Phonetics and Phonology

An Introduction to the Science of Speech

Chen Gafni

Document version: 1.0 (12 January 2025)

Important!

This document will be continually updated and improved. Download the latest version at: https://github.com/openphonbook. To suggest corrections and contribute content, please email me: chen.gafni@gmail.com

Table of Contents

List of Figures

List of Tables

Preface

This book aims to provide a comprehensive introduction to phonetics and phonology with emphasis on the relation between the fields. In addition to the more “traditional” topics, such as articulation and phonological analysis, the book discusses topics such as emotive speech and sound symbolism, which rarely appear in general phonetics and phonology textbooks. Another advantage of this book is the use of many media files, which make the contents of the book more vivid.

Despite the close relation between phonetics and phonology, they are distinct subjects in the eyes of the students, since they require different sets of skills. Phonetics is more about memorizing schematic processes, while phonology requires more analytical skills. The book includes many exercises designed to enhance all those necessary skills.

The idea of the book came about in 2020 during the first months of the covid-19 pandemic, which forced all academic teaching to go online. At the time, I was a new lecturer teaching “Phonetics and Phonology”, a first-year BA course, at the Department of English Literature and Linguistics at Bar-Ilan University. Facing overwhelmed and frustrated students, I started preparing summaries to help them cope with the unexpected challenges of online learning. These summaries grew up to become the present book.

After several years of writing and editing, I realized that writing such a comprehensive book is a bit too much for one person, especially since I was facing some personal challenges at the time. Therefore, I decided to turn to the scientific community and ask for help in exchange for making the book free to the public. The book will be stored in an online repository (see cover page), and I welcome contributions from additional authors.

Chen Gafni
January 2025

Acknowledgments

I would like to express my appreciation and gratitude to all the people who made me fall in love with the sounds of language. First and foremost, I would like to thank my teachers at Tel-Aviv University – Dr. Evan Cohen (phonetics), Dr. Galit Adam (phonological development), and especially Prof. Outi Bat-El (phonology), who was also my MA advisor. I owe special thanks to the late Prof. Reuven Tsur, with whom I took a fascinating nine-year journey in the realms of cognitive poetics. Also, I would like to thank Prof. Lior Laks, who gave me the opportunity to teach at Bar-Ilan University (first as a teaching assistant and later as a lecturer). I also thank my teaching assistant at Bar-Ilan University, Tamar Fenster-Roth, for commenting on the manuscript and for helping me test my pedagogical ideas in class. Finally, I thank all my students, whose questions and errors helped me sharpen my teaching as well as my own understanding of the material.

Version history

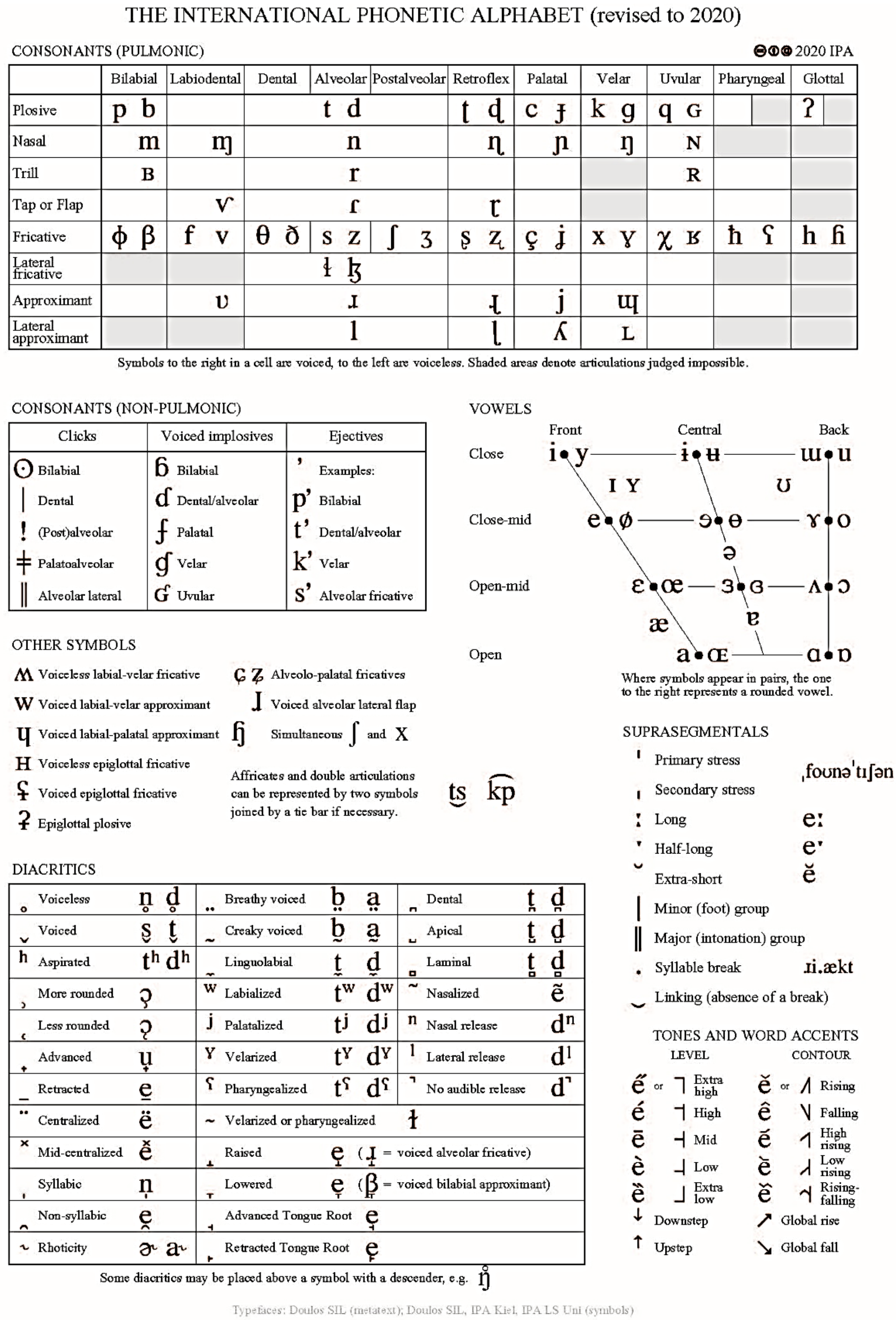
- 1.0 (12 January 2025): first published version of the book – 19 chapters in five parts.

Contributors

Chen Gafni, PhD – the author and editor of the document.

A former lecturer at Bar-Ilan University, The Hebrew University of Jerusalem, and Ben-Gurion University of the Negev. Developer of the software Child Phonology Analyzer. Main research interests: phonological development, acoustic phonetics, emotive speech, sound symbolism.

The International Phonetic Alphabet (IPA)



Part I: Introduction

Chapter 1 Overview

This book concerns two distinct, but closely related fields: phonetics and phonology. The purpose of this introductory chapter is to give the reader a general sense of what these fields are about. Section 1.1 provides general definitions of the terms *phonetics* and *phonology* and discusses the relations between the fields. Section 1.2 discusses the importance of studying phonetics and phonology. Section 1.3 reviews various topics that are of interest to phoneticians and phonologists.

1.1. Phonetics and phonology

As the names suggest, ***phon***etics and ***phon***ology have a shared interest: speech sounds (*phone* from Ancient Greek φωνή, means 'sound, voice') . The fields differ in the aspects of speech sounds that are in the focus of research (although the distinction between the fields is not always clear). **Phonetics**, the scientific study of speech, concerns the physical aspects of speech and communication. It is divided into three main branches: **Articulatory phonetics** studies the speech organs and mechanisms of speech production. **Acoustic phonetics** studies the physical properties of speech sounds. **Perceptual** (or, **auditory**) **phonetics** studies the processing and interpretation of speech sounds by the ears and the brain. Figure 1.1 illustrates the relations among the main branches of phonetics.

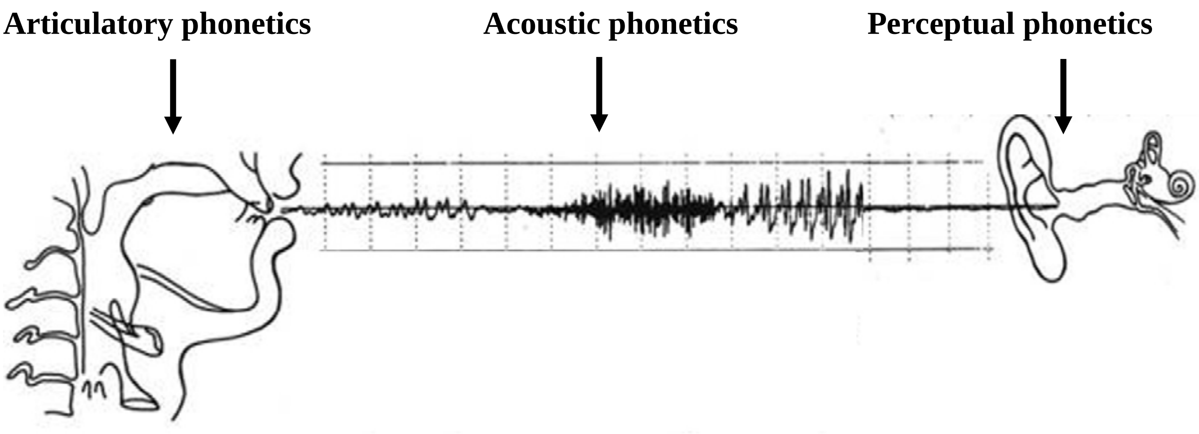


Figure 1.1 The branches of phonetics

Phonology is a branch of linguistics dealing with the relations among speech sounds in particular languages and in languages generally. Certain approaches to phonology are concerned with the cognitive aspects of speech sounds, namely the mental representation of speech sounds (in the "mind" of the speaker) and the rules governing the interactions among them. In other words, phonology is a theory of **grammar** in the descriptivistic sense – it is the set of linguistic rules concerning speech sounds that speakers acquire spontaneously and use intuitively (and subconsciously).

Superficially, phonetics and phonology are differentiated by research methods. Studying the articulation of speech sounds is relevant mainly to phoneticians, while analyzing sound patterns in written records is a typical practice in phonology. However, the two fields are not really separable. For example, phoneticians are interested in the innate physiological components underlying the production and perception of speech, which are shared by all humans. However, articulation and perception are also affected by linguistic experience. In other words, the phonology of a given language affects, to some extent, the way speakers of that language produce and perceive speech. In the opposite direction, phonological processes and historical sound changes are typically studied by theoretical phonologists. Yet, these changes often have an articulatory or perceptual basis. Although it is more convenient to discuss phonetics and phonology separately, I emphasize the link between them, whenever possible.

1.2. Motivation

In the broader context of linguistics and language studies, phonetics and phonology often take a secondary place to other linguistic branches, mainly syntax and semantics. It is likely because those aspects that focus on the structure (syntax) and meaning (semantics) of the message seem more important to humans than aspects focusing on sound. Yet, one should remember that speech sounds are the main carriers of the message in spoken language. Thus, understanding the mechanisms of sound production and perception is vital to understanding how language works.

Since the main goal of communication is transmitting thoughts and intentions, it is easily understood why the meaning components of the message are more accessible to humans than the properties of the sounds carrying the message. Consequently, the dimension of sound becomes prominent mainly when it interferes with communicating the message. First, early child speech is often unintelligible as infants are learning to use sound perception and production mechanisms. Second, accented speech often constitutes a barrier in communication between speakers of different dialects and languages. Third, various forms of hearing and language impairments (e.g., apraxia of speech) affect the person's ability to produce or perceive certain aspects of speech sounds. Finally, spelling errors (e.g., confusing *bed* and *bad*) may also represent difficulties related to speech sounds, especially when made by non-native speakers.

1.3. Research topics

This section reviews a variety of research topics in phonetics and phonology, including advanced topics that are not typically covered in general university courses. The purpose of this review is to expose the readers to current research trends and, hopefully, spark their interest in phonetics and phonology. The review of each topic begins with a general description of the topic, followed by a short summary of one specific study. I selected studies that I found interesting, but there are many other studies that would be just as good exemplars.

1.3.1. **Articulatory phonetics**

This field uses a variety of instruments, such as ultrasound and MRI, to study the speech organs and their movements during speech. This is useful for understanding processes of both normal and impaired speech production. One of the major problems in imaging studies of speech production is automatic detection of the tongue contour. Since the tongue is involved in the production of most speech sounds, determining its position and shape is critical for articulation studies. However, due to certain technical limitations, the tongue is not always clearly visible in the acquired images. A study by Xu et al. (2016) attempts to solve this problem with a sophisticated computer algorithm that can follow the motion of tongue contours over long ultrasound image sequences. Figure 1.2 illustrates automatic contour tracking of the tongue of a single speaker during a long stream of speech.

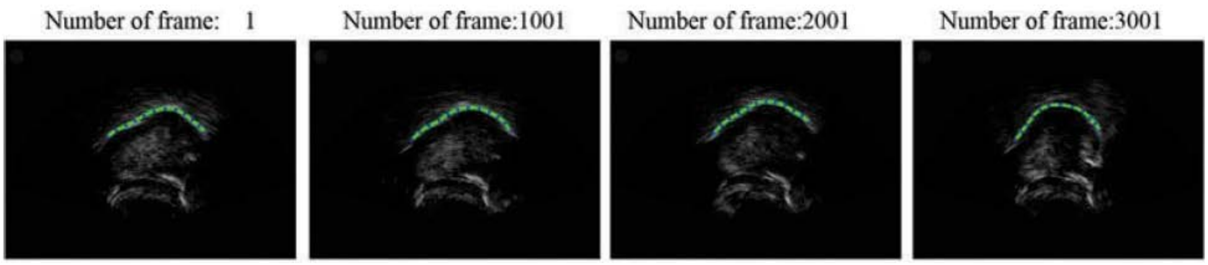


Figure 1.2 Automatic tongue contour tracking in ultrasound image sequence. The dashed lines are the automated contours (From: Xu et al., 2016)

1.3.2. Speech perception

The perception of speech is influenced by the properties of our auditory system (ears and brain), but also by our linguistic knowledge (stored in other parts of the brain). Researchers use a variety of experimental techniques to study the various factors involved in speech perception. In one such experimental setting, participants listen to pairs or triplets of meaningless words (also called **nonwords**) and their ability to discriminate the stimuli is tested. Since the words are unknown to the participants they must rely on their implicit phonological knowledge to make the judgments.

Dupoux, Hirose, Kakehi, Pallier, & Mehler (1999) used this method to demonstrate the well-known phenomenon of **categorical speech perception**, namely that speech perception is affected by the contrasts available in the speaker's native language. In their study, native speakers of Japanese and native speakers of French listened to triplets of nonwords like *ebzo*, *ebuzo*, and *ebuuzo* (*uu* represents a long vowel). They were asked to decide whether the third nonword in each triplet was the same as the first or the second nonword.

Dupoux et al. found that speakers of Japanese (but not French) had trouble differentiating nonwords like *ebzo* and *ebuzo*. This finding was attributed to the syllable inventory of Japanese, which has severe limitations on word-mid consonant sequences like *bz*, while there are no such limitations in French. By contrast, speakers of French (but not Japanese) had trouble differentiating nonwords like *ebuzo* and *ebuuzo*. This finding was attributed to the fact that French does not contrast long and short vowels (but Japanese does).

1.3.3. Speech acoustics

Analyzing the acoustic properties of recorded speech can serve various theoretical and practical purposes, from identifying emerging phonological changes to characterizing intonation patterns and investigating impaired speech. For example, a study by Ferragne & Pellegrino (2010) analyzed the acoustic properties of vowels in 13 accents of British English. Figure 1.3 illustrates the difference in the vowel of *hide* in two recordings by the same speaker. Such variation, when found within a group of speakers, could be a sign of a larger on-going language change. With the help of acoustic analysis, linguists can arrive at a precise description of variation within and across speakers and dialects. This information is also useful for various applications that require voice recognition.

