} → {nasal, Position} : _ {stop, Position}

Note that I am using the Prolog convention that constants start with lower-case letters, variables with upper-case letters.

This rule suggests that *anple* should not be an English word, since if we input the appropriate phonemes, the output is *ample*:

np
= /n/ /p/
= {nasal, alveolar} {stop, bilabial, voiceless}
→ {nasal, bilabial} {stop, bilabial, voiceless} (by applying the rule above)
= [m] [p]
= *mp*

Starting from the sequence {nasal} {stop}, the only possible outputs are [mp], [mb], [nt], [nd], [ŋk] and [ŋg]. Thus the rule correctly predicts the existence of words like *ample*, *amble*, *antler* and *handle* (all pronounced as spelt), and *ankle* and *angle* (where /n/ becomes [ŋ]).

Other assimilation rules are covered in the Exercises. Deciding on the appropriate rules for a given language is (hopefully) a task for a linguist rather than a computer scientist, whose role is to implement the rules in a computer program. However, it is important to understand the nature of phonological rules, and studying some simple examples seems a good way of acquiring and demonstrating this understanding.

Exercises

Don't feel you need to tackle all of these! I suggest as a minimum you try (1) – (5) and a selection of the others. The object is to reinforce your understanding of the concepts: phoneme, allophone and phonological rule.

1. How many phonemes are there in (a) *Keith* (b) *coughs*? What are they in the IPA? In each case try to demonstrate the correctness of your answer by finding words differing by only one of the phonemes you have identified.

2. You are NOT expected to know the IPA symbols; the table given in the Appendix will be provided if and when necessary. However it is useful to have some practice in using them. Study the following phonetic transcription of a verse of Lewis Carroll's poem *The Walrus and the Carpenter*. The transcription corresponds to my 'careful' pronunciation. Write down the normal English spelling. If your pronunciation differs from mine, write down an amended transcription in the IPA.

   ðə tʰaɪm hæz kʰʌm ðə wɔlrʌs sɛd
   tʰu tʰɔk ɒv mɛni θɪŋz
   ɒv ʃuz ænd ʃɪps ænd silɪŋ wæks
   ɒv kʰæbɪdʒɪz ænd kʰɪŋz
   ænd waɪ ðə si ɪz bɔɪlɪŋ hɒt
   ænd wɛðə pʰɪgz hæv wɪŋz

3. Consider the prefix *in* with the meaning 'not' followed by an English word beginning with one of the 6 English stops. (For example, *in* + *defensible* = *indefensible* meaning 'not defensible'.) Make sure that the word after *in* can occur as a separate word and that the meaning is 'not' rather than motion as in *input*. Do these words show the 'nasal assimilation' rule developed above?

4. *Aston* is usually pronounced [æstən]; *Asda* [æzdə]. Assuming that both words contain the phoneme /s/ (i.e. that the phonemic representations are something like /æstɒn/ and /æsdə/), suggest an appropriate phonological rule to generate the correct pronunciation. Try generalizing your initial rule. Can you find other examples to fit your rule? (In this and following exercises, try to write the rules in both English and feature set notation.)

5. Assume that the standard way of forming the plural of an English noun is to add the phoneme /z/ (NOT /s/) to the end of the word. Consider words ending in one of the 6 English stops. Construct phonological rule(s) to generate the correct pronunciation. For example, the plural of *bid* is [bɪdz] as expected but the plural of *bit* is [bɪts].

6. In German, words whose spelling ends with *d* are pronounced with [t], as are words whose spelling ends with *t*. Many such words have inflected forms in which *e* is added with the (approximate) pronunciation [ə]. These are pronounced as spelt. Thus:

   das Bund ist bunt = [das bʊnt ɪst bʊnt] (the bundle is colourful)
   bunte Bunde = [bʊntə bʊndə] (colourful bundles)

   Write appropriate phonological rule(s).

7. Another kind of phonological rule actually DELETES phonemes from the output. Consider English words whose spelling suggests that they end in a nasal phoneme followed by a stop (e.g. *lend*). The nasal assimilation rule suggests that the nasal and the stop will have the same position of articulation so that only 6 endings are possible. Is this correct? Write rule(s) to generate the correct pronunciation of all the endings you find.

8. The table below shows the pronunciation of some verbs and their negatives in one dialect of Modern Greek (spoken rapidly).

[anɛvazo]	I lift up	[ðɛn anɛvazo]	I don't lift up
[ɛtrɛksa]	I ran	[ðɛn ɛtrɛksa]	I didn't run
[viazomɛ]	I'm in a hurry	[ðɛ viazomɛ]	I'm not in a hurry
[lɛs]	you say	[ðɛ lɛs]	You don't say
[bɛno]	I enter	[ðɛ bɛno]	I don't enter
[pirazi]	it matters	[ðɛm birazi]	it doesn't matter
[trɛksanɛ]	they ran	[ðɛn drɛksanɛ]	they didn't run
[katalava]	I understood	[ðɛŋ gatalava]	I didn't understand

Assume that the negative of a verb is formed by preceding it by the phoneme string /ðɛn/. Suggest rule(s) to predict the pronunciation of the negative form of the verb.

9. There is a general tendency in languages to make successive phones 'more similar'. For example, we have seen that sequences such as /nk/ may be mapped to [ŋk], where the position of articulation of both phones is velar. A different kind of rule INSERTS phones into the output to improve the similarity between successive phones. For example, in the sequence /mk/ we move from a voiced bilabial (nasal) to an voiceless velar (stop). English names with this sequence, such as *Tomkin*, are often pronounced as though spelt *Tompkin*. The sequence is then voiced bilabial (nasal) to voiceless bilabial (stop) to voiceless velar (stop), now with only one voicing or position change with each new phone. Another example is *Hamton* which may yield *Hampton*.

   a) Write the appropriate rule(s) for the sequence /m<stop>/.

   b) The alternation of *Tomson* and *Tompson* suggests that this rule can be generalized to /m<fricative>/. Try it.

10. Consider the following information about Spanish in the dialect spoken around Madrid.

<u>Spelling</u>	<u>Pronunciation</u>	<u>Meaning</u>
dato	[dato]	fact, piece of information, datum
dardo	[darðo]	dart
dolar	[dolar]	dollar
drama	[drama]	drama
duda	[duða]	doubt
madre	[maðre]	mother
claridad	[klariðað]	brightness, clarity

([a], [o] and [e] are PURE vowels which occur in Spanish and in SAE but not in SEE which only uses them in the diphthongs [aʊ], [oʊ] and [eɪ].)

   a) Assuming that /d/ is a phoneme in Spanish, and that [d] is the same phone as in English (which the Collins Spanish Pocket Dictionary says that it is), write rule(s) to generate the correct pronunciations of the words given above.

   b) In fact, in the relevant dialect of Spanish, the initial phone in words like *dato* is a DENTAL stop ([d̪] in IPA), not an alveolar stop as in English. In other words, the tip of the tongue starts on the back of the upper teeth rather than on the tooth ridge. A feature set definition for this phone is {stop, dental, voiced}.

   Revise your rules from (a) to generate pronunciations using this phone (e.g. *duda* should give[d̪uða]).

   c) Suppose the rule(s) from (b) are generalized to ALL voiced stops. Predict the pronunciation of the /g/ phonemes in *gafas* (=a pair of glasses) and *paga* (=payment) and the /b/ phonemes (spelt *v*) in *via* (=route) and *dividir* (=to divide). You may end up with phones which do not occur in English. (To check your answer, you need a Spanish speaker from the right area of Spain!)

11. Construct a similar example to those given here, ideally based on a different language or dialect of English. (If you know French, for example, consider the pronunciation of final consonants in phrases such as *petit enfant, petit livre* or *petit oiseau*. Speakers of different English dialects could attempt to produce phonological rules mapping SEE to their speech, assuming the phonemic representation is stored in SEE. Thus, for example, /bɑθ/ → [bæθ] in northern English English.) I would be interested to see any such examples.

Appendix

IPA	*SEE Examples*	*ASCII*	*Partial Feature Set*
[i]	h**ee**l, m**e**	IY	{vowel,voiced}
[ɪ]	h**i**t	IH	{vowel,voiced}
[e]	*SAE b**ai**t*	EY	{vowel,voiced}
[ɛ]	m**e**t, h**ea**d	EH	{vowel,voiced}
[æ]	h**a**t	AE	{vowel,voiced}
[a]	*SAE f**a**ther, p**o**t*	AA	{vowel,voiced}
[ə]	**a**bout, aft**er**, f**er**n	AX	{vowel,voiced}
[ʌ]	**u**p, f**u**n	UX	{vowel,voiced}
[u]	s**oo**n	UW	{vowel,voiced}
[ʊ]	p**u**t, f**oo**t	UH	{vowel,voiced}
[o]	*SAE b**oa**t*	OW	{vowel,voiced}
[ɔ]	f**or**k, t**au**t	AO	{vowel,voiced}
[ɒ]	h**o**t	OH	{vowel,voiced}
[ɑ]	b**a**th, b**ar**	AH	{vowel,voiced}
[eɪ]	w**ai**t, c**a**ke	EI	{vowel,voiced}
[aɪ]	k**ite**, b**uy**	AY	{vowel,voiced}
[ɔɪ]	c**oi**n, t**oy**	OY	{vowel,voiced}
[oʊ]	bone, open	OU	{vowel,voiced}
[aʊ]	c**ow**, **ou**t	AW	{vowel,voiced}
[ɪə]	**ear**, sh**eer**	IA	{vowel,voiced}
[ɛə]	**air**, sh**are**	EA	{vowel,voiced}
[ʊə]	t**our**	UA	{vowel,voiced}
[p]	s**p**in	p	{stop,bilabial,voiceless}
[b̥]	**b**oo	b	{stop,bilabial,voiced}
[t]	s**t**op	t	{stop,alveolar,voiceless}
[d̥]	**d**og	d	{stop,alveolar,voiced}
[k]	s**c**an	k	{stop,velar,voiceless}
[g̊]	**g**ate	g	{stop,velar,voiced}
[m]	**m**at	m	{nasal,bilabial,voiced}
[n]	**n**ot	n	{nasal,alveolar,voiced}
[ŋ]	ki**ng**	NG	{nasal,velar,voiced}
[f]	**f**at	f	{fricative,labiodental,voiceless}
[v]	**v**at	v	{fricative,labiodental,voiced}
[θ]	**th**umb	TH	{fricative,dental,voiceless}
[ð]	**th**at	DH	{fricative,dental,voiced}
[s]	**s**at	s	{fricative,alveolar,voiceless}
[z]	**z**ip	z	{fricative,alveolar,voiced}
[ʃ]	me**sh**	SH	{fricative,palatal,voiceless}
[ʒ]	mea**s**ure	ZH	{fricative,palatal,voiced}
[h]	hot	h	{fricative,glottal}
[tʃ]	**ch**air	CH	{affricative,palatal,voiceless}
[dʒ]	e**dg**e, **j**am	JH	{affricative,palatal,voiced}
[l]	**l**ot	l	{approximant,voiced}
[r]	**r**ot	r	{approximant,voiced}
[j]	**y**awn	y	{approximant,voiced}
[w]	**w**in	w	{approximant,voiced}

Notes

bilabial = both lips
labio-dental = upper teeth and lower lip
dental = tongue tip and upper teeth
alveolar = tongue tip and tooth ridge
palatal = tongue and hard palate
velar = tongue and soft palate