CHAPTER 1
What is phonology?

PREVIEW

This chapter introduces phonology, the study of the sound systems of language. Its key objective is to:

• explain the difference between physical sound and "a sound" as a discrete element of language

• highlight the tradeoff between accuracy and usefulness in representing sound

• introduce the notion of "sound as cognitive symbol"

• present the phonetic underpinnings of phonology

• introduce the notion of phonological rule

KEY TERMS
sound
symbol
transcription
grammar
continuous nature of speech

Phonology is one of the core fields composing the discipline of linguistics, which is the scientific study of language structure. One way to understand the subject matter of phonology is to contrast it with other fields within linguistics. A very brief explanation is that phonology is the study of sound structure in language, which is different from the study of sentence structure (syntax), word structure (morphology), or how languages change over time (historical linguistics). But this is insufficient. An important feature of the structure of a sentence is how it is pronounced – its sound structure. The pronunciation of a given word is also a fundamental aspect of the structure of the word. And certainly the principles of pronunciation in a language are subject to change over time. So phonology has a relationship to numerous domains of linguistics.

An important question is how phonology differs from the closely related discipline of phonetics. Making a principled separation between phonetics and phonology is difficult – just as it is difficult to make a principled separation between physics and chemistry, or sociology and anthropology. While phonetics and phonology both deal with language sound, they address different aspects of sound and serve different purposes. Phonetics deals with "actual" physical sounds as they are manifested in human speech, and concentrates on acoustic waveforms, formant values, duration measured in milliseconds, or amplitude and frequency. Phonetics also deals with the physical principles underlying the production of sounds, such as vocal tract resonances, or the muscles and other articulatory structures used to produce those resonances. Phonology, on the other hand, is an abstract cognitive system dealing with rules in a mental grammar: principles of subconscious “thought” as they relate to language sound. Phonology is responsible for selecting the proper sound categories for the individual sounds in the morphemes of an utterance, given a certain context. Phonetics is responsible for giving physical form to those abstract cognitive sounds.

1.1 Phonetics – the manifestation of language sound

From the phonetic perspective, “sound” refers to time-varying mechanical pressure waves and the sensations arising when a pressure wave strikes the ear. In physical sound, the wave changes continuously, which can be graphed as a waveform showing the amplitude on the vertical axis and time on the horizontal axis. Figure 1 displays the waveform of a pronunciation of the word “wall”, with an expanded view of the details of the waveform at the center of the vowel between “w” and “ll”.

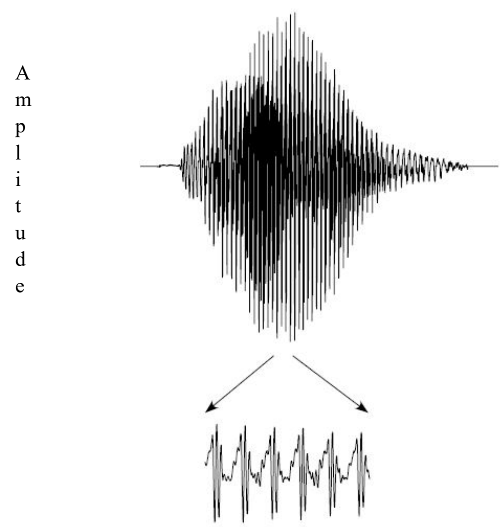


FIGURE 1

Figure 2 provides an analogous waveform of a pronunciation of the word “will”, which differs from “wall” just in the choice of the vowel.

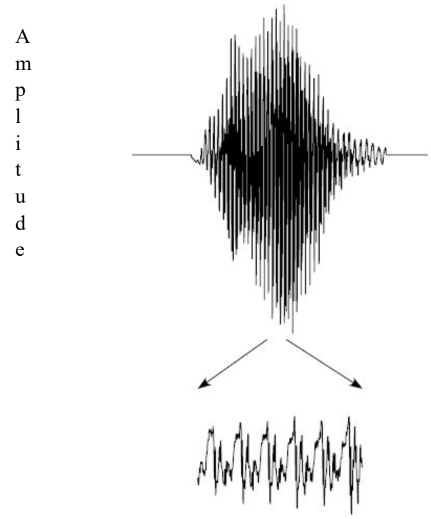


FIGURE 2

Inspection of the expanded view of the vowel part of these waveforms reveals subtle differences in the overall shape of the time-varying waveforms, which is what makes these words sound different.

It is difficult to directly characterize those physical differences from the waveform, but an important analytical tool of phonetics, the **spectrogram**, provides a useful way to describe the differences, by reducing the absolute amplitude properties of a wave at a sequence of ‘exact’ times, to a set of (less precise) amplitude characteristics in different frequency and time regions. In the spectrogram below, the vertical axis represents frequency in Hertz (Hz) and darkness represents amplitude. Comparing the spectrograms of “wall” and “will” in Figure 3, you can see that there are darker bands in the lower part of the spectrogram, and the frequency at which these bands occur – known as **formants** – is essential to perceptually distinguishing the vowels of these two words. Formants are numbered from the bottom up, so the first formant is at the very bottom.

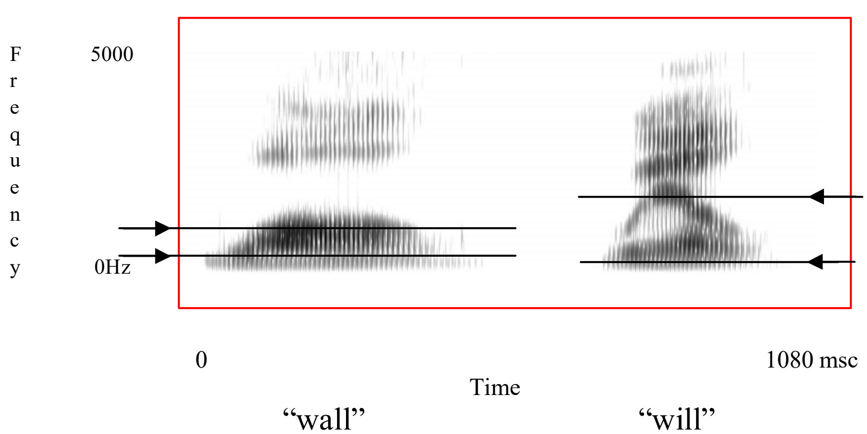


FIGURE 3

In “wall” the first two formants are very close together, occurring at 634 Hz and 895 Hz, whereas in “will” they are far apart, occurring at 464 Hz and 1766 Hz. In fact, in “wall”, the formants are so closed together that that they effectively merge into one blur in the picture, and we have to rely on another analytic method of LPC analysis to distinguish the formants. In “will”, on the other hand, the spectrogram clearly shows the second formant rising then falling. The mechanical reason for the difference in these sound qualities is that the tongue is in a different position during the articulation of these two vowels. In the case of the vowel of “wall”, the tongue is relatively low and retracted, and in the case of “will”, the tongue is relatively fronted and raised. These differences in the shape of the vocal tract result in different physical sounds coming out of the mouth.

The physical sound of a word’s pronunciation can be highly variable, as we see when we compare the spectrograms of three pronunciations of “wall” in Figure 4: the three spectrograms are obviously different, even though there is an overall similarity.

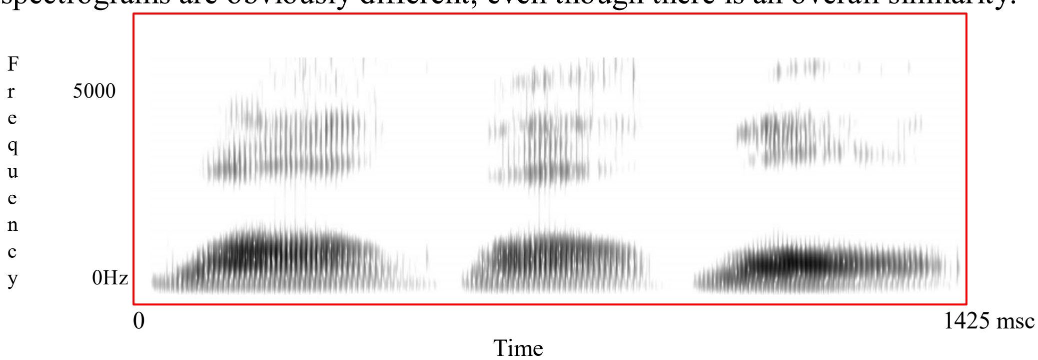


FIGURE 4

The first two pronunciations are produced at different times by the same speaker, differing slightly in where the first two formants occur (634 Hz and 895 Hz for the first token versus 647 Hz and 873 Hz for the second), and they differ in numerous other ways such as the greater amplitude of the lower formants in the first token. In the third token, produced by a second (male) speaker of the same dialect, the first two formants are noticeably lower and closer together, occurring at 541 Hz and 617 Hz.

Physical variation in sound also arises because of differences in surrounding context. Figure 5 gives spectrograms of the words “wall”, “tall” and “lawn”, with grid lines to identify the portion of each spectrogram in the middle which corresponds to the vowel.

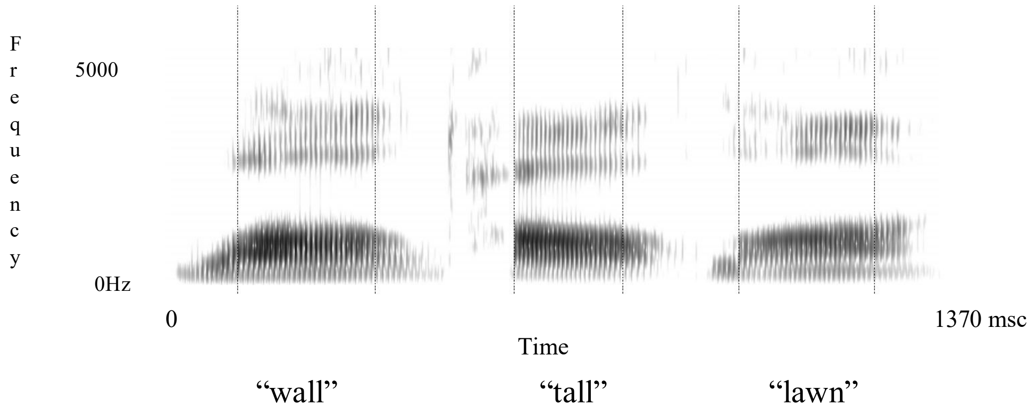


FIGURE 5

In “wall”, the frequency of the first two formants rapidly rises at the beginning and falls at the end; in “tall”, the formant frequencies start higher and falls slowly; in “lawn”, the formants rise slowly and do not fall at the end. A further important fact about physical sound is that it is continuous, so while “wall”, “tall” and “lawn” are composed of three sounds where the middle sound (*qua* cognitive category) in each word is the same one, there are no sharp physical boundaries between the vowel and the surrounding consonants.

A common type of continuous phonetic process is **coarticulation**, where an aspect of the production of one sound overlaps the production of surrounding sounds. We see this in two English utterances, ‘I scream’ and ‘I scheme’, which are the same except for the inclusion of “r” in ‘scream’. In this dialect of English, the consonant *r* is strongly rounded. However, that lip rounding extends significantly before and after *r* itself. In the spectrogram of these utterances, we can see the influence of this lip rounding on neighboring segments, by how *r* causes a lowering of resonance frequencies.

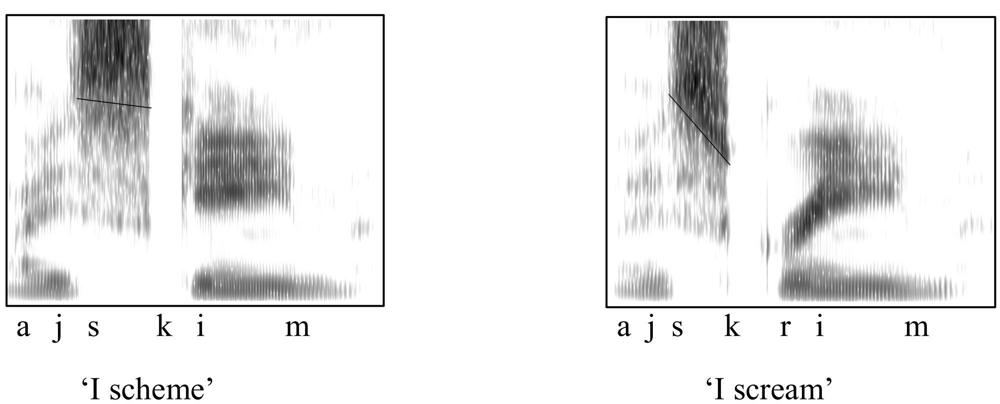


FIGURE 6

The mostly-blank part roughly in the middle is the consonant [k]. In ‘scheme’, what follows is the vowel [i] and the first two dark horizontal bands are F_1 and F_2, at 334 Hz and 2260 Hz. In ‘scream’, what follows is the consonant [ɹʷ], whose F_1 and F_2 start at 361 and 1222 Hz – lip rounding causes significant lowering of the second formant. F_2 then rises continuously to 2057 Hz in the following vowel [i]. This happens because [ɹʷ] is rounded, therefore it has low F_2, and while [i] is not rounded, the change from round to non-round happens gradually, overlapping the production of [i] significantly. We also see an influence of [ɹʷ] on the preceding [s]. As indicated by the line in 'scheme', the darker higher amplitude portion of [s] is relatively constant in frequency, reflecting the intrinsic resonance properties of [s]. In contrast, in 'scream', the resonance frequency of [s] falls substantially from the beginning to ending of [s], because of the increasing anticipatory lip-rounding of [s] before [ɹʷ] (we cannot detect a lowering effect on [k] since it is essentially silence until its release). On the web page you can hear the full utterances, plus excised [s], where the falling pitch of [s] in 'scream' is very evident.

1.2 Phonology: the mental representation of sound

The goal of science is the discovery of principles that explain the nature of things, so we need to be able to organise things in the world into different types. In linguistics, we cannot directly inspect pressure waves produced by speakers of a language, we need to create a stable representation of that physical sound from which we can learn about the speech-related properties of language. A very common current way to do this is to make a digital recording, to use some simple hardware (for example a microphone and a computer) to reduce physical sound to a series of numbers. This way, we have a potentially permanent capturing of that sound, and we can compare and classify utterances for similarities and differences. Aided by the companion web page https://languagedescriptions.github.io/IP3/ch1web.html, we will look at different ways of representing language sounds, looking at different ways to represent two words of Logoori (a Bantu language of Kenya), which translate to English 'dog' and 'new'. The ultimate goal is to say how these utterances are similar, and how they are different, which will then inform us somewhat about this language.

The web page provides the full set of numbers that are the digital representation of these recordings, below is a fractional sample of the first, middle and last 3 numbers.

'dog'	'new'
-34	-4
-36	-5
-29	-9
⋮	⋮
-2709	514
-2497	1111
-2079	2040
⋮	⋮
-472	681
-424	576
-349	443

TABLE 1

You could listen to these recordings, but I recommend first trying to understand what these words might be based on the other evidence provided. The wall of numbers (9,183 for ‘dog’, 9,191 for ‘new’) supplemented with the information that the data has a sampling rate of 22050 Hz, is almost completely uninformative, though you can determine that it lasts 416.462 msc. Clearly, we need better visualization.

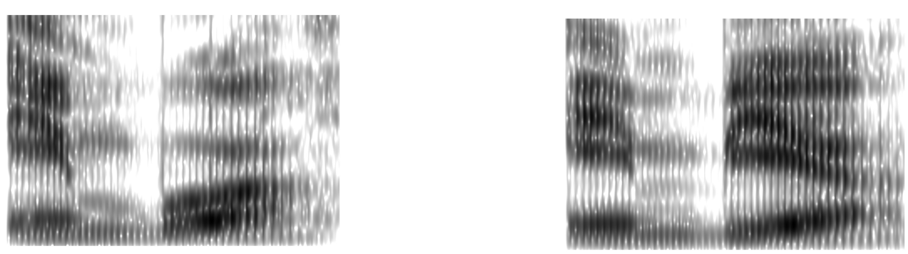
We can also graph these numbers, creating a waveform of the utterances.

‘dog’ ‘new’

FIGURE 7

An expert phonetician could tell you a very little bit about these utterances from these pictures. While a waveform picture is more informative than a long table of numbers, it is rather imprecise. Time and amplitude information is very hard to extract from the picture, though it is easier to perceive the whole representational object compared to the table of numbers. Given a reasonable amount of training in reading waveform pictures, it might be possible to surmise that the utterance is composed of a vowel-nasal-vowel sequence, but which vowels and nasals?

A more informative display is a spectrogram, which transforms the table of numbers into another bunch of numbers, in this case a 116 × 204 matrix of (floating-point) numbers (23,664 numbers).



‘dog’ ‘new’

FIGURE 8

An expert phonetician could tell you more about these words from these pictures. But how do we talk about the differences systematically, not using vague descriptions like “a bit darker”, “a bit further up the picture” or “a bit further to the right”? Again, these pictures are more informative than waveform pictures or tables of nunbers, but it is difficult to say exactly when something happens (the horizontal axis) or at what frequency it happens (the vertical axis), or at what amplitude (relative darkness).

Our next step is to compute formant values, again a list of numbers, but this type of analysis is much more informative and compact, thus useful. We can reduce each of these words to 228 numbers, which is 76 triples of formant numbers: F_1, F_2, F_3. As you can see on the web page, these utterances have been stripped to a more manageable size, plus we can be much more definite about the time and frequency properties that we are talking about. A portion of the table of formants is given below.

0.02	459	2198	2863	491	2151	2849
0.04	451	2174	2885	475	2155	2851
0.06	464	2156	2897	481	2139	2845
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.48	620	1556	3132	691	1415	2290
1.5	702	1530	2991	666	1437	2246
1.52	636	1598	3025	688	1429	2274
		‘dog’			‘new’	

TABLE 2

Unfortunately, in the course of throwing away information in order to compact the representation of these utterances and make the information more usable, we have significantly compromised the fidelity of the representation – this set of formant values can be re-converted into a sound waveform, but one that poorly represents the original utterances (play the re-synthesized sounds from the web page).

The problem we faced in seeking a useful representation of the language sounds of these words in terms of physical properties is that we were using the wrong tool for classification, namely physical analysis. These numeric tools can tell us how the sound categories of a language are realized, but they do not tell us how the mind organizes external sound into cognitive categories. We need symbolic representations of physical sound as they are used as cognitive building blocks of language. Therefore, we simply reduce these two recordings to a tiny set of technically-defined symbols specialially developed for language, and represent these utterances as [ímbwá] ‘dog’ and [ímbjá] ‘new’. How we converting the continuous stream of speech sound into discrete units, a transcription, is the topic of Chapter 2.

1.3 The concerns of phonology

You should now understand what distinguishes phonetics from phonology: phonology is the study of cognitive computations performed on categories based on sound, phonetics is the study of how those categories are physically realized. Now we consider some specific aspects of sound structure that would be part of a phonological analysis.

The sounds of a language. One aspect of phonology investigates what the individual sounds of a language are. We could take note in a description of the phonology of English that we have the sound [θ] as in ‘**th**ing’, which is lacking in German and French, and while English lacks the vowel [ø], that vowel exists in German in words like *schön* ‘beautiful,’ also in French (spelled *eu,* as in *jeune* ‘young’), or Norwegian (*øl* ‘beer’).

Sounds in languages are not just isolated atoms; they are part of a system. The systems of stops in Hindi and English are given in (1). The symbol ^{h} indicates that the consonant is aspirated, and ʈ, ɖ in the third column of Hindi indicates a retroflex stop, a sound type lacking in English – discussed in chapter 2.

(1)

Hindi stops				English stops		
p	t	ʈ	k	p	t	k
p^h	t^h	$ʈ^h$	k^h	p^h	t^h	k^h
b	d	ɖ	g	b	d	g
b^h	d^h	$ɖ^h$	g^h			

The stop systems of these languages differ in three ways. English does not have a series of voiced aspirated stops like Hindi [b^h d^h $ɖ^h$ g^h], nor does it have a series of retroflex stops [ʈ $ʈ^h$ ɖ $ɖ^h$]. Furthermore, the phonological status of the aspirated sounds [p^h t^h k^h] is different in the languages, as discussed in chapter 3, in that they are basic lexical facts of words in Hindi, but are the result of applying a rule in English.

Rules for combining sounds. Another aspect of language sound which a phonological analysis may take account of is that in any language, certain combinations of sounds are allowed, but other combinations are systematically impossible. The fact that English has the words [gɹik] 'Greek', [gɹejt] 'grate', [gɹʌd^ʒ] 'grudge', [bɹɛd] 'bread' is a clear indication that there is no rule against words beginning with the consonant sequence [gɹ]. Similarly, there are many words which begin with *gl*, such as [glu] 'glue', [glɪf] 'glyph', [glæns] 'glance', [glɪmɹ̩] 'glimmer' showing that there is no rule against words beginning with *gl*. It is also a fact that there is no word *[glɪk][1] in English. The question is, why is there no word **glick* in English? The best explanation for this is simply that it is an accidental gap. Not every logically possible combination of sounds following the rules of English phonology is found as an actual word of the language.

While the nonexistence of *glick* in English is accidental, the exclusion from English of certain other imaginable but nonexistent words is based on a rule of the language. There are words that begin with *sn* like *snake*, *snip* and *snort*, also numerous words beginning with [sm], for example *small, smite, smidgen, smell*, but no words words beginning with *gn* or *gm,* thus **gnick*, **gnark*, **gniddle, *gmelt, *gmite* are not words of English. While there are words spelled with *gn*, such as *gneiss, gnostic*, they are pronounced without [g] – [naɪs], [nɔstɪk]. Moreover, native speakers of English have a clear intuition that hypothetical **gnick*, **gnark*, **gniddle, *gmelt, *gmite* could not be words of English, whereas speakers have no such intuition about accidentally non-existent *glick*. A description of the phonology of English would provide a basis for characterizing the fact that English words can start with [gl, gɹ] but not [gn, gm].

Rule-governed variations in pronunciation. A phonological analysis especially explains variations in the pronunciation of words-parts (morphemes). For example, there is a very general rule of English phonology which dictates that the plural suffix on nouns is pronounced as [ɨz], represented in spelling as *es*, when the preceding consonant is one of a certain set of consonants including [ʃ] (spelled *sh*) as in *bushes*, [t^ʃ] (spelled as *ch*) as

[1] The asterisk is used to indicate that a given word is non-existent or wrong.

in *churches*, and [dʒ] (spelled *j*, *ge*, *dge*) as in *cages*, *bridges*. This pattern of pronunciation is not limited to the plural, so despite the difference in spelling, the possessive suffix *s*[2] is also subject to the same rules of pronunciation: thus, plural *bushes* is pronounced the same as the possessive *bush's*, and plural *churches* is pronounced the same as possessive *church's*.

This is the sense in which phonology is about the sounds (countable) of language. From the phonological perspective, a "sound" is a specific unit in a language which combines with other specific units according to certain rules, and which are realized as physical sound. What phonology is concerned with is how sounds behave in a grammar.

Summary

Phonetics and phonology both study language sound. Phonology examines language sounds as mental units, encapsulated symbolically for example as [æ] or [g], and focuses on how these units function in grammars. Phonetics examines how symbolic sound is manifested as a continuous physical phenomenon. The conversion from the continuous external domain to mental representation requires focusing on the information that is important, which is possible because not all physical properties of speech sounds are cognitively important. One of the goals of phonology is then to discover exactly what these cognitively important properties are, and how they function in expressing regularities about languages.

Exercises

The first three exercises are intended to be a framework for discussion of the points made in this chapter, rather than being a test of knowledge and technical skills.

1. Examine the following true statements and decide if each best falls into the realm of phonetics or phonology, and why.
   a. Sound in the word *frame* changes continuously.
   b. The word *frame* is composed of four segments.
   c. Towards the end of the word *frame*, the velum is lowered.
   d. The last consonant in the word *frame* is a bilabial nasal.
2. Explain what a "symbol" is; how is a symbol different from a letter?
3. Why would it be undesirable to use the most precise representation of the physical properties of a spoken word that can be created under current technology in discussing rules of phonology?

---

[2] This is the "apostrophe s" suffix found in The child's shoe, meaning 'the shoe owned by the child.'

Further reading

Ashby & Maidment 2005; Isac & Reiss 2008, Johnson 1997; Ladefoged & Johnson 2010 Liberman 1983; Stevens 1998.