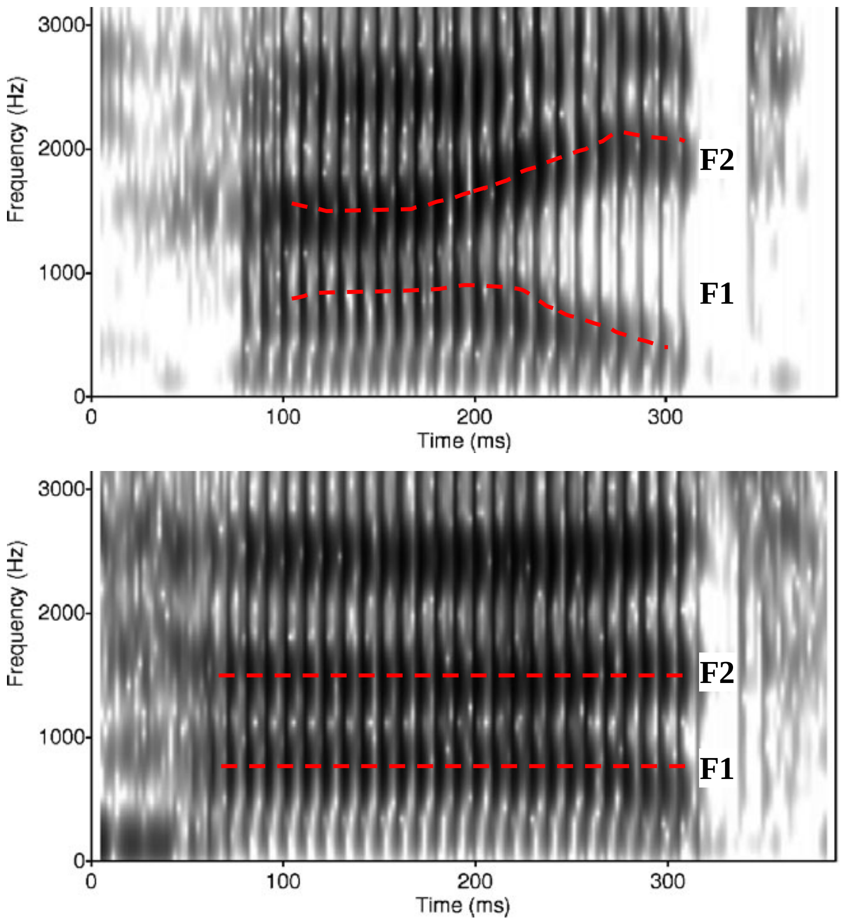


Figure 1.3 Acoustic analysis of the vowel of hide in two British accents. The dashed lines mark formant frequencies (see Chapter 11). The top panel illustrates a diphthong (a vowel of changing quality [ai]). The bottom panel illustrates a monophthong (a vowel with a constant quality) (adapted from: Ferragne & Pellegrino, 2010).

1.3.4. **Historical phonology and sound change**

Analyzing historical written documents and audio recordings can demonstrate how the phonology of language changes over long time periods. Such longitudinal analysis can reveal general linguistic principles and help us understand current language trends (and predict future linguistic changes). For example, Harrington, Palethorpe, & Watson (2000) analyzed nine Christmas broadcasts made by Queen Elizabeth II between 1952 and 1988. They found a shift in the quality of vowels over the course of time, and concluded that the queen's accent has shifted towards the "mainstream" accent of the British population. Figure 1.4 illustrates changes in the acoustic properties of some of Queen Elizabeth II's vowels. The ellipses mark the distribution of formant frequencies of each type of vowel (see Chapter 11). A longer/wider ellipse means greater variability. Overlapping ellipses suggest similar pronunciation of different vowels.

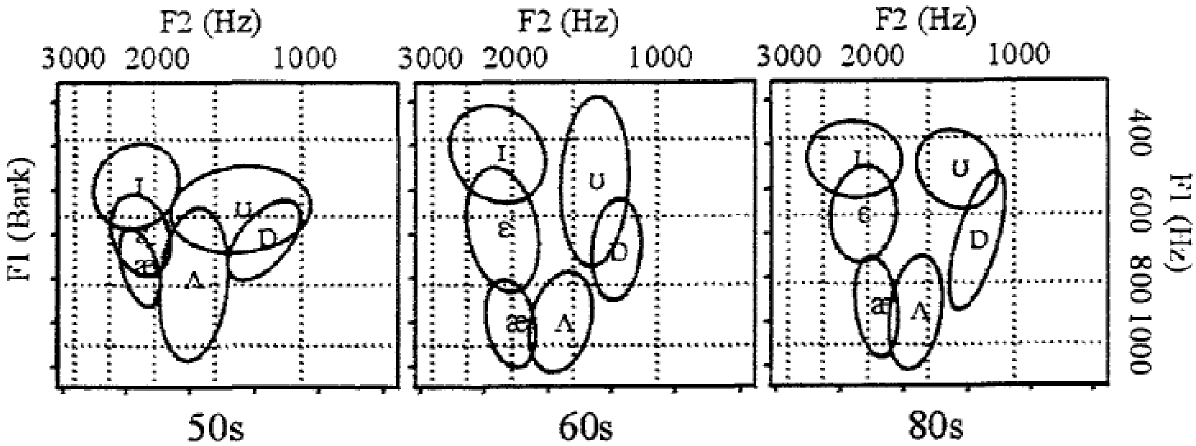


Figure 1.4 Changes in the acoustic distribution of some of Queen Elizabeth II's vowels from the 1950s to the 1980s. Ellipses include data points within two standard deviations of the mean. (From: Harrington et al., 2000)

1.3.5. **Typology and cross-linguistic comparisons**

Studies in this field focus on the similarities and variation across languages. Patterns and phenomena found in various (especially unrelated) languages could reflect fundamental cognitive and physiological principles shared by all humans. In particular, the extent of cross-linguistic variation can suggest what is linguistically possible or impossible. A study by Rose & Walker (2004) examined a phonological process found in many languages: **consonant harmony**. For example, in the language of Kikongo, the perfective active suffix /-idi/ is pronounced [-ini], when preceded by a nasal consonant like [m] and [n] (compare: [n-suk-i**d**i] 'I washed' and [tu-kun-i**n**i] 'we planted'). This can be described as a nasal assimilation of /d/ to another nasal consonant.

This pattern has been observed in other languages as well, while other, theoretically possible, patterns are not attested in any language (for example, in no language does /d/ change into [b] when the word contains another [b] at a distance, e.g., /sub-i**d**i/ → *[sub-i**b**i]; the asterisk indicates an inexistent form). Theoretical phonologists attempt to construct a theoretical model that would account for the existence of certain phenomena cross-linguistically and for the absence of other theoretically possible phenomena. For example, Rose

& Walker propose a theoretical model that incorporates the similarity and distance between consonants to predict when assimilation between consonants is more or less likely to occur.

1.3.6. Loanword phonology

The main practice in theoretical phonology is to analyze words from a given language and construct a theory that explains how sounds behave in that language. One of the best sources of data is **loanwords –** words that were borrowed from other languages. When such words are introduced to the target language, their pronunciation is often modified. The changes in pronunciation provide direct evidence to the active phonological principles of the target language. Discovering these principles is more difficult when considering only words that are native to the target language.

Paradis & Lacharité (1997) analyzed the phonological form of a large number of words borrowed from French to several African languages and also words borrowed from English to Quebec French. They found that the phonological adaptation of loanwords follow similar principles across languages: speech sounds (segments) not found in the target language are usually preserved with minimal changes. In addition, when a loanword contains a structure (such as syllable) that is ill-formed in the target language, a segment (mainly a vowel) is usually inserted to repair the ill-formed structure. Occasionally ill-formed segments are deleted when embedded within a higher level ill-formed structure. For example, the French word *civil* (pronounced [sivil]) is borrowed into the Western African language Fula as [siwil]. The consonant [v] does not exist in Fula and is replaced by [w]. The same consonant is deleted in [waju], the borrowed form of *voyou* ‘bum’ (pronounced: [vwaju]), to avoid a sequence of two consonants at the beginning of the word. Finally, the word [tabl] ‘table’ is borrowed into Fula as [taabal]. A vowel [a] is inserted between [b] and [l] to avoid a sequence of two consonants at the end of the word. Beyond the descriptive survey of loanwords, the goal of theoretical phonology, as in the case of Paradis & Lacharité, is to propose a theory that would account for the chosen repair strategies in different loanwords.

1.3.7. Sociophonetics

Sociophonetics studies the phonological expression of sociological variables (socioeconomic status, gender, etc.). In a classic study in sociophonetics (Labov, 1972), salespersons in three department store in New-York were asked for the location of some department (e.g., “where are the women’s shoes?”), when the expected answer was “fourth floor”. Importantly, the selected stores represented and appealed to customers of different social classes. This was indicated by factors such as the locations of the stores as well as their advertising and price policies. The study revealed that the language of the salespersons was in correlation with social class. Salespersons in the high-ranked store pronounced "r" in "fourth floor" more than salespersons in the low-ranked store ("fouth floo"). Since then, many studies have demonstrated the importance of social factors in phonological and phonetic studies.

1.3.8. Phonological development

Phonology (speech sounds, intonation, etc.) is the earliest aspect of language acquired by infants. For example, studies show that newborn babies can distinguish speech in their native language from speech in a foreign language based on prosodic cues such as rhythm and intonation (e.g., Mehler et al., 1988). Remarkably, it was found that children across the world acquire their native language along similar paths. This is important both for identifying typical and atypical language development and also for constructing and testing linguistic theories.

Early language acquisition studies were typically longitudinal diary studies, in which parents documented the development of language of their own child over a long period of time. Nowadays, most researchers conduct cross-sectional studies: one-time recording of many children at different ages. This is considered a more practical (but less direct) way to study language acquisition.

In one of the most famous longitudinal studies, Smith (1973) documented and analyzed the speech development of his son for two years. In addition to illustrating the acquisition of segments and syllable types, the study also investigated the relation between competence (in the sense of linguistic knowledge) and performance. For example, Smith demonstrated that his son was able to differentiate the words *mouse* and *mouth* (=competence) when he was still pronouncing them the same way (=performance) (the child performed successfully when asked to bring a flash card bearing a picture of a mouth or a picture of a mouse).

1.3.9. Phonetics of emotion

Spoken language is used not only for exchanging information about things in the world, but also to communicate the speaker's emotions, moods, and attitudes. This is done not only through the selection of words, but also via various aspects of speech such as intonation and voice quality. Studying the acoustic properties of emotive speech (i.e., speech that conveys emotions) can shed light on the way we express our emotions and how we understand the emotions of others.

Bänziger & Scherer (2005) used the popular technique of emotion portrayals to study the acoustics of emotive speech. In this experimental paradigm, professional actors are asked to read some text while displaying certain emotions. The acoustic properties of the spoken sentences are analyzed to obtain vocal profiles of various emotions. In addition, the recordings are presented to listeners, in order to investigate whether they are able to identify the intended emotions and find which vocal cues they use to make their judgments.

Bänziger & Scherer collected emotions portrayals for four emotions: *joy*, *anger*, *fear*, and *sadness*. The simulated emotions also varied in intensity. They found that emotional arousal was the best predictor of the results: high arousal emotions (despaired sadness, elated joy, panic fear, and hot anger) were produced with higher and more variable pitch than low arousal emotions (sadness, calm joy, and anxious fear). Such findings are helpful for the research of emotions and also for the development of various applications.

1.3.10. Sound symbolism

One of the most prominent features of languages is the **arbitrariness of the linguistic sign**. That is, the meaning of words does not depend by and large on their structure. For example, some equivalents for the word *dog* in several languages are: *chien* in French, *Hund* in German, *perro* in Spanish, and *kelev* in Hebrew. It is clear that these words are arbitrary signs: there is nothing about any of them that makes it particularly suitable to describe dogs. The arbitrariness of the linguistic sign has become one of the pillars of modern linguistics, at least since Ferdinand de Saussure's (1916) *Course in General Linguistics*.

However, there is a growing number of studies that show that the sounds of a word can contribute to its meaning in a non-arbitrary way. Such a non-arbitrary relation between sound and meaning is called **sound symbolism**, or **sound iconicity**. One of the most striking types of evidence for sound symbolism as an active cognitive principle is the **sound-shape symbolism**. In a popular demonstration of this phenomenon, known as the "bouba/kiki effect", people are shown two abstract shapes: one with a smooth, curvy contour, and one with a sharp, angular contour (see Figure 1.5). They are told that, in some unknown language, one of these two figures is called "bouba" and the other is "kiki", and they are asked to guess which is which. Several studies have shown that the vast majority pick "kiki" as the angular shape and "bouba" as the curvy shape, even though they have never seen these figures before. This effect is attributed to some kind of abstract similarity between the shapes and the sounds, and suggest that the "arbitrariness of the linguistic sign" hypothesis should be revised as 'the linguistic sign is usually arbitrary'.

Figure 1.5 Shapes from the "bouba/kiki" experiment (From: Ramachandran & Hubbard, 2001)

Chapter 2 Phonetic transcription

Researchers in the language sciences often study aspects of spoken language and wish to report their findings to a broad audience which may include speakers of different languages. For that purpose they need to transform spoken words in the language into a written form that represents the way these words are pronounced. This representation is called **phonetic transcription**.

Phonetic transcription: a representation of the **pronunciation** of spoken language.

In order to make phonetic transcription effective, we need some kind of standard system to represent the pronunciation of spoken words. Such a system is called a **phonetic alphabet**. It is a set of **phonetic symbols –** symbols that represent speech sounds (speech sounds are also known as **phones** or **segments**). The most important requirement of the phonetic alphabet is that the phonetic symbols will be unambiguous. In other words, every phonetic symbol should represent a single sound, and different sounds should be represented by different symbols. In addition, the phonetic alphabet should be elaborate and flexible enough to allow for the representation of (potentially) all the speech sounds used in any human language.[1]

Phonetic alphabet: a set of symbols representing speech sounds. The phonetic alphabet is constructed such that each symbol represents a single sound, and each sound has a unique phonetic representation.

The requirements listed above make it clear why we need to use a specialized system instead of the conventional writing system (**orthography**) of some language, such as English. Obviously, the orthography of any given language was designed to represent the sounds of that particular language and not the sounds used in other languages. Consider, for example, the Latin script. It serves as the basis of the writing systems of numerous languages of different language families. However, since languages vary considerably in terms of their sound inventories, the Latin script is used differently in different languages. For example, the letter ⟨j⟩ represents different sounds in Standard European English, French, and Spanish (as in the names Jacob, Jacques, and Juan in the respective languages). In particular, the sound of ⟨j⟩ in

[1] Several disciplines focusing on a single language or language family (e.g., Semitic languages) have developed their own phonetic alphabets which include only symbols relevant to the studied languages.

each of these languages is not used (or used marginally) in the other two languages. In addition, even within the same language, a given letter may represent different sounds (e.g., the sound of ⟨u⟩ in the words *put* and *but*).

Moreover, since many languages that use the Latin script have more sounds than distinct letters, the script had to be adapted for each individual language to allow for a more faithful representation of its spoken form. Such adaptations are typically made by adding diacritical marks above and below letters (e.g., á, ç, ñ), or by assigning a special phonetic value to certain letter sequences (e.g., the sound of the sequence ⟨sh⟩ in the word *shine* is not the same as the sequence of sounds of the individual letters ⟨s⟩+⟨h⟩).

Orthography: the writing system of a given language. The basic units of orthography are **letters**, or **graphemes**.

One problem with these adaptations is that they are not done consistently across languages, or even within the same language. For example, in Hungarian, acute accent placed over ⟨u⟩ (i.e., ⟨ú⟩) indicates a long [u] sound, but when placed over ⟨a⟩ (i.e., ⟨á⟩), the accent changes not only the length of the vowel but also its quality.[2] In other languages, accent markers can indicate different phenomena as stress and tone. Thus, there is a need for a standard system to represent, in a consistent manner, the various sound categories used in all human languages, the number of which is in the order of several thousand! (Moran & McCloy, 2019)

Notes on notation:

- The written form of words is indicated in italic font (e.g., *cat*).
- Words in languages other than English are followed by the English translation enclosed between single quotation marks (e.g., the French word *oui* ‘yes’).
- **Angle brackets** ⟨ ⟩ denote single letters or sub-word letter sequences when discussing their pronunciation.
- **Square brackets** [ ] denote the phonetic form of any linguistic expression.
- **Slashes** / / denote hypothesized “underlying” phonetic form (see Chapter 14). Used when discussing phonological processes that transform the hypothesized mental phonetic representation of a word into its spoken form.

[2] I assume the reader is more-or-less familiar with the concepts of vowels and consonants, but I will provide formal definitions of these terms in later chapters.

Example: the hypothesized underlying representation phonetic of the word *bank* is /bænk/, while its spoken phonetic form can be represented by [bæŋk]. The second letter of *bank* is denoted by ⟨a⟩, its underlying mental representation is denoted by /æ/, and its pronunciation is denoted by [æ]. The third letter of *bank* is denoted by ⟨n⟩, its underlying mental representation is denoted by /n/, and its pronunciation is denoted by [ŋ], due to the influence of the following [k].

2.1. Transcription systems

Various phonetic alphabets are being used in linguistic research. This book adopts the **International Phonetic Alphabet (IPA)**, published by the International Phonetic Association (also IPA), which is one of the most popular ones (among phoneticians, at least). The full IPA chart can be found at the beginning of this book, and will be described in detail in the following parts of this book. Among other transcription systems, one notable alternative is the **[North] Americanist Phonetic Alphabet (APA**, or **NAPA)**, especially popular in linguistic domains such as syntax and semantics. Table 2.1 describes some common notation differences between IPA and APA (see full APA chart in Pullum & Ladusaw, 1996).

Table 2.1 Common notation differences between IPA and APA

IPA	APA	Description
d͡ʒ	j	The sound of ⟨j⟩ in *jacket*
j	y	The sound of ⟨y⟩ in *yard*
ʃ	š	The sound of ⟨sh⟩ in *shine*
t͡s	c	The sequence of sounds of ⟨t⟩+⟨s⟩ in languages that treat it as a single consonant, e.g., the sound of ⟨z⟩ in German words like *zwei* ‘two’
t͡ʃ	č	The sound of ⟨ch⟩ in *check*
y	ü	A rounded high front vowel, e.g., the sound of ⟨u⟩ in French *sur* ‘on, upon’
ʒ	ž	The sound of ⟨s⟩ in *pleasure*

In addition, it is noteworthy that many languages has some kind of “r-like” sound. IPA has several symbols that indicate different pronunciations of this sound (e.g., [r], [ʁ], [ɹ]), while APA uses [r] to represent all these variants (at least to the extent that there are no two contrastive r-like sounds in the same language).

2.2. Broad and narrow transcriptions

The idea of phonetic transcription is to represent the way words are pronounced. While this may sound simple, doing phonetic transcription is actually not a very well-defined task. In principle, there is an infinite number of ways to pronounce any given word (see more in Chapter 14). For example, any change in the position and shape of the tongue during speech affects the quality of the produced sound to some extent. This means that, in theory, there is an infinite number of ways to transcribe a word (see also a note on this issue in Section 3.5).

In practice, human transcribers (or humans in general) cannot perceive every tiny change in the pronunciation of words and speech sounds. Fortunately, most pronunciation variations are not meaningful for linguistic research. Thus, one can set an upper limit to the level of precision of phonetic transcription, according to the research goals.

In general, we distinguish two approaches to phonetic transcription. **Narrow (allophonic) transcription** captures as many aspects of a specific pronunciation as possible (as long as they are perceivable and relevant to the research goals). For example, some possible allophonic transcriptions of *emphatic* are [ɛɱfæɾɪk] and [ɛɱfætʰɪk]; where [ɱ] represents the pronunciation of ⟨m⟩ before ⟨f⟩, ⟨v⟩ and ⟨ph⟩; [ɾ] and [tʰ] represent the pronunciation of ⟨t⟩ between vowels in some American and British dialects of English, respectively.

By contrast, **broad (phonemic) transcription** captures only aspects of the pronunciation that are enough to differentiate the word from other words in the language.[3] For example, phonemic transcription of *emphatic* would be something like [emfætɪk]. Here, we transcribed ⟨m⟩ and ⟨t⟩ as [m] and [t], respectively, since the different pronunciations of ⟨m⟩ and ⟨t⟩ are not contrastive in English, and the [t] and [m] symbols are more available than [ɱ], [ɾ] and [ʰ]. On the other hand, we have less freedom with the transcription of vowels because of the various contrastive pronunciations of English vowel letters (though people often use less accurate phonetic transcriptions of vowels, such as [e] instead of [ɛ], for the sake of convenience).

Narrow and broad transcription are general approaches. There is no one standard way of doing either kind. Ultimately, the choice of transcription method depends on the research needs (and also on the available resources), but one is advised to establish explicit transcription criteria and follow them consistently.

[3] The terms *phonemic* and *allophonic* will be explained in Chapter 14.

2.3. Reliability and accuracy of phonetic transcription

As explained in Section 2.2, the outcome of phonetic transcription tends to be variable, when prepared by different transcribers and for different purposes. When preparing phonetic transcriptions, one should be as explicit as possible regarding the conventions used and the qualifications of the transcribers. Moreover, when consulting published transcriptions, one should be aware of potential factors of variability and inaccuracy, which may, or may not, be acknowledged by the authors. Below, I list some of these factors.

First, different authors use different transcription systems. Some, typically in general linguistics, adhere to standard general systems such as IPA and APA. Other linguists, who focus on the descriptive grammar of specific language families (e.g., Semitic linguistics), often use a standard transcription system developed for the specific language family. Yet, some authors develop ad hoc phonetic alphabets for their research (this is probably more common in old texts). Furthermore, while some authors make explicit reference to their transcription conventions, others may not. This is crucial, since many symbols represent different entities in different systems (see, for example, Table 2.1). Thus, when reading texts that contain phonetic transcriptions, it is important to identify the conventions used in the transcriptions.

Second, the accuracy of transcriptions is largely affected by the competence of the transcribers. This point is especially relevant for accounts of indigenous languages made by outsider researchers. Often, these were missionaries who learned the language-in-question during their mission. Being non-native speakers and, perhaps, lacking formal training in phonetics, these researchers produced transcriptions whose quality depended on their phonological awareness and aptitude, as well as on the amount of exposure to the ambient language.

Third, typographical constraints also contribute to variability of transcriptions across texts. This is especially true for texts published before the digital era when, e.g., diacritical marks were not easily produced in print.

Fourth, variability of transcriptions may reflect different language variations, such as dialect and period. Phonology textbooks, in particular, do not always specify when and where a given dataset was collected (or even if the informants were native speakers of the language-in-question). Thus, a language label (e.g., “Spanish”) attached to a dataset may be misleading without additional specifications.

Finally, phonological data listed in textbooks was often altered deliberately, for pedagogical reasons. Occasionally, the authors of such textbooks acknowledge that the data was altered, but do not provide reference to the unaltered data. In other cases, authors do not disclose this fact, and one can only learn that the data was altered by comparing it to other sources or by consulting with native speakers of the language-in-question. Moreover, some datasets are often copied from one textbook to another (possibly with additional changes made in each transition) such that their integrity becomes more obscure with time.

In this book, I use phonological datasets from many resources. For the sake of transparency, I always provide a reference to the original source (except for transcriptions of Modern Hebrew, which I have prepared myself). I refrain from making any changes to the transcriptions other than adapting them uniformly to IPA. Finally, I made an effort to use only datasets that seemed authentic to me, though I cannot guarantee that all the data in this book is accurate.

2.4. Using IPA symbols on the computer

While the Latin script is the source of many phonetic symbols, the IPA charts include many additional symbols that are not readily available on standard keyboards. This section provides some practical information for doing phonetic transcription on the computer.

2.4.1. Fonts installed on the computer

With only a few exceptions, most IPA symbols can be found in many standard character sets (e.g., Times New Roman). Latin letters (e.g., b) found on standard western keyboards are included in the 'Basic Latin' section of the character set. Additional symbols used in various writing systems can be found in the 'Latin-1 Supplement' (e.g., æ), 'Latin Extended' (e.g., ŋ), and 'Greek' (e.g., θ) sections. Some standard character sets come with a built-in 'IPA Extensions section', which include special phonetic letters not included in other sections. Finally, various diacritical marks are included in the 'Spacing Modifier Letters' and the 'Combining Diacritical Marks' section.

In addition to standard fonts included in many systems, there are also fonts especially designed for phonetic transcription. Two such popular fonts are Charis SIL and Doulos SIL. To use these fonts, download and install them according to the instructions.

To use phonetic symbols in Microsoft Office apps go to the 'Insert' menu and select 'Symbol'. Select the desired symbol in the following dialog:

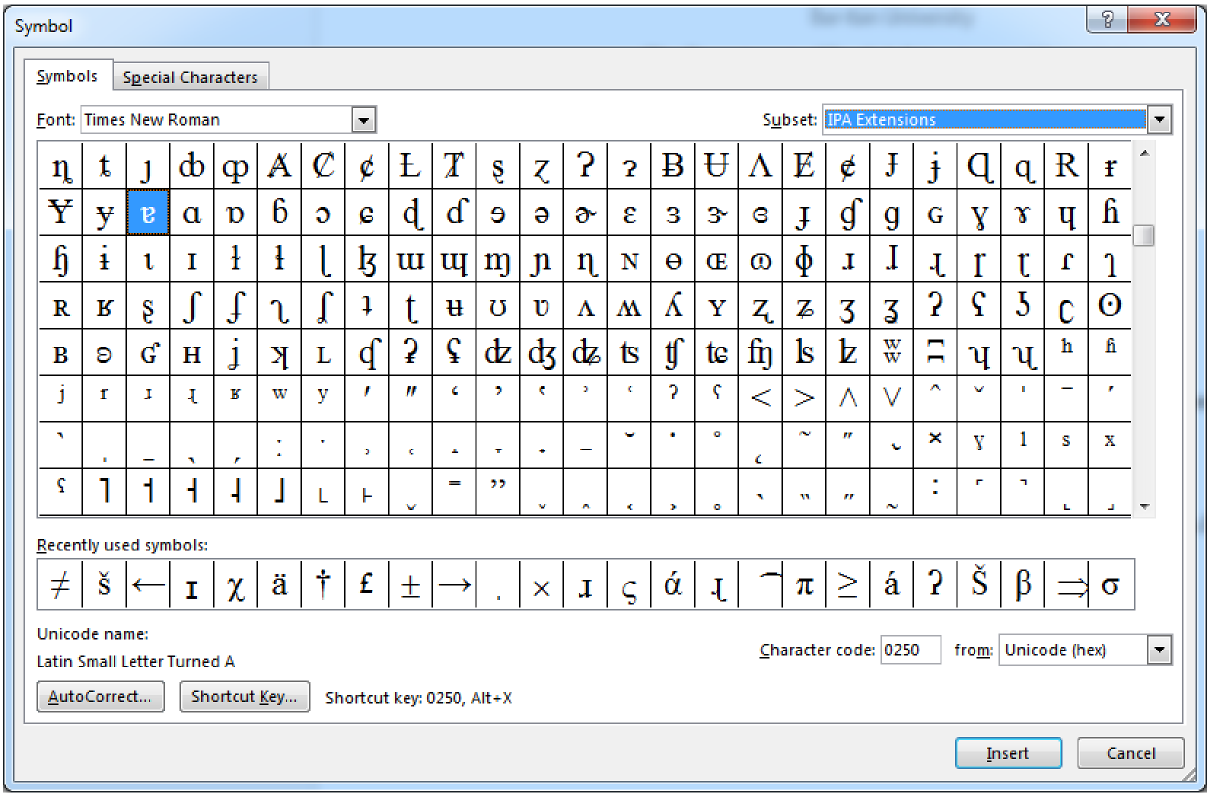


Figure 2.1 Inserting IPA symbols to Microsoft Office documents

2.4.2. **Virtual IPA keyboard in apps and websites**

Some websites and apps include a virtual IPA keyboard that facilitates the process of phonetic transcription. For example, you can type in IPA symbols with the virtual keyboard found in: ipa.typeit.org/full, then copy-and-paste them to your document

2.5. Practical phonetic transcription

This section describes some basic practical principles of phonetic transcription. The principles are demonstrated using a simplified version of transcription guidelines for English words.

DISCLAIMER: doing phonetic transcription requires a lot of training and excellent language skills. It is probably best to rely on native speakers of the relevant language in this task. The present section focuses on the methodology of phonetic transcription. It is included in this book because it is part of the requirements in the university course I taught. Since the majority of my students were not native speakers of English (and neither am I), the level of precision aimed for in this section is probably below the standards used in linguistic research. This is especially true in the case of transcribing vowels, which would be extremely difficult for non-native speakers. Moreover, due to the considerable variation in pronunciation across dialects of English (which affects mainly vowels), there is no one correct way to transcribe English words. Thus, one should not treat this section as a reliable guide to "how to do phonetic transcription of English words".

2.5.1. **Principles of phonetic transcription**

The main principle of phonetic transcription is that **transcription should reflect pronunciation rather than orthography**. This principle has several practical consequences. First, symbols should be used according to their definition in the phonetic alphabet, not according to how they are used in the written language. In particular, if a letter is pronounced differently in different words, it will be represented by different symbols in the phonetic transcription. For example, the pronunciation of ⟨c⟩ is different between the words *cat* and *nice*. Thus, [c] cannot be used in the phonetic transcriptions of both words (in fact, we will use [k] and [s] to represent the pronunciation of ⟨c⟩ in *cat* and *nice*, respectively).

The second consequence is that **silent letters should not be transcribed at all**. This includes cases like ⟨c⟩ in ⟨ck⟩ sequences (e.g., [bæk] 'back'), the sequence ⟨gh⟩ in certain words (e.g., [naɪt] 'night'), double letters pronounced as a single short sound (e.g., ⟨nn⟩ in *dinner* [dɪnɚ]), and so on. Third, "non-combinatorial" letter combinations should be transcribed according to the phonetic convention **– one symbol per sound, rather than one symbol per letter** (e.g., ⟨sh⟩ is transcribed as [ʃ] in [ʃaɪn] 'shine'). An "exception" to the last rule applies when diacritics are used. In particular, a tie-bar can be used to combine two phonetic letters when the combination represents a single sound in the language. For example, the sequence of letters ⟨ch⟩ in *chair* is transcribed in IPA as [t͡ʃ].

Finally, the pronunciation of certain letters changes from context to context due to phonological processes. In principle, as explained in Section 2.2, there could be numerous different pronunciations for each letter. In practice, we will take into account only alternative pronunciations that are perceptible and consistent enough. One common example of this principle is voicing assimilation. For example, the ⟨-s⟩ plural suffix in English is pronounced as [s] after voiceless consonants (e.g., [kæts] 'cats') and as [z] after voiced consonants and vowels (e.g., [dɔgz] 'dogs').

2.5.2. Simplified transcription of English

1. Many English letters representing consonants are used in IPA (more-or-less) in the same way they are used in English orthography. Note, however, that some English letters have more than one possible pronunciation (for many reasons, see for example paragraph 4 below), while the corresponding IPA symbol has one fixed pronunciation. Table 2.2 lists IPA symbols that are more-or-less equivalent to English consonant letters, together with some representative English words demonstrating the usage of these symbols.

Table 2.2 IPA symbols equivalent to English letters

IPA	Examples
b	**b**uy, ca**b**
d	**d**og, hea**d**, la**dd**er
f	**f**an, lea**f**
g	**g**oat, ba**g**
h	**h**igh, a**h**ead
k	s**k**y
l	**l**ie, f**l**y, nai**l**
m	**m**y, s**m**ile, ja**m**
n	**n**ight, s**n**ake, ca**n**
p	**p**ie, s**p**y, ca**p**
s	**s**ing, gra**ss**
t	**t**ie, s**t**ick, ca**t**, la**tt**er
v	**v**ine, lea**v**e
w	**w**ine, s**w**im
z	**z**oo

Notes:

i. The pronunciation of letters in boldface is represented by the corresponding IPA symbol.

ii. [ɡ] is an official IPA symbol, representing the sound of ⟨g⟩ in *good*. [g] is not an official IPA symbol and should not be used, in principle. However, many people use it instead of [ɡ] for the sake of convenience.

2. Some English consonants are represented by IPA symbols that are not English letters or that are used differently in English. The IPA symbols are listed on Table 2.3 together with some representative English words demonstrating the usage of these symbols.

Table 2.3 IPA symbols that are not used, or used differently than the English letters

IPA	Examples
j	**y**es, hallelu**j**ah
ð	**th**is, brea**the**, fa**th**er
θ	**th**ink, ma**th**
ʃ	**sh**ine, ca**sh**, emo**ti**on
ʒ	plea**s**ure, gara**ge**
ɻ / ɹ	**r**ay, t**r**y, ve**r**y
t͡ʃ	**ch**air, ca**tch**
d͡ʒ	**g**iant, ba**dge**, **j**acket
ŋ	si**n**g, si**n**k

Notes:

i. The pronunciation of letters in boldface is represented by the corresponding IPA symbol.

ii. [ɻ] represents general "American" pronunciation of ⟨r⟩, while [ɹ] represents general "British" pronunciation. The IPA [r] is restricted more-or-less to some Scottish and Welsh dialects.

3. Some IPA symbols are English letters but represent sounds that are not used in most dialects of English (or, in other words, the usage of these symbols in English is different from their IPA denotations). Therefore, these letters should **NOT** be used in IPA transcription of English words: c, q, x, y.

4. Phonological processes: as mentioned above, the pronunciation of speech sounds can change in the context of certain sounds due to phonological processes (see more in Part IV of this book). Phonological processes can alter the common pronunciation of some letters, such that the pronunciation of a certain letter may be represented by different IPA symbols in different words. Some notable examples of such phonological processes include:
   - Voicing assimilation of the /-d/ past suffix and the /-s/ plural suffix to the preceding consonant (e.g., [wɑːk**t**] 'walke**d**' vs. [ɻʌb**d**] 'rubbe**d**', [dɔɡ**z**] 'dog**s**' vs. [kæt**s**] 'cat**s**').
   - Place assimilation of /n/: the pronunciation of /n/ changes to [ŋ] before the consonants [ɡ] and [k] (e.g., [ɪ**ŋ**ɡlɪʃ] 'E**n**glish', [sɪ**ŋ**k] 'si**n**k').

5. English vowels letters (⟨a⟩, ⟨e⟩, ⟨i⟩, ⟨o⟩, ⟨u⟩) are less compatible with their IPA definitions. Because the transcription of English vowels is a complicated issue, I do not focus on it in this book. There are only a few principles that I insist on regarding the transcription of English vowels. First, unpronounced letters should not be transcribed. For example, word-final ⟨e⟩ in words like *cake* and *while* is not pronounced (presumably) in any English dialect. Second, a single phonetic letter should be used to transcribe a single vowel quality (e.g., *eat* can be transcribed as [iːt] or [it] but not as [eat]).[4] Third, as mentioned in paragraph 3 above, [y] should not be used in the transcription of English words. When ⟨y⟩ is used as a vowel in English words, it should be transcribed with [i] (e.g., in *really* and *biology*) or [aɪ] (e.g., in *shy*). The IPA usage of [y] (see Chapter 7) is not compatible with the use of the English letter ⟨y⟩.

   With that said, to give the reader some "taste" of the vowel system of English, I list below IPA symbols representing vowels used in various dialects of English. The list is based on General American English pronunciation (Peterson & Barney, 1952). Some complementary information regarding dialectal variations was obtained from Wiktionary (https://en.wiktionary.org).

Table 2.4 IPA symbols for English vowels

IPA	Examples	Dialects
iː	h**ee**d	
ɪ	h**i**d	
ɛ	h**ea**d	
ə	**a**head	
æ	h**a**d	
ɑ	f**a**ther	GenAm, RP
ɔ	d**o**g, b**a**ll	GenAm
ɒ	d**o**g	RP
ʊ	h**oo**d	
uː	wh**o**	
ʌ	h**u**t	
ɜː	h**ea**rd	RP
ɝ	h**ea**rd	GenAm

Notes:

i. The IPA symbols represent the pronunciation of the boldfaced letters in the corresponding words.

[4] In fact, there is some dialectal variability, where certain vowels are pronounced as monophthongs (i.e., have a single quality) in some dialects and as diphthongs in other dialects. In basic transcription training, it is recommended to use examples that are not influenced by such variability.

ii. The pronunciation of vowels vary considerably across dialects of English. The ‘Dialects’ column specifies the relevant dialects for each pronunciation (GenAm = General American, RP = Received Pronunciation (Standard British)). This information was retrieved from Wiktionary. When no dialects are specified, the information applies at least to General American English.

6. **Diphthongs**: diphthongs are vowels whose quality changes within the syllable (see Chapter 7). Examples include ⟨i⟩ in *five*, ⟨a⟩ in *fame*, and ⟨o⟩ in *no*. There are alternative ways to transcribe such vowels (each alternative relies on a different assumption but both transcriptions reflect the pronunciation equally well).
   - [aj] = [ai]: the sound of ⟨i⟩ in *five* ([fajv] = [faiv]).
   - [ej] = [ei]: the sound of ⟨a⟩ in *fame* ([fejm] = [feim]).
   - [ow] = [ou]: the sound of ⟨o⟩ in *no* ([now] = [nou]).

2.5.3. Practice

Transcribe the following English words using the International Phonetic Alphabet (IPA):

1. big
2. fish
3. bench
4. class
5. thought
6. this
7. united
8. linked
9. psychology
10. measures

Solution to sample exercise

	IPA	**Explanation**
1.	[big]	Notice the use of the official IPA symbol [ɡ].
2.	[fiʃ]	One symbol per sound (⟨sh⟩ = [ʃ]).
3.	[bent͡ʃ]	Use the tie bar ͡ to combine two phonetic letters to represent a single consonant.
4.	[klas]	Double letter (⟨ss⟩), single sound.
5.	[θot]	Many silent letters.
6.	[ðis]	Different pronunciations of the same sequence of letters (⟨th⟩, cf. 5).
7.	[junaited] / [junajted]	Sound without letter ([j]). Two ways for transcribing diphthongs ([ai] = [aj])
8.	[liŋkt]	Different pronunciations of the ⟨-ed⟩ suffix, assimilations of /n/ to a following [ɡ] or [k], assimilation of /d/ after voiceless consonants like [k].
9.	[saikolod͡ʒi]	More silent letters.
10.	[meʒeɻz] / [meʒeɹz]	"American" vs. "British" pronunciations of ⟨r⟩.

Part II: Articulatory phonetics

Articulatory phonetics studies the production mechanisms of speech. Understanding these mechanisms is important to understanding various phenomena, including phonological processes, certain speech disorders, and vocal expression of emotions. Articulation research utilizes various instruments to observe speech organs in action and measure their activity.

Imaging techniques such as videostroboscopy, Real-time MRI, X-ray, and Ultrasound can produce videos of moving organs, allowing exact determination of the duration and speed of articulation, as well as the exact configuration of the **articulators** (Juslin & Scherer, 2005). Other devices measure the electrical activity of speech organs to assess articulation. For example, Electroglottography (EEG) measures the degree of contact between the vibrating vocal cords during voice production, while electropalatography (EPG) monitors the contact between the tongue and hard palate. Additional techniques include electromyographic (EMG) measurement of muscle activity and assessment of breathing patterns via various devices (e.g., thermistors that measure temperature differences between inhaled and exhaled air, and strain gauges that measure chest movement).

Articulator (speech organ): Any specific part of the vocal tract which may be involved in the articulation of a segment.
Vocal tract: the pathway through which the air flows during speech.

This part of the book describes the articulatory processes underlying the production of speech sounds. The descriptions are qualitative and focus on the position and motion patterns of speech organs. Configurations of speech organs are illustrated on snapshots extracted from animation clips that demonstrate the articulation of speech sounds. The animations were obtained from the *Seeing Speech* project (Lawson et al., 2018) (a list of links to the animations used in this book appears in Appendix C). The descriptions are also accompanied by links to real-time MRI video clips (also from the *Seeing Speech* project) demonstrating the articulation of speech sounds by a real human.

Chapter 3 Consonants – basic description

A **consonant** is a segment (speech sound) whose articulation involves a significant obstruction to air flow in the vocal tract (Trask, 1996). Examples include: [b], [k], [m], and [w]. In this chapter we will cover the basic terms needed to understand consonants. In Chapter 4, we will explore various types of consonants in more detail.

3.1. Place of articulation

Let's start with some short exercises. Try to produce the sound [d] (as in the name of the letter *d*) and pay attention to the process of articulation. You can also watch a real-time MRI demonstration of the articulation of [d] here. You might have noticed that the front part of the tongue briefly touches an area behind the upper teeth (or even the teeth themselves), and then retracts, while producing an "explosive" burst of air. Now, try to produce the sound [g] (as in the word *good*) and pay attention to the process of articulation (watch). You might have noticed that the tongue briefly touches an area at the roof of the mouth, and then retracts, while producing an "explosive" burst of air.

Note! Symbols and words enclosed between square brackets (e.g., [d]) represent the pronunciation of the corresponding speech sounds and words, according to the conventions of the phonetic alphabet.

Let's compare the articulation of [d] and [g]. Both consonants involve a brief contact between the tongue and the upper part of the mouth that ends with an "explosive" burst of air. In other words, the articulation mechanism is similar. What distinguishes these consonants is the place of contact between the tongue and upper surface of the mouth (the part of the tongue involved in the articulation is also different). In [d], the tongue touches an area behind the upper teeth called the **alveolar ridge** , in [g] the back part of the tongue touches an area called the **velum** , or **soft palate** (In the next chapter we will elaborate on these anatomical landmarks). We say that the articulation of [d] and [g] differs with respect to **place of articulation**.

Place of articulation: the place in the vocal tract where airflow is interrupted during speech production.

Place of articulation is one of the major features of consonants. The place of articulation of [d] is the **alveolar ridge** – this is where the tongue touches the roof the mouth. [d] is an **alveolar** consonant. The place of articulation of [g] is the **velum** – this is where the tongue touches the roof the mouth. [g] is a **velar** consonant.

Now, let's examine the articulation of [b] (watch). Like [d] and [g], [b] involves contact between two articulators and has an "explosive" burst quality. The difference between the consonants is, again, with respect to place of articulation (i.e., the place of contact). [b] involves contact between the lips. It is a **bilabial** consonant (bilabial = involving the two lips). Figure 3.1 illustrates the place of articulation of [b], [d], and [g].

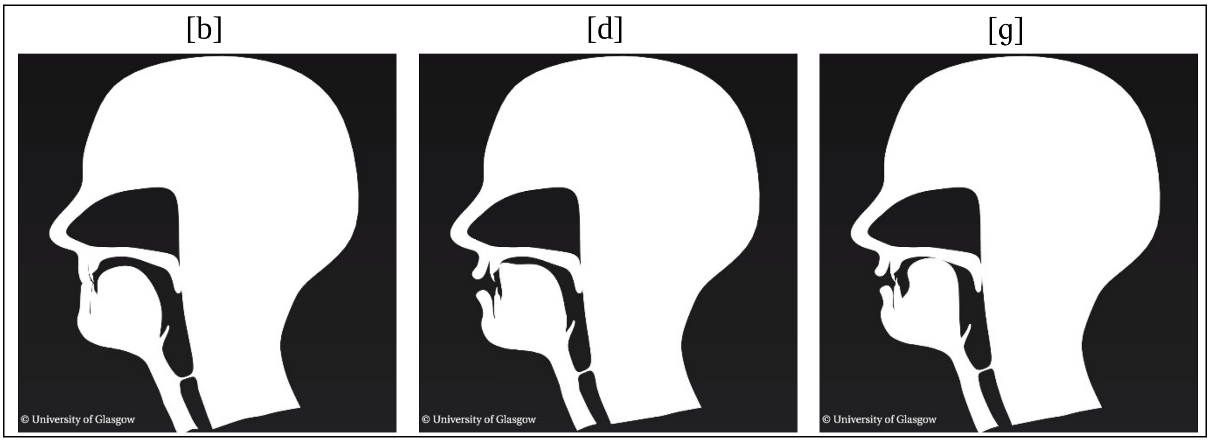


Figure 3.1 Place of articulation. The figure shows the place of articulation of three consonants: [b] is a bilabial consonant (involving contact between the lips), [d] is an alveolar consonant (involving contact between the front part of the tongue and the alveolar ridge), and [g] is a velar consonant (involving contact between the back part of the tongue and the velum).

3.2. Manner of articulation

Next, consider the articulation of [n] (watch). You might have noticed that, like [d], [n] involves contact between the front part of the tongue and the alveolar ridge. Yet, [n] does not have the "explosive" quality we observed in [d]. Instead, [n] involves "smooth" and continuous sound. This means that the difference between [d] and [n] is not due to the place of articulation. Rather, there is some difference in the mechanism of sound production in the two consonants. If you look closely at the video of [n], you can see a "valve" opening at the upper back part of the mouth at the same time the tongue touches the alveolar ridge. This "valve" is the velum (soft palate). When the velum is down, air coming up from the lungs can flow into the nose and out. Thus, despite the fact that the tongue is blocking the airflow in the mouth, nasal airflow enables the continuous sound of [n]. So, we say that the articulation of [d] and [n] differs with respect to **manner of articulation**.

Manner of articulation: the way in which airflow is interrupted during speech production.

Manner of articulation is one of the major features of consonants. The manner of articulation of [d] is **plosive**, or **oral stop** (or, simply **stop**) – it involves full contact between articulators, during which airflow through the mouth is stopped completely. When the obstruction is released it is accompanied by an explosive burst of air (see more details in Chapter 5). The manner of articulation of [n] is **nasal**. It involves full contact between articulators, which block airflow through the mouth. At the same time, the velum is lowered, allowing air to flows through the nose instead of the mouth.

Now, consider the articulation [z] (watch). Like [d] and [n], [z] seems to involve contact between the front part of the tongue and the alveolar ridge. Yet, it sounds very different from both [d] and [n]. Like [n], [z] involves continuous sound production, which means that airflow is not blocked completely. In fact, if you try pronouncing [z], you might notice that the front part of the tongue is not pressed against the alveolar ridge (the contact observed in the video probably involves the sides of the tongue rather than its tip). Rather, there is some gap between the tip of the tongue and the alveolar ridge which allows air to flow continuously. But, since this gap is narrow, air particles flowing through the gap tend to collide with each other. The result is an audible friction noise. Thus, the articulation mechanism of [z] is different than those of [d] and [n]. In other words, we have, again, a difference in the manner of articulation. [z] is a **fricative** – it involves close approximation between the articulators, leaving a narrow gap through which air flows continuously with a friction noise. Figure 3.2 illustrates the manner of articulation of [d], [n], and [z].

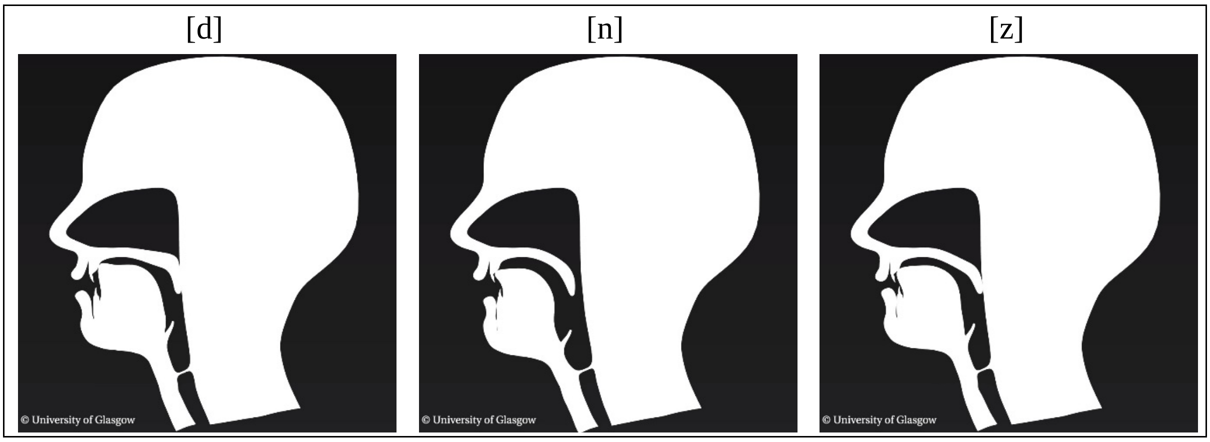


Figure 3.2 Manner of articulation. The figure shows the manner of articulation of three consonants: [d] is a plosive consonant, [n] is a nasal consonant, and [z] is a fricative consonant.

3.3. Voice

As a final exercise, place your fingers around the center of your throat, while pronouncing a long [z] sound. Now, try this exercise again, but this time pronounce a long [s] sound. First, you might notice that both consonants have a similar noisy quality, suggesting a similar manner of articulation (both are fricatives). Second, the configuration of the mouth is more or less the same – the tip of the tongue comes into close approximation with the alveolar ridge. This suggests that the consonants have the same place of articulation (both are alveolar). So what is the difference? You might have felt some vibrations in your throat when you pronounced [z], but not when you pronounced [s].

To understand where these vibrations come from, let us examine Figure 3.3, which illustrates the throat when observed from above. The figure shows some structures located above the **trachea** (the tube leading air in and out of our lungs). In particular, note a pair of bright stripes labeled 'Vocal fold'. The **vocal cords** (also known as **vocal folds**) are a pair of tissues at the top of the trachea. The vocal cords may vibrate during speech production. When they vibrate, we produce a **voiced** sound. If the vocal cords don't vibrate during speech production, we produce a **voiceless** (or, **unvoiced**) sound. Thus, we say that the contrast between [z] and [s] is due to **voice** (or, **voicing**). [z] involves vibrations of the vocal cords – it is voiced, while [s] involves no such vibrations and, is therefore, voiceless (or, unvoiced).

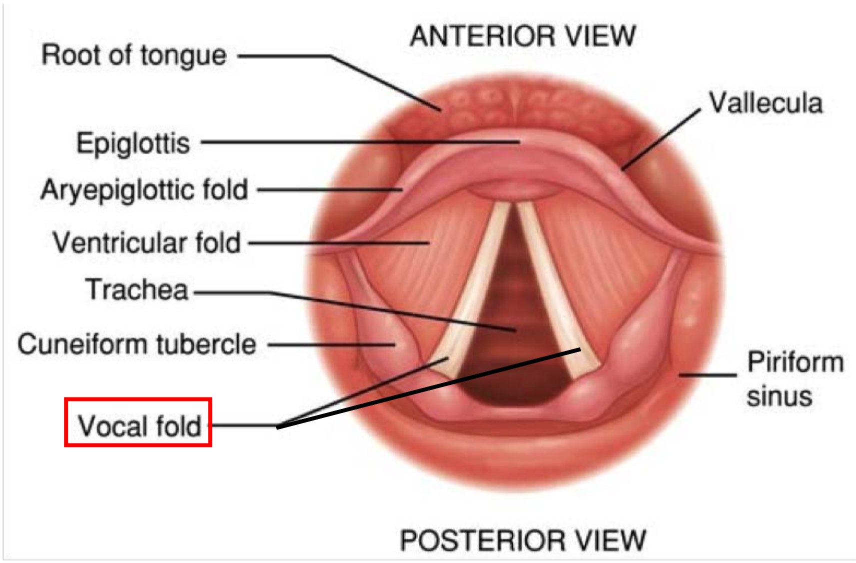


Figure 3.3 The throat from above. The vocal cords (vocal folds) are a pair of tissues at the top of the trachea. Vibrating vocal cords produce voiced sounds, while non-vibrating vocal cords during speech production produce voiceless (unvoiced) sounds.

Voice (or Voicing): The presence or absence of vibration of the vocal cords during an articulation.

To summarize, **consonants** are speech sounds whose articulation involves a significant obstruction to airflow in the vocal tract. Consonants are typically classified according to three main features:

- **Place of articulation**: the place in the vocal tract where airflow is interrupted during speech production.
- **Manner of articulation**: the way in which airflow is interrupted during speech production.
- **Voice (Voicing)**: the state of the vocal cords during speech production (vibrating/not vibrating).

Chapter 4 describes the various places of articulation in more detail, while Chapter 5 discusses the various manners of articulation.

3.4. The IPA chart

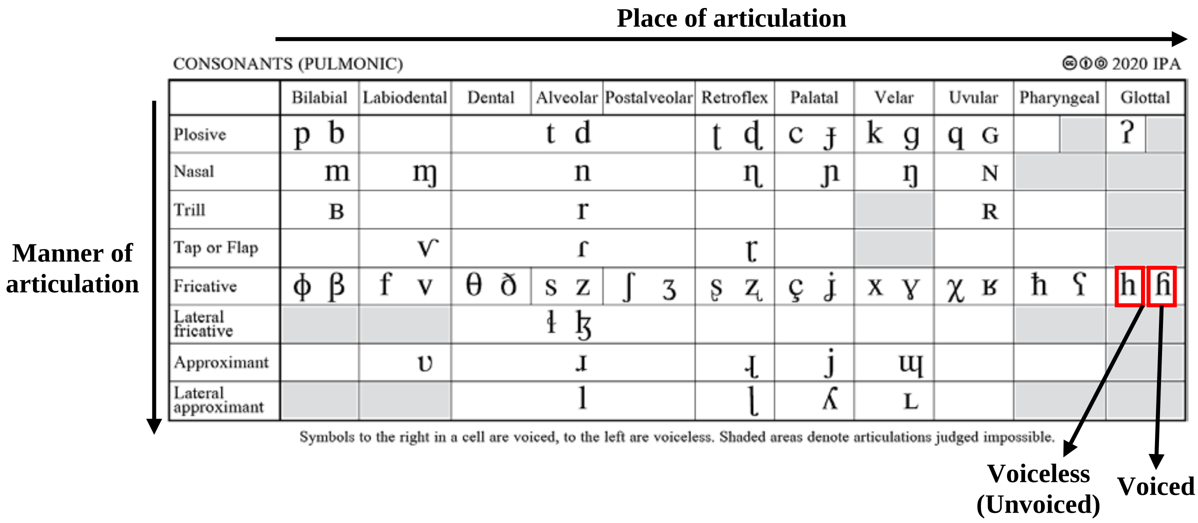
The CONSONANTS (PULMONIC) [5] table of the IPA (International Phonetic Alphabet) sheet organizes the consonants according to their articulatory properties (Figure 3.4). The columns of the table organize consonants by Place of articulation. The rows of the table organize consonants by Manner of articulation. Finally, the horizontal alignment of symbols within cells indicate their voice – symbols aligned to the right represent voiced consonants, symbols aligned to the left represent voiceless consonants.

The conventional way to read the symbols in the Consonants chart is by specifying the articulatory features of the consonants in the following order: [voice] [place of articulation] [manner of articulation]. For example, [p] is a voiceless bilabial plosive, [s] is a voiceless alveolar fricative, and [ʎ] is a voiced palatal lateral approximant.

We can see that some cells in chart are empty. For example, there are no symbols in the intersection of the 'Bilabial' column and the 'Approximant' row. In addition, some cells contain two symbols – one aligned to the right, one aligned to the left – while other cells contain only one symbol which is aligned to the right. For example, there is no left aligned symbol in any of the cells of the 'Nasal' row. These existence of these empty slots in the chart means that the IPA has no designated symbols for representing the corresponding consonants, presumably because they are relatively rare cross-linguistically (e.g., voiceless nasals). Consonants with no designated symbol can be represented by combinations of some consonant symbol and a diacritical mark from the DIACRITICS section of the IPA chart (see more below).

Finally, some cells in the chart are gray-shaded. They represent articulatory combinations judged to be physically impossible. For example, glottal nasal consonants are considered to be impossible since they require complete closure of the **glottis** (the opening between the vocal cords), while allowing continuous nasal airflow. However, complete closure of the glottis means that air does not flow out of the trachea and, thus, cannot reach the nasal cavity.

[5] The term *Pulmonic* ('related to the lungs') refers to airstream mechanism. See more in Chapter 8.



	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Figure 3.4 Consonants (Pulmonic).This section of the IPA chart organizes consonants according to their Place of articulation (columns), Manner of articulation (rows), and voice (horizontal alignment within cells).

3.4.1. Representing missing segments with diacritics

With the help of the DIACRITICS section of the IPA we can represent consonants that have no designated letter. As can be seen in Figure 3.4, many cells in the IPA consonant chart contain a symbol for a voiced consonant, but not a symbol for its voiceless counterpart. The missing voiceless consonants can be represented by a combination of a voiced consonant and the voiceless under-ring diacritic. For example, a voiceless alveolar nasal can be represented by the voiceless under-ring diacritic placed under the voiced alveolar nasal [n], i.e., [n̥].

Other diacritics modify the manner of articulation of the hosting consonant. The lowered sign placed under a fricative can be used to represent an approximant in the same place of articulation. For example, a voiced uvular approximant can be represented by a combination of a voiced uvular fricative [ʁ] and the lowered sign, i.e., [ʁ̞]. The raising sign can be used to turn a lateral approximant into a lateral fricative in the same place of articulation (e.g., a velar lateral fricative [ʟ̝]).

Some diacritics modify the place of articulation of the hosting consonant. The dental diacritic can be used with alveolar consonants to represent dental consonants (e.g., a voiceless dental plosive [t̪]), and with bilabial consonants to represent labiodental consonants (e.g., a voiced labiodental plosive [b̪]). The retracted diacritic can be used to turn alveolar consonants to postalveolar (e.g., postalveolar nasal [n̠]) and velar to uvular (e.g., uvular approximant [ɰ̠]). Finally, the extra-short breve, which can be found in the SUPRASEGMENTALS section of the IPA, can turn a plosive or a trill into a tap or flap (e.g., a voiced uvular tap or flap [ɢ̆] or [ʀ̆]).

In rare cases, a consonant symbol needs to be combined with more than one diacritic in order to represent another consonant. For example, a voiced uvular lateral fricative can be represented by modifying the symbol for a voiced velar lateral approximant by the retracted and raised symbols, i.e., ʟ̠̝.

3.5. A note on descriptive accuracy

The labels used in descriptions of sounds are categorical. For example, a "voiceless alveolar fricative" would be used as a label for any consonant produced by bringing the front part of the tongue into close approximation to the alveolar ridge while not producing vibrations in the vocal cords. However, such labels are not absolute descriptions of how the sound is actually pronounced. A "voiceless alveolar fricative" (or, simply [s]) may exist in two different languages/dialects, but may be pronounced somewhat differently by speakers of the respective languages/dialects.

Different pronunciations of the same descriptive label can be due to slightly different configurations of the articulators. Different configurations can result from different shape of the tongue or the lips, or a slight shift in where the tongue comes into contact/approximation to the upper surface of the mouth. A change in configuration of the articulators affects the flow of air during speech, thus resulting in different sounds.

One may ask, whether using categorical labels in descriptions of sounds is justified, given that the label does not always reflect pronunciation accurately. Indeed, when producing computerized models of articulation, more accurate descriptions may be called for. In such cases, exact numerical values might be more appropriate than categorical labels and descriptions. However, in phonological research, which typically focuses on linguistic contrasts, infinitely accurate descriptions are not very feasible, or even useful (see also Section 2.2). One may be perfectly correct in using the label [s] for any voiceless alveolar fricative, as long as the language/dialect in question does not contrast two such consonants. If a language contrasts two such consonants (e.g., one produced with the tip, the other with the blade of the tongue), diacritical marks can be added to express this contrast (e.g., s̺ and s̻, respectively).

Chapter 4 The articulators and places of articulation

This chapter reviews the main articulators involved in the production of speech sounds and describes the various places of articulation of consonants according to the IPA chart. In addition to the articulators, which regulate the air flow, it is important to mention to pathway through which the air flows during speech, i.e., the **vocal tract**, illustrated in Figure 4.1.

Most speech sounds use air coming out from the **lungs** (see other types in Chapter 8). The air flows through the **trachea** (which is also used for inhaling air) and arrives at the **larynx** (the "voice box"). This complex structure contains the vocal cords, which are responsible for the voicing feature of speech sounds (see Chapter 3). The structure of the larynx is illustrated in Figure 3.3 and its general location in the body is indicated at the bottom of Figure 4.1.[6] The space above the larynx is the **pharynx**, located at the back of the mouth. Air flowing up the pharynx can proceed along the **oral tract (oral cavity)** and the **nasal tract (nasal cavity)** on its way out.

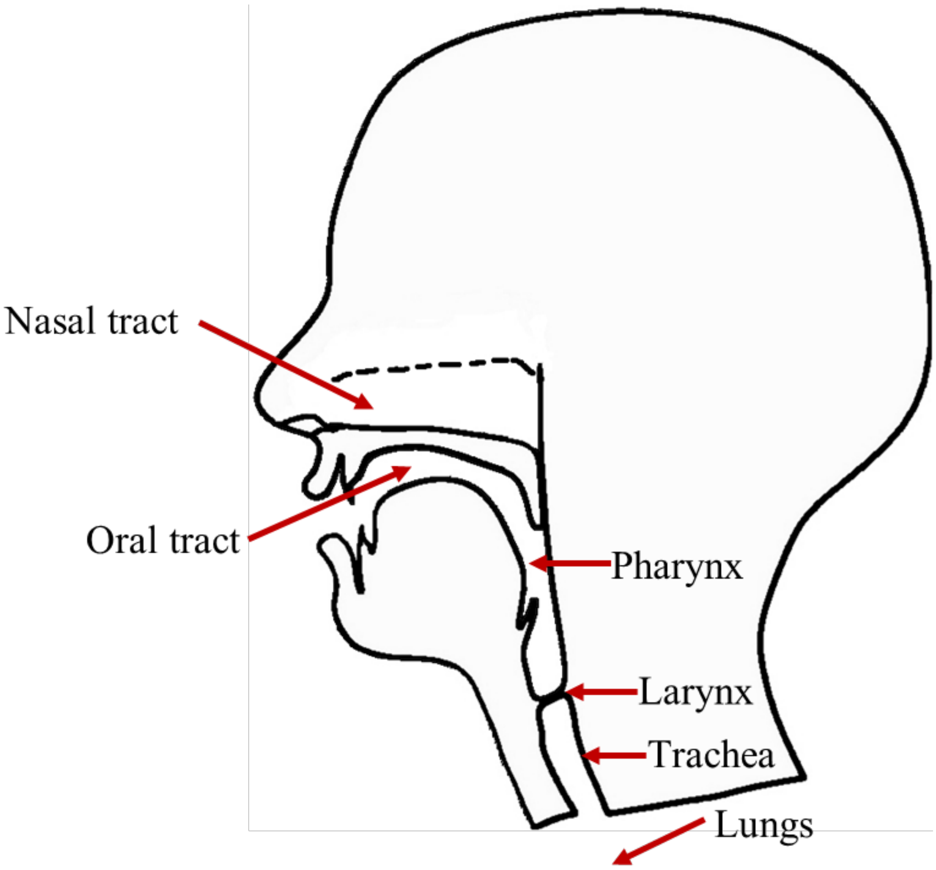


Figure 4.1 The vocal tract (adapted from: https://www.seeingspeech.ac.uk/ipa-charts)

[6] In general, the larynx is involved in the process of **phonation** (voicing is a sub-type of which) and contributes to the overall voice quality of the individual (see Chapter 9).

Vocal tract: The pathway through which air flows or can flow during the production of speech, running from the **lungs** through the **trachea**, the **larynx** and the **pharynx** and out through the **oral** and **nasal cavities**. (Trask, 1996)

The articulation of a consonant involves an interaction between two articulators (at least). The moving articulator is called the **active articulator**, and the static articulator is called the **passive articulator**. Figure 4.2 labels the passive articulators, while Figure 4.3 labels the active articulators. The passive articulators include the **upper lip**, the **teeth**, the upper surface of the mouth, the wall of the **pharynx**, and, to some extent, the **glottis**. Note that, although both lips can move, the upper lip, attached to the upper jaw, is rather fixed to its position and, thus, considered a passive articulator. The active articulators include the **lower lip**, different parts of the tongue and, to some extent, the walls of the pharynx, the glottis.

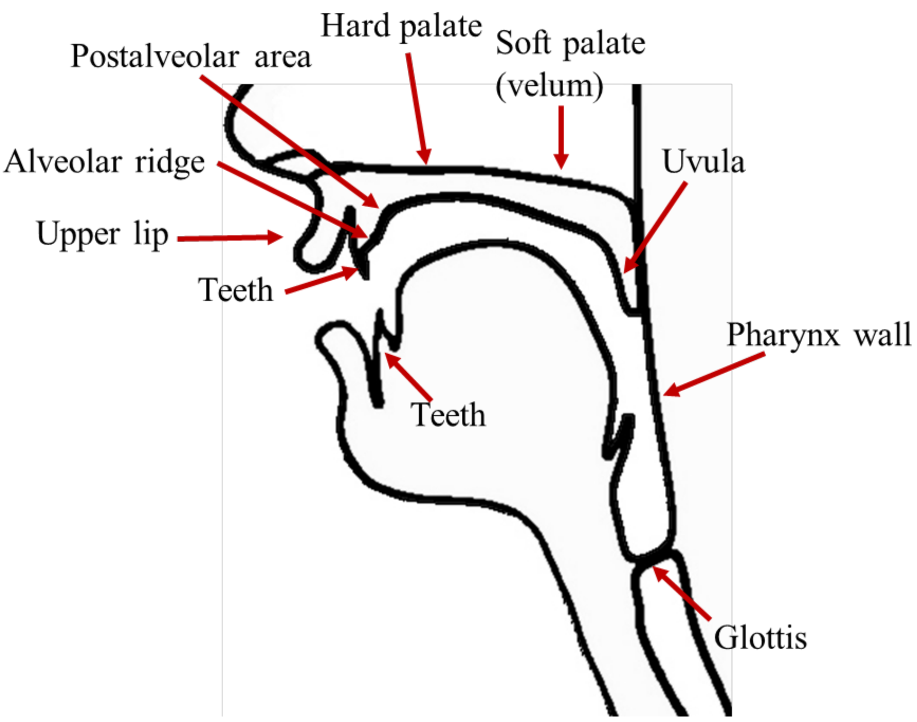


Figure 4.2 The passive articulators (adapted from: https://www.seeingspeech.ac.uk/ipa-charts)

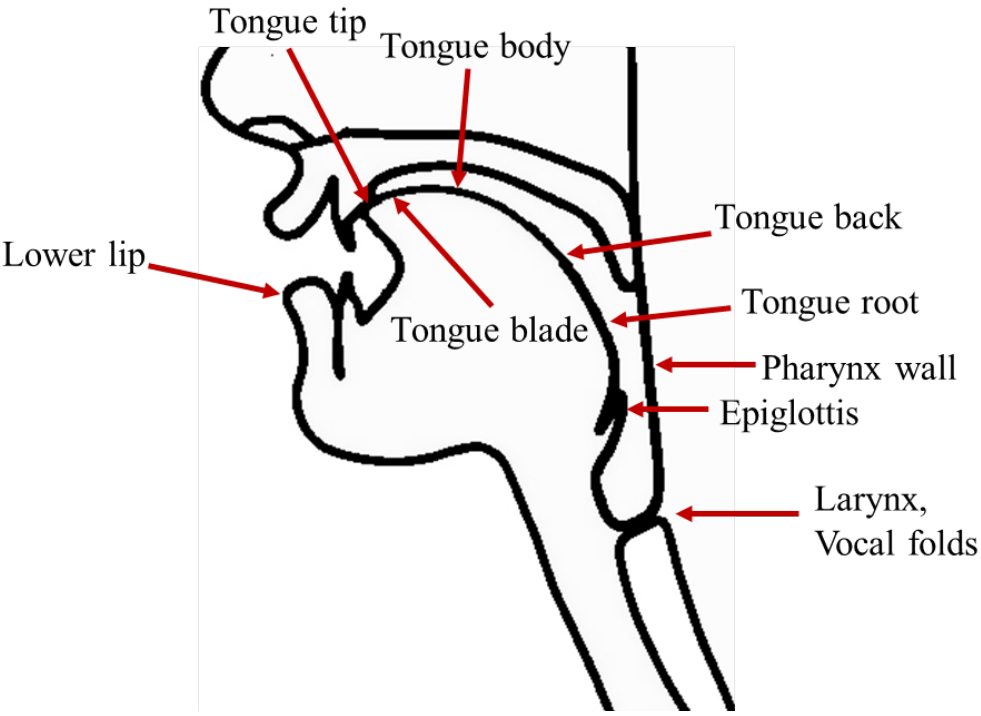


Figure 4.3 The active articulators (adapted from: https://www.seeingspeech.ac.uk/ipa-charts)

The remaining of this chapter reviews the articulators and the places of articulation associated with them. The descriptions are accompanied by animation snapshots taken from the *Seeing Speech* project (Lawson et al., 2018). Each figure labels the relevant active (A) and passive (P) articulators. Note that all figures demonstrate contact between the articulators. However, other degrees of interaction are possible at all places of articulations. Varying degrees of interaction give rise to different manners of articulation. These are discussed in Chapter 5.

Figure 4.4A illustrates **bilabial** place of articulation. As the name suggests, bilabial involves some interaction between the lips. As noted above, the upper lip is the passive articulator, while the lower lip is the active articulator. Examples of bilabial consonants include: [b], [p], and [m]. **Labiodental** consonants, like [f] and [v], are illustrated in Figure 4.4B. They involve interaction between the lower lip (active) and the upper teeth (passive).

Dental consonants involve interaction between the tongue and the upper teeth. Dental articulation is illustrated in Figure 4.5A. The IPA chart contains two symbols in the Dental column. However, these symbols actually represent **interdental** consonants, in which the tongue protrudes between the teeth, as illustrated in Figure 4.5B. The IPA [θ] and [ð] symbols represent interdental fricatives, which represent the sound of ⟨th⟩ in English words like *think* and *that*, respectively. Thus, there are no "true" dental consonants in the IPA chart. To represent dental consonants, you can use a phonetic letter representing an alveolar consonant (see below) of the same manner and voice, and place the subscript bridge diacritic under it. For example, a voiceless <u>dental</u> fricative can be represented by [s̪], to contrast it with the voiceless <u>interdental</u> fricative [θ].[7]

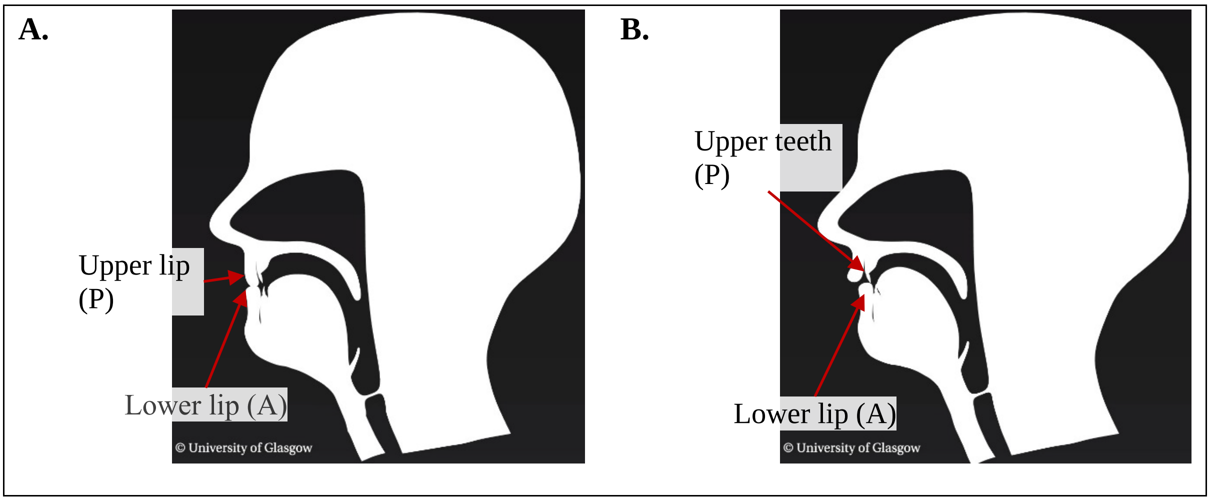


Figure 4.4 Bilabial (A) and Labiodental (B) consonants

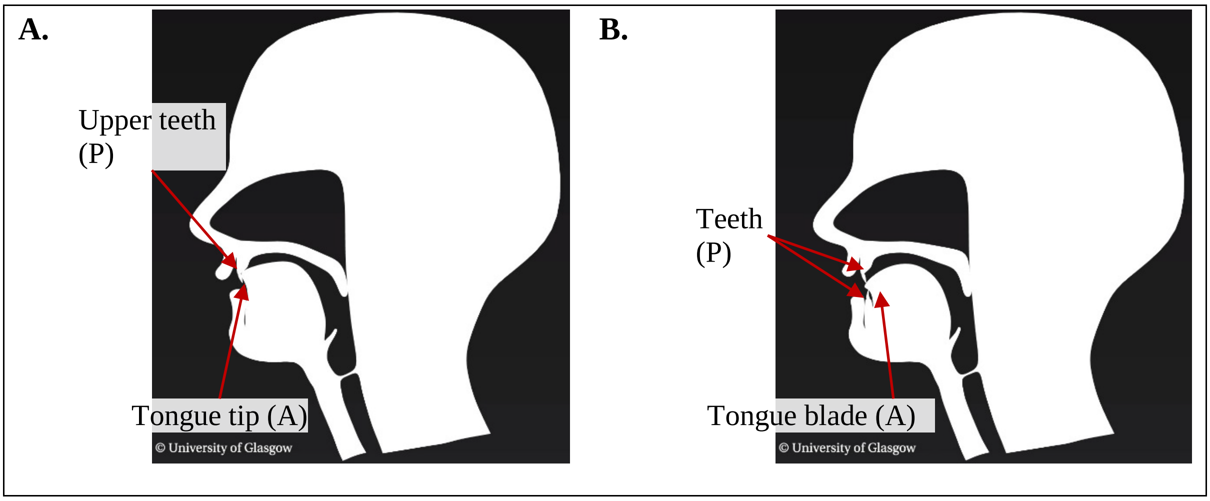


Figure 4.5 Dental (A) and Interdental (B) consonants

The active articulator in dental consonants is either the **tip of the tongue** or the **blade of the tongue**, a small area just behind the tip. In interdental consonants the active articulator is

[7] In many languages (including some dialect of English), consonants like /t/, /d/, /s/ are dental. However, since there is rarely any contrast between alveolar and dental consonants, the subscript bridge diacritic is often omitted for convenience.

the blade of the tongue. The tip and the blade of the tongue are the smallest but most flexible parts of the tongue. The IPA consonant chart does not have designated letters for consonants produced with the tip or the blade of the tongue, since they are not contrastive in most languages. However, some languages do contrast consonants based on the exact part of the tongue involved in the articulation. When the active articulator is the tip of the tongue, the consonant is referred to as **apical**, while consonants produced with the blade of tongue are called **laminal**. Apical and laminal articulations can be indicated in transcription using special diacritics – inverted subscript bridge for apical (e.g., [t̺]), and a subscript square for laminal (e.g., [t̻]).

The articulation of most consonants takes place at the upper surface of the mouth, which can be divided into several sections. The **alveolar ridge** is the rather horizontal surface immediately behind the upper teeth. Despite its small area it is the site of some of the most common consonants, like [t], [s], and [n]. Consonants articulated at the alveolar ridge are called **alveolar**. The back part of the alveolar ridge, where the inclination of the surface changes between a flat and a sharp angle, is called the **postalveolar area**, and consonants produced at that location are called **postalveolar** (Postalveolar consonants are also sometimes called **palato-alveolar**, or **alveo-palatal**. See Ladefoged & Johnson (2010) for discussion of these terms.). Common postalveolar consonants are [ʃ] and [ʒ], which represent the sound of ⟨s⟩ in the words *sure* and *pleasure*, respectively. The active articulator in alveolar and postalveolar consonants is either the tip or the blade of the tongue. Figure 4.6 illustrates the configuration of the articulators in alveolar and postalveolar consonants.

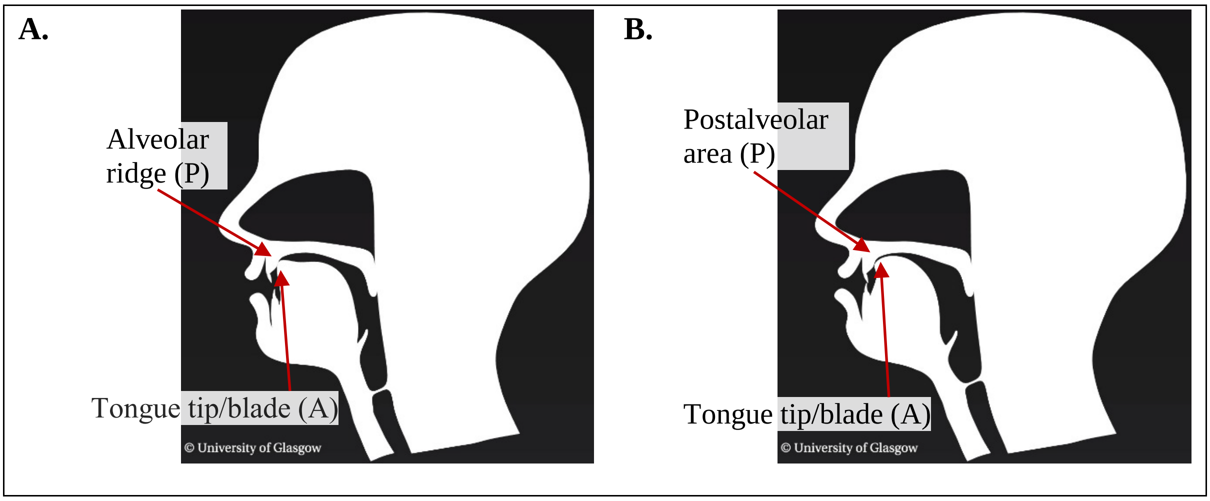


Figure 4.6 Alveolar (A) and Postalveolar (Palatoalveolar) (B) consonants

The next place of articulation according to the IPA chart is, **retroflex**. Unlike other terms, the term *retroflex* reflects the active articulator rather than the passive articulator. Retroflex consonants are produced with a curled tongue, such that the tip of the tongue or the bottom surface of the tongue (the tongue underblade) interacts with the roof of the mouth (see Figure 4.7A). The passive articulator in retroflex consonants is usually the postalveolar area or the hard palate. In certain varieties of American English ⟨r⟩ is typically retroflexed [ɻ], while Hindi contrasts dental and postalveolar retroflex plosives. As can be seen, retroflex consonants are typically represented by letters of alveolar consonants modified by the retroflex hook.

Table 4.1 Dental and retroflex plosives in Hindi (Indo-European; India) (Ladefoged, 2005)

	Dental			Retroflex (postalveolar)		
Voiced	d̪al	'lentil'	(listen)	ɖal	'branch'	(listen)
Voiceless	t̪al	'beat'	(listen)	ʈal	'postpone'	(listen)

Palatal consonants are produced by an interaction of the **body of the tongue** (or, **front of the tongue**) and the **hard palate** (see Figure 4.7B). The hard palate is the hard surface of the roof of the mouth, extending backwards from the postalveolar area, and the body of the tongue is defined as the part of the tongue that lies underneath the hard palate when the tongue is at rest. English has one palatal consonant, [j], which represents the sound of ⟨y⟩ in the word *yes* is a palatal consonant. Hungarian has several additional palatal consonants, as exemplified in Table 4.2.

Table 4.2 Palatal consonants in Hungarian (Uralic; Hungary) (Ladefoged, 2005)

Voiced plosive	ɒɟɒ	'his brain'	(listen)
Voiceless plosive	ɒcɒ	'father'	(listen)
Nasal	ɒɲɒ	'mother'	(listen)

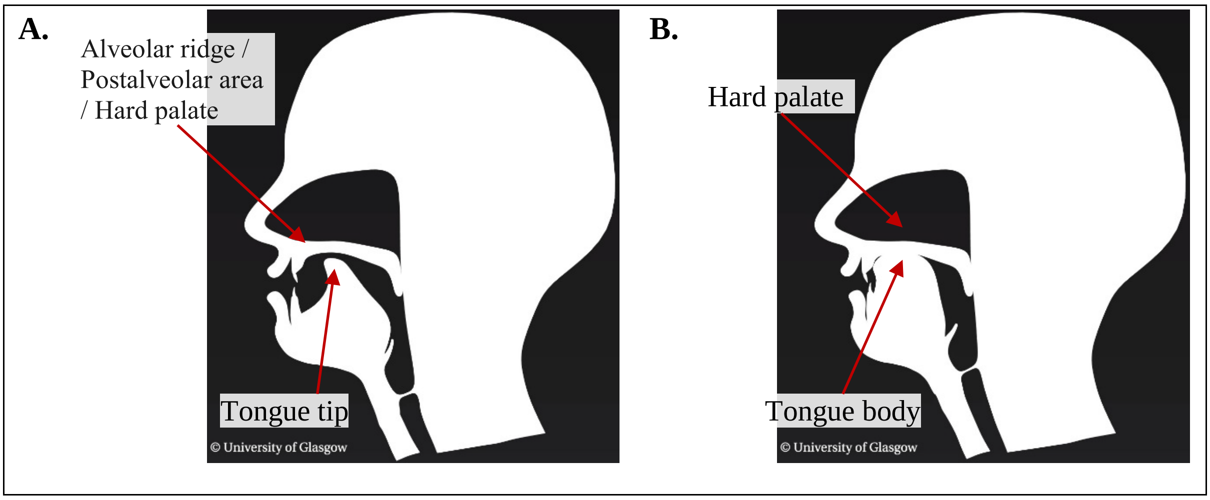


Figure 4.7 Retroflex (A) and Palatal (B) consonants

At the back of the hard palate, the surface changes from a hard tissue into a soft, flexible tissue called the **soft palate**, or **velum**. The part of the tongue that lies underneath the velum when the tongue is at rest is called the **back of the tongue**. Accordingly, **velar** consonants are produced by an interaction of the back of the tongue and the velum (see Figure 4.8A). Examples for velar consonants include [k], [ɡ] (as in the word *good*), and [ŋ] – the typical pronunciation of ⟨n⟩ before the consonants [k], [ɡ] (e.g., in the words *bank* and *king*). Note that, while the velum typically functions as a passive articulator, it also has an active role in the articulation of nasal consonants and vowels – lowering the velum allows air to flow through the nasal cavity in these segments (see also Sections 3.2, 5.2, and 7.2).

At the back end of the velum there is a piece of hanging tissue, called the **uvula**. Accordingly, **uvular** consonants are produced by an interaction of the back of the tongue and the uvula (see Figure 4.8B). Uvular consonants are found, for example, in Modern Hebrew:

Table 4.3 Uvular consonants in Modern Hebrew (Semitic; Israel)[8]

Voiced approximant [ʁ̞]	ʁ̞oʃ	‘head’	(listen)
Voiceless fricative [χ]	mɛlɛχ	‘king’	(listen)

Note: there is no IPA symbol for a voiced uvular approximant. This consonant can be represented by putting the lowering diacritic under the symbol for a voiced uvular fricative [ʁ].

[8] The recordings were made by the author of this book. There is some debate about the identity of these consonants, and there is also some inter-speaker variation in their pronunciation. I follow Kreitman (2010) in classifying them as uvular.

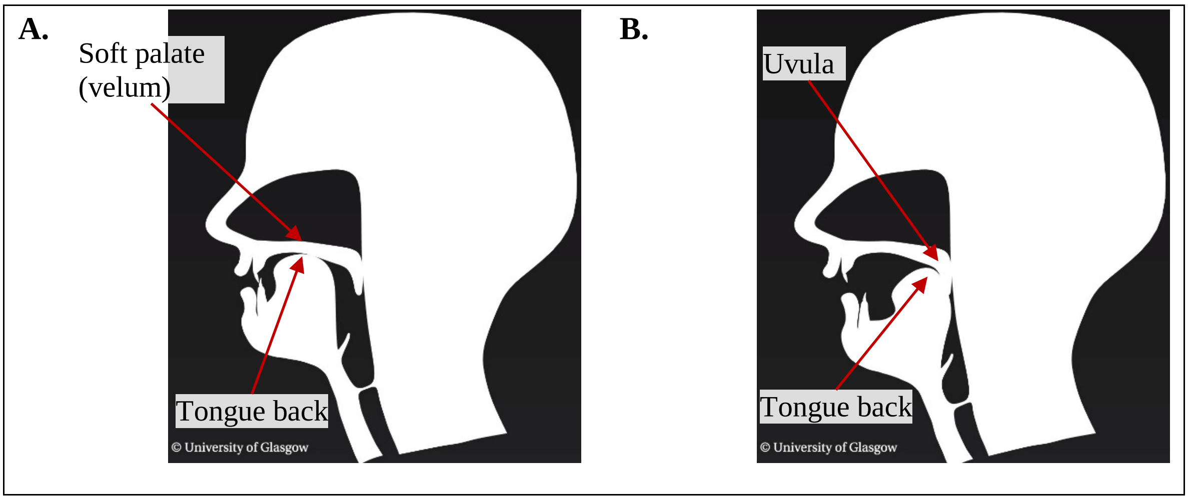


Figure 4.8 Velar (A) and Uvular (B) consonants

The next articulation site is the **pharynx** wall – the back wall of the oral cavity. Two places of articulation are associated with the pharynx. In **pharyngeal** consonants, the active articulator is the **root of the tongue** – the farthest back part of the tongue which is opposite to the back wall of the pharynx.[9] In **epiglottal** consonants, the active articulator is the **epiglottis**, a cartilage attached to the lower part of the root of the tongue.[10] The articulation of pharyngeal and epiglottal consonants involves retraction of the active articulator towards the back wall of the pharynx (see Figure 4.9). Aghul has both pharyngeal and epiglottal consonants, as exemplified in Table 4.4 and Table 4.5. Note that epiglottal consonants are listed in the OTHER SYMBOLS section of the IPA sheet, rather than in the consonant chart.

Table 4.4 Pharyngeal consonants in Aghul (Northeast Caucasian; Dagestan) (Ladefoged, 2005)

Voiced fricative [ʕ]	muʕar	'bridges'	(listen)
Voiceless fricative [ħ]	muħar	'barns'	(listen)

Table 4.5 Epiglottal consonants in Aghul (Northeast Caucasian; Dagestan) (Ladefoged, 2005)

Voiceless fricative [ʜ]	mɛʜɛr	'wheys'	(listen)
Voiceless plosive [ʡ]	sɛʡ	'measure'	(listen)

[9] The pharynx itself can constrict slightly and serve as an active articulator (see Section 6.2.4)

[10] While the epiglottis is rarely used for linguistic purposes, it has a very important biological function. During swallowing, the root of the tongue pushes the epiglottis down over the larynx, partially blocking food from entering the trachea.

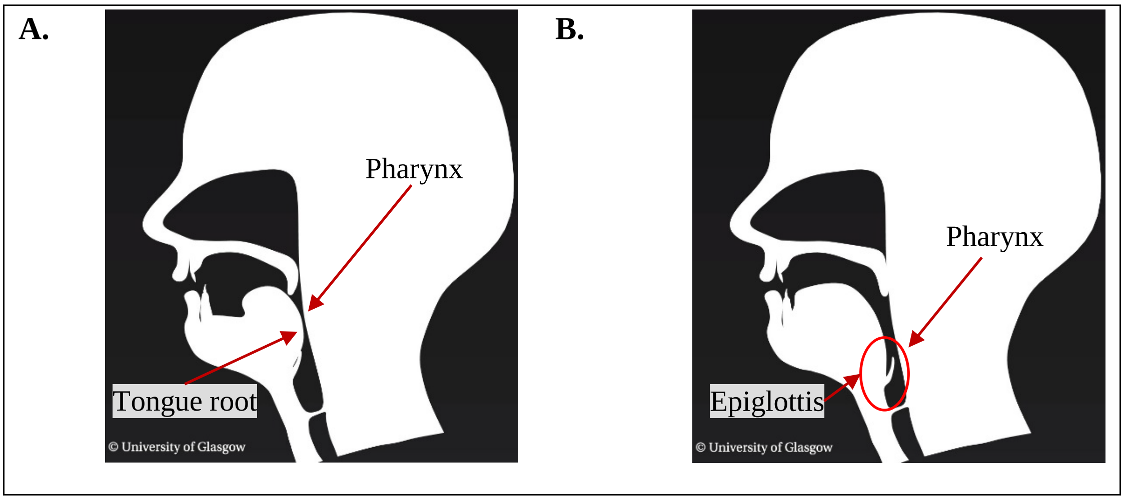


Figure 4.9 Pharyngeal (A) and Epiglottal (B) consonants

The last place of articulation is **glottal**, associated with the action of the **vocal cords**. The passive articulator in these consonants is the **glottis**, the V-shaped space between the vocal cords (since the glottis is not a physical object, it is not an articulator in the normal sense of the word). Glottal consonants are produced by constriction of the glottis, that is, by bringing the vocal cords closer together. Figure 4.10 shows two phases of glottal consonants – constricted and open. The most common glottal consonant is the voiceless glottal fricative [h], as in *head*. In certain dialects of English, ⟨t⟩ is often pronounced as a glottal stop [ʔ] (e.g., *city* [sɪʔiː]) (Trudgill, 2016). One way to emphasize the difference between the phrases *a name* and *an aim* is by inserting a glottal stop between *an* and *aim*.

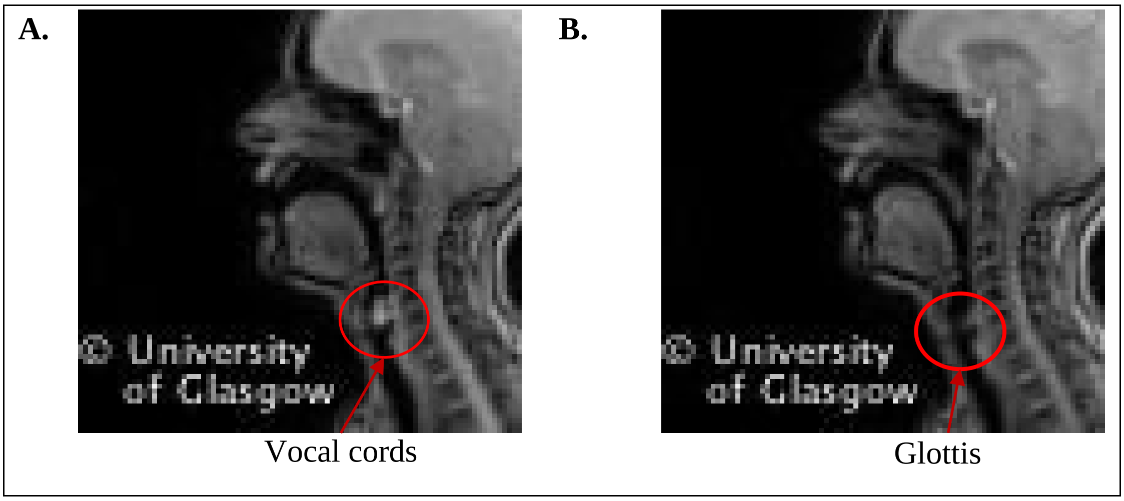


Figure 4.10 Glottal consonants. (A) Constricted glottis and (B) open glottis

Table 4.6 summarizes the various places of articulation and the articulators associated with each of them (see additional distinctions in Ladefoged & Maddieson, 1996). The table classifies the various places into several groups of major places of articulation. **Labial** consonants involve the lower lip as the active articulator. **Coronal** consonants are produced by the tip or the blade of the tongue, while **dorsal** consonants are produced with the back of the tongue. Note that palatal consonants are sometimes classified as coronal and sometimes as dorsal (Ladefoged & Johnson, 2010). **Radical** consonants involve the root of the tongue (or the epiglottis, which is attached to the root). Finally, the label **laryngeal** applies only to glottal consonants (recall that the glottis is the opening between the vocal cords, which are contained in the larynx). The main motivation underlying the classification into major place groups is phonological – consonants in the same major group tend to behave similarly across languages. This point will be demonstrated in Part IV of this book.

Table 4.6 Places of articulation

Major place group	**Place of articulation**	**Active articulator**	**Passive articulator**
Labial	Bilabial	Lower lip	Upper lip
	Labiodental	Lower lip	Upper teeth
Coronal	Interdental	Tongue blade	Teeth
	Dental	Tongue tip	Upper teeth
	Alveolar	Tongue tip/blade	Alveolar ridge
	Postalveolar	Tongue tip/blade	Postalveolar area
	Retroflex	Tongue tip	Postalveolar area / Hard palate
Coronal/dorsal	Palatal	Body (front) of tongue	Hard palate
Dorsal	Velar	Back of tongue	Soft palate (velum)
	Uvular	Back of tongue	Uvula
Radical	Pharyngeal	Root of tongue	Pharynx wall
	Epiglottal	Epiglottis	Pharynx wall
Laryngeal	Glottal	Vocal cords	Glottis

Chapter 5 Manner of articulation

This chapter reviews in detail the mechanisms underlying the various manners of articulation of consonants. Each manner is illustrated by a snapshot or a series of snapshots from animations from the *Seeing Speech* project (Lawson et al., 2018). The descriptions are accompanied by links to real-time MRI video clips (also from the *Seeing Speech* project) demonstrating the articulation of consonants by a real human. The mechanisms underlying different manners of articulation are compared to highlight similarities and differences between them.

5.1. Plosives

We begin the review with **plosive** manner of articulation, also known as **oral stop** (or simply, **stop**). In general, the articulation of plosives has three phases:

1. **Closure**: The active articulator touches the passive articulator, blocking airflow through the oral cavity.
2. While the **closure is held**, air flowing out from the lungs accumulates in the oral cavity behind the contact point of the articulators. As a result, **air pressure increases in the oral cavity**.
3. **Release**: The active articulator is pulled away abruptly from the passive articulator. The release allows the air compressed inside the oral cavity to expand and flow out. Since the release is abrupt, air flows out in an "explosive" burst.

Figure 5.1 illustrates the closure and release phases of a voiced bilabial plosive [b], produced by pressing the lower lip against the upper lip (watch).

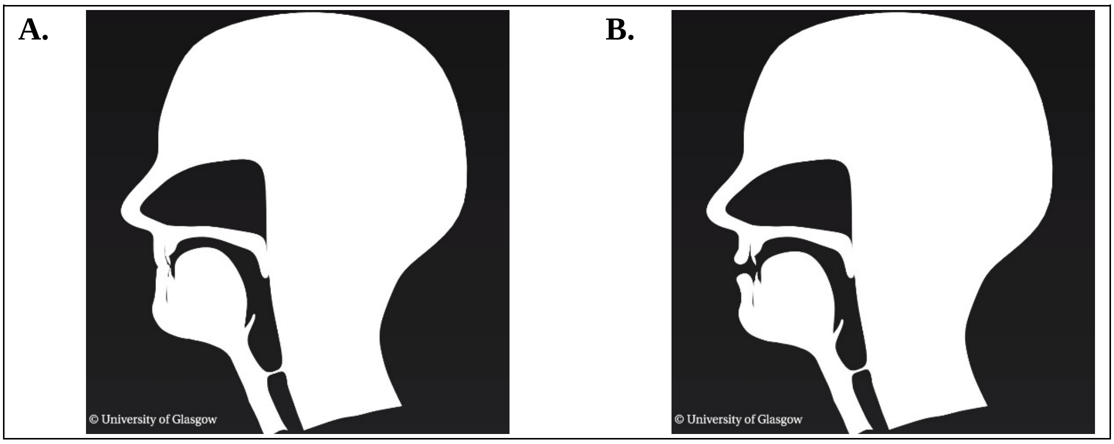


Figure 5.1 Bilabial Plosive. (A) closure (B) release.

5.2. Nasals

Similarly to plosives, **nasal** (or, **nasal stop**) consonants involve full contact between two articulators, which blocks airflow through the oral cavity. However, in nasals, the velum is lowered during closure, allowing air to flow through the nasal cavity. As a result, when the oral closure is released, there is no pressure difference between air inside and outside the oral cavity, so no burst is produced after the release. Figure 5.2 illustrates the difference in the closure phase between a bilabial oral stop [b] and a bilabial nasal stop [m] (watch). Notice the position of the velum (marked by an arrow) in the two consonants.

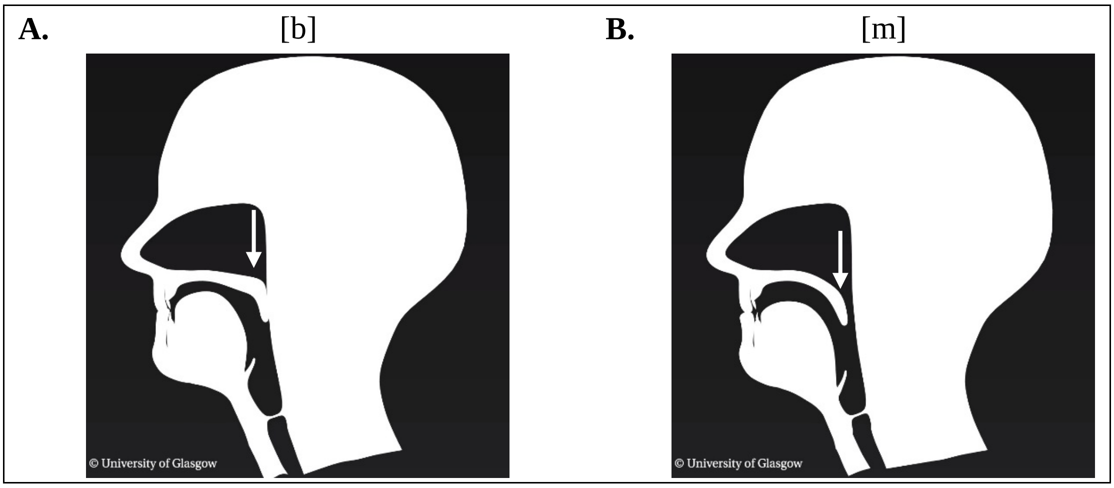


Figure 5.2 Oral vs. nasal stop. Oral closure in (A) a bilabial oral stop [b] and (B) a bilabial nasal stop [m]. The velum is marked by an arrow. Lowered velum allows air to flow through the nose in nasal consonants.

5.3. Fricatives

By contrast to plosives, **fricative** consonants do not involve contact between articulators (at least not contact that block oral air flow completely). Rather, the active articulator comes into close approximation to the passive articulator, leaving only a narrow gap in the oral cavity through which air can flow out. Forcing air to flow through this narrow gap increases the contact among air particles. Thus, when air particle flow they rub against each other, producing friction noise (hence, the term *fricative*). Figure 5.3 illustrates the difference in maximal degree of interaction between a (voiceless) velar stop [k] (watch) and a (voiceless) velar fricative [x] (watch).[11] Note that as a consequence of this difference, airflow is continuous in fricatives and discontinuous in plosives.

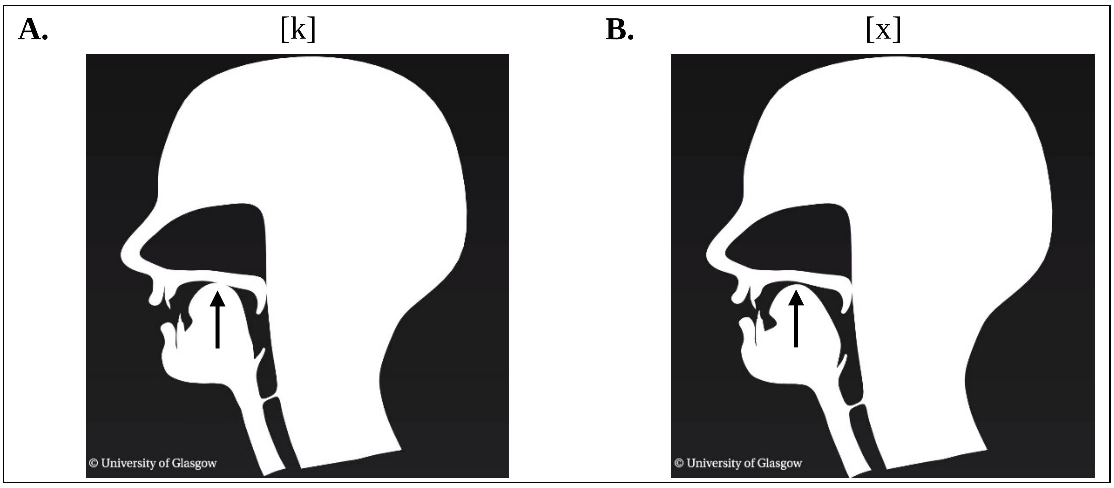


Figure 5.3 Plosive vs. fricative. The maximal degree of interaction between articulators. (A) Plosives, such as [k], involve full contact between articulators. (B) Fricatives, such as [x], involves close approximation between articulators. The arrow points to the point of contact in the plosive and to the point of maximal approximation in the fricative.

5.4. Affricates

Affricate consonants combine the closure feature of plosives and the noisy quality of fricatives. More precisely, affricates involve full closure of the oral cavity and a gradual release of the closure. The first two articulatory phases of plosives and affricates are identical: they begin when the active articulator touches the passive articulator, blocking airflow through the oral cavity. During closure, air accumulates and compresses behind the point of contact, resulting in increased air pressure in the oral cavity. The difference between plosives and affricates comes at the final phase of release.

In plosives, this phase is relatively short and abrupt resulting in an explosive burst of air. By contrast, in affricates the active articulator is pulled away slowly from the passive articulator. The air trapped inside the mouth will flow out to equalize the pressure with the surrounding environment, like in plosives. However, due to the gradual nature of the release phase in affricates, some of the air will flow out while the articulators are still in close

[11] The voiceless velar fricative is found in languages such as Russian (corresponding to the letter ⟨x⟩) and Spanish (corresponding to the letter ⟨j⟩).

approximation, thus producing a friction noise, like in fricatives. Watch a real-time MRI demonstration of an alveolar plosive [t] and an alveolar affricate [t͡s].

Since affricates combine properties of plosives and fricatives, the IPA does not use designated symbols to represent them. Instead, affricates are represented by a sequence of a plosive symbol and a fricative symbol with a **tie-bar** above or below them, e.g., [t͡s] or [t͜s]. The tie-bar symbols can be found in the OTHER SYMBOLS section of the IPA sheet.

The use of a tie-bar indicates that the sequence of symbols forms a single consonant rather than a sequence of two consonants. Note that adding the tie-bar does not necessarily have articulatory or auditory consequences ([t͡s] may sound exactly like [ts]). Rather, it indicates how native speakers of the relevant language perceive the sound(s). For example, the sequence [t]+[s] appears in English words like *cats* [kæts]. However, native speakers of English always perceive this sequence as two separate consonants. Thus, using a tie-bar with [ts] would be inappropriate in English (e.g., *[kæt͡s]). By contrast, in Modern Hebrew, [t]+[s] sequences can be either a single consonant (e.g., *צב* [t͡sav] 'tortoise') or a sequence of two consonants (e.g., *תשומת לב* [tsumɛt lɛv] 'attention', rather than *[t͡sumɛt lɛv]).[12]

Note that affricates are always represented by a plosive and a fricative in the same place of articulation or in adjacent places of articulation that involve the same active articulator.[13] When the place of articulation of the plosive and fricative parts are not identical, according to the IPA, the place of articulation of the affricate is determined by the fricative part. Thus, [t͡ʃ] (the sound of ⟨ch⟩ in *check*) and [d͡ʒ] (the sound of ⟨j⟩ in *jacket*) are postalveolar affricates. This is because the point of contact tends to be closer to the place of articulation of the fricative than to the place of articulation of the plosive (if you pronounce [t] and then [t͡ʃ], you might notice that the place of contact between the tongue and upper surface of the mouth shifts a bit to the back in [t͡ʃ]). In order to transcribe such affricates more precisely, one can use a diacritic to modify the place of articulation of the plosive (e.g., attach the "Retracted" diacritic to [t] in order to represent a voiceless postalveolar affricate [t̠͡ʃ]).

The following scheme compares the articulation of plosives and affricates:

---

[12] Although there is no auditory difference, we can determine that [ts] is not a single consonant in [tsumɛt lɛv] since [t] is a prefix and is not pronounced in other morphologically related phrases, such as *לשים לב* [lasim lɛv] 'to pay attention' (*[lat͡sim lɛv]).

[13] Possible non-identical place combinations include: bilabial + labiodental, alveolar + (inter)dental, alveolar + postalveolar, velar + uvular.

Plosive	**Affricate**
Closure in the oral cavity	Closure in the oral cavity
↓	↓
Air pressure increases	Air pressure increases
↓	↓
Sudden release: explosion	**Gradual** release: friction

5.5. Taps and flaps

While affricates differ from plosives in the duration of release, **tap** and **flap** consonants differ from plosives in the duration of closure. Like plosives, they involve contact between articulators, but the contact in taps and flaps is very brief. As a result, less pressure is accumulated during the closure, and the release is less dramatic.

The IPA does not distinguish taps from flaps, possibly because no language contrasts taps and flaps at the same place of articulation (Hall, 1997: 106). However, their articulatory mechanism is slightly different. A tap involves a rapid "forward-backward" movement of the active articulator (typically the tip of the tongue) – it moves directly towards the passive articulator, strikes it, and moves back to its rest position. In flap consonants, the active articulator is withdrawn from its rest position and then moves back rapidly, in a slightly different trajectory, such that it strikes the passive articulator on the way back to its rest position. Watch a real-time MRI demonstration of an alveolar tap/flap [ɾ], which is a common phonetic realization of ⟨t⟩ between vowels in American English (e.g., [wɑːɾɚ] 'water') and of ⟨r⟩ between vowels in Spanish (e.g., *caro* [kaɾo] 'expensive').[14]

5.6. Trills

In **trill** consonants, the active articulator vibrates near the passive articulator, causing the air between the articulators to vibrate, as well. The produced sound is similar to that of multiple taps, but it is generated by airstream rather through contact between the articulators. ⟨rr⟩ in Standard Spanish (as in *perro* [pero] 'dog') is an alveolar trill [r] (watch). There are also bilabial trills produced by vibrations of the lower lip (watch). The Kele language spoken in the easterly section of inland Manus Island, New Guinea has both alveolar and bilabial trills.

[14] Ladefoged & Maddieson (1996: 230-232) discuss the differences between the American alveolar flap and the Spanish dental tap.

Table 5.1 Trills in Kele (Austronesian; New Guinea) (Ladefoged, 2005)[15]

Bilabial [ʙ]	^{m}ʙulim	‘face’	(listen)
Alveolar [r]	^{n}ruwin	‘bone’	(listen)

5.7. Approximants

Another manner of articulation that involves no contact between articulators is **approximant**. In these consonants, the active articulator only slightly approaches the passive articulator, leaving enough space for the air to flow relatively freely. Approximants can be viewed as another point on the continuum of maximal degree of interaction between the articulators: moving from full contact in plosives (and others) to close approximation in fricatives and on to slight approximation in approximant. Figure 5.4 illustrates the difference in maximal degree of interaction between a (voiced) retroflex fricative [ʐ] and a (voiced) retroflex approximant [ɻ] (watch). [ɻ] is a common pronunciation of ⟨r⟩ in some varieties of American English (Collins & Mees, 2003). [ʐ] is found, for example in Russian (e.g., *ЖУК* [ʐuk] ‘beetle’) (Hamann, 2004).

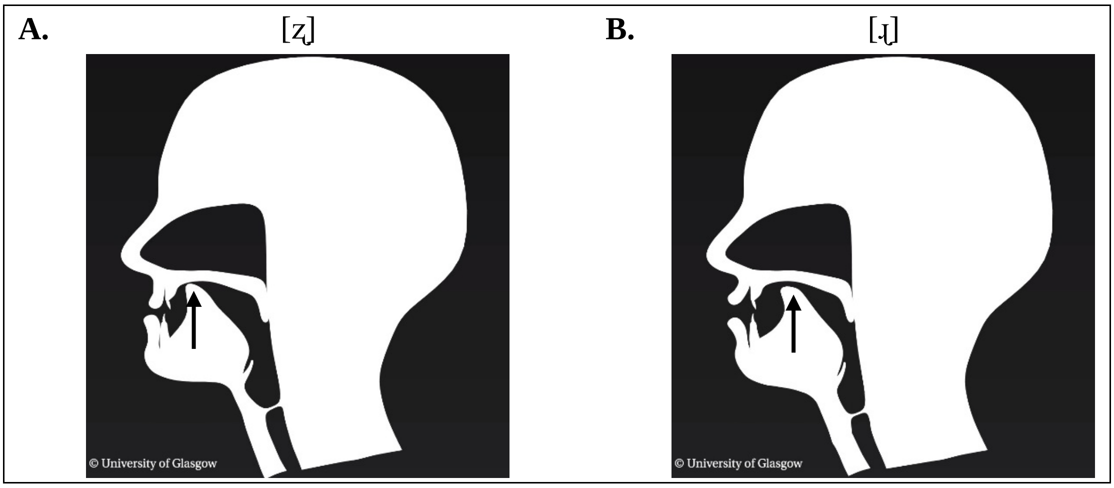


Figure 5.4 Fricative vs. approximant. The maximal degree of interaction between articulators. (A) Fricatives, such as [ʐ], involve close approximation between articulators causing air to flow through the oral cavity with friction. (B) Approximants, such as [ɻ], involve slight approximation between articulators allowing air to flow through the oral cavity relatively freely. The arrows point to the points of maximal approximation.

[15] [^{m}ʙ] and [^{n}r] are **prenasalized** consonants – consonants with an initial brief period of air flow through the nasal cavity.

5.8. Lateral consonants

The last two manners of articulation involve simultaneous full contact in the oral cavity and oral air flow. In these **lateral** consonants, part of the tongue makes contact with the upper surface of the mouth along its median line, thus blocking airflow through the medial part of the oral cavity. However, the tongue is configured in such a way that air can flow along the sides of the tongue. **Lateral approximants**, like [l] (e.g., in *luck*) are the most common lateral consonants. These consonants involve free air flow along the sides of the tongue. In **lateral fricatives**, the oral tract along the sides of the tongue is narrowed, causing the air to flow with friction. Lateral fricatives are found, for example in Zulu:

Table 5.2 Lateral fricatives in Zulu (Bantu; South Africa) (Ladefoged, 2005)

Voiced [ɮ]	ɮálà	'play (imp.)'	(listen)
Voiceless [ɬ]	ɬânzà	'vomit'	(listen)

5.9. Additional classes

In addition to the formal classes of manner of articulation, there several additional terms that intersect with some of these classes. These additional classes group consonants that are either a sub-class of one of the major manner classes, or have different manners, yet they seem to have similar properties (e.g., in terms of perception or phonological behavior).

In the context of fricatives and affricates it is worth mentioning the sub-class of **strident** (or, **sibilant**) consonants. This sub-class includes alveolar and postalveolar fricatives and affricates, such as: [s], [ʃ], [z], [ʒ], [t͡s], [t͡ʃ], and [d͡ʒ]. These are rather "loud" and noticeable consonants. The voiceless stridents are especially interesting. They are used cross-linguistically and cross-culturally to calm down and get the attention (e.g., shushing) of other people (see Chapter 11). In addition, voiceless stridents have a unique phonological behavior: sequences of a voiceless strident plus another consonant (e.g., [sport]) are found in many languages that do not allow other sequences of fricative + consonant (e.g., *[fport], *[hport]).

The approximants [w] (as in English *win*) and [j] (the sound of ⟨y⟩ in *yes*) are also commonly known as **glides** (or, **semivowels**). Articulatorily and acoustically they are similar to the vowels [u] and [i], respectively, but they function as consonants.

The term **rhotic** refers to "r-like" consonants, including the alveolar trill [r], the alveolar tap [ɾ], the retroflex approximant [ɻ], the uvular fricative [ʁ] (typical pronunciation of ⟨r⟩ in

French), and more. Finally, the term **liquid** is a conventional name for the combined group of lateral approximants ('l-sounds') and rhotics ('r-sounds').

5.10. Summary

The following figures summarize the contrasts among several manners of articulation. Figure 5.5 compares various manners of articulation by the degree of obstruction in the oral tract. Figure 5.6 and Figure 5.7 compare various manners of articulation with completer closure according to the duration of closure and release, respectively.

Plosive, Affricate, Nasal	High	
Fricative	↑	Lateral fricative
Approximant	Low	Lateral approximant

Figure 5.5 Manners of articulation by degree of obstruction in oral tract

Plosive	Long
	↑
Tap or Flap	Short

Figure 5.6 Manners of articulation by duration of closure

Affricate	Long
	↑
Plosive	Short

Figure 5.7 Manners of articulation by duration of release

5.11. Practice

Describing the articulation of consonants based on their articulatory features (place, manner, and voice) is an important skill for any phonetician. In this section, I demonstrate how to write a schematic description of the articulation of various consonants. In general, the practice proceeds along the following lines:

(i) Find the consonant's symbol in the phonetic chart.

(ii) Identify the consonant's place of articulation (column in IPA chart) and determine the active and passive articulators (see Table 4.6).

(iii) Identify the consonant's manner of articulation (row in IPA chart).

(iv) Describe the articulation according to the manner of articulation (use the schemes outlined in this chapter). The description should refer to the actions of the active articulator with respect to the passive articulator, and indicate how the airflow is affected by the interaction of the articulators.

(v) Identify the consonant's voice (IPA cells: left – voiceless, right – voiced) and describe the state of the vocal cords (vibrating/not vibrating).

It is also useful to practice the naming convention of consonants: [voice] [place of articulation] [manner of articulation].

5.11.1. Plosives

Plosives have three phases: (i) closure – the active articulator touches the passive articulator blocking airflow through the oral cavity; (ii) air pressure increases in the oral cavity behind the point of contact; (iii) release – the active articulator is pulled away abruptly causing the air inside the oral cavity to flow out in an "explosive" burst.

Example – [k]:

- Name: voiceless velar plosive.
- Articulators: passive: velum, active: tongue back.
- Description: The back of the tongue touches the velum blocking airflow through the oral cavity. Air pressure increases in the oral cavity behind the point of contact. The back of the tongue is then lowered abruptly, causing the air inside the oral cavity to flow out in an "explosive" burst. The vocal cords don't vibrate.

5.11.2. **Nasals**

In nasals, the active articulator touches the passive articulator blocking airflow through the oral cavity. At the same time, the velum is lowered, allowing air to flow through the nasal cavity.

Example – [ɱ]:

- Name: voiced labiodental nasal.
- Articulators: passive: upper teeth, active: lower lip.
- Description: the lower lip touches the upper teeth blocking airflow through the oral cavity. At the same time, the velum is lowered, allowing air to flow through the nasal cavity. The vocal cords vibrate.

5.11.3. **Fricatives**

In fricatives the active articulator comes into close approximation to the passive articulator, forming a narrow gap between the articulators. Air flows through the gap with friction.

Example – [β]:

- Name: voiced bilabial fricative.
- Articulators: passive: upper lip, active: lower lip.
- Description: The lower lip comes into close approximation to the upper lip, forming a narrow gap. Air flows through the gap with friction. The vocal cords vibrate.

5.11.4. **Affricates**

Affricates have three phases: (i) closure – the active articulator touches the passive articulator blocking airflow through the oral cavity; (ii) air pressure increases in the oral cavity behind the point of contact; (iii) release – the active articulator is pulled away slowly causing the air inside the oral cavity to flow out with friction.

Example – [d͡ʒ]:

- Name: voiced postalveolar affricate.
- Articulators: passive: postalveolar area, active: tongue tip/blade.
- Description: the tongue tip/blade touches the postalveolar area blocking airflow through the oral cavity. Air pressure increases in the oral cavity. The tongue tip/blade is then pulled away slowly, causing the air inside the oral cavity to flow out with friction. The vocal cords vibrate.

5.11.5. **Taps and flaps**

In a **tap**, the active articulator moves directly towards the passive articulator, strikes it, and moves back to its rest position. In **flap** consonants, the active articulator is withdrawn from its rest position and then moves back rapidly, in a slightly different trajectory, such that it strikes the passive articulator on the way back to its rest position.

Example – [ɾ]:

- Name: voiced alveolar tap/flap.
- Articulators: passive: alveolar ridge, active: tongue tip/blade.
- Description: **Tap**: the tongue tip/blade briefly touches the alveolar ridge and is withdrawn back along the same path. **Flap**: the tongue tip/blade is drawn from neutral position and strikes the alveolar ridge on its way back (along a different path). The vocal cords vibrate.

5.11.6. **Trills**

In trill consonants, the active articulator vibrates near the passive articulator, causing the air between the articulators to vibrate, as well.

Example – [ʙ]:

- Name: voiced bilabial trill.
- Articulators: passive: upper lip, active: lower lip.
- Description: The lower lip vibrates near the upper lip, causing the air between the lips to vibrate, as well. The vocal cords vibrate.

5.11.7. **Approximants**

In approximants , the active articulator slightly approaches the passive articulator, allowing air to flow relatively freely between the articulators.

Example – [j]:

- Name: voiced palatal approximant.
- Articulators: passive: hard palate, active: tongue body.
- Description: The body of the tongue approaches the hard palate, allowing air to flow relatively freely between them. The vocal cords vibrate.

5.11.8. **Lateral approximants**

In lateral approximants, the active articulator touches the passive articulator, blocking airflow through the medial part of the oral cavity. The air flows freely along the sides of the tongue.

Example – [l]:

- Name: voiced alveolar lateral approximant.
- Articulators: passive: alveolar ridge, active: tongue tip/blade
- Description: The tongue tip/blade touches the alveolar ridge, blocking airflow through the medial part of the oral cavity. The air flows freely along the sides of the tongue. The vocal cords vibrate.

5.11.9. **Lateral fricatives**

In lateral fricatives, the active articulator touches the passive articulator, blocking airflow through the medial part of the oral cavity. In addition, the oral tract is narrowed along the sides of the tongue causing the air to flow with friction.

Example – [ɮ]:

- Name: voiced alveolar lateral fricative.
- Articulators: passive: alveolar ridge, active: tongue tip/blade
- Description: The tongue tip/blade touches the alveolar ridge, blocking airflow through the medial part of the oral cavity. The oral tract is narrowed along the sides of the tongue causing the air to flow with friction. The vocal cords vibrate.

Chapter 6 Complex consonants

The previous chapters described the articulation of consonants that have a single place of articulation, that is, consonants that involve interference to the airflow in a single point along the vocal tract. This chapter introduces two additional types of consonants that involve simultaneous interference to airflow in two places.

6.1. Double articulation

Doubly-articulated consonants are produced by simultaneous interference to airflow of the same manner in two different places. Some doubly-articulated consonants have designated IPA symbols, which can be found in the "OTHER SYMBOLS" section of the IPA sheet. Most famous of these is the voiced labial-velar approximant [w] (watch), as in *win*. As can be seen in Figure 6.1A, [w] involves approximation of the articulators in two places: the lips are rounded and the back of the tongue approaches the velum.[16] Compare this to an approximant with a single articulation at the lips and no significant raising of the tongue (Figure 6.1B).

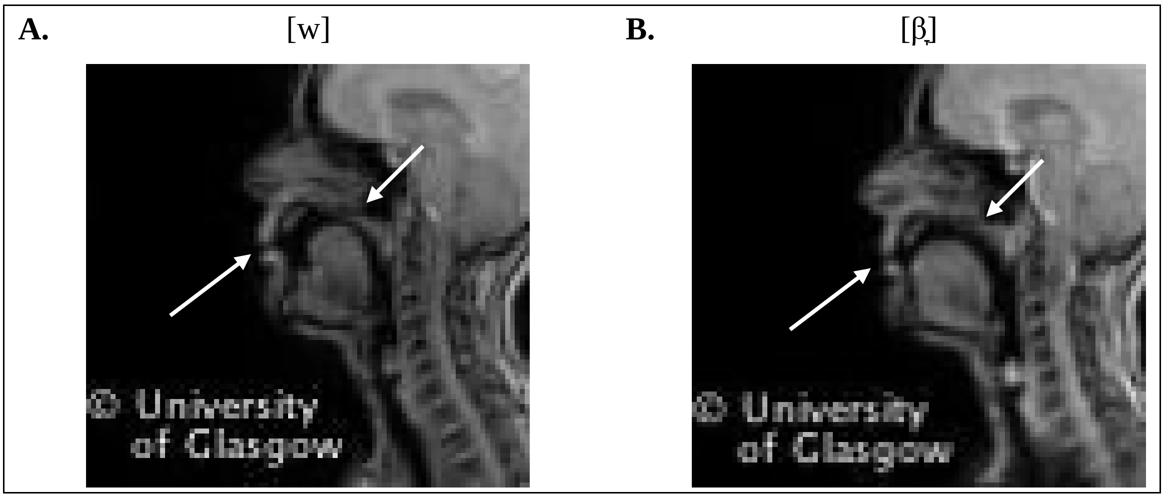


Figure 6.1 Double vs. single articulation. (A) The voiced labial-velar approximant [w] involves slight approximation between the lips and also between the back of the tongue and the velum. (B) The voiced bilabial approximant [β̞] involves slight approximation between the lips and no significant raising of the tongue.[17] The lowering sign ˕ is used in combination with the voiced bilabial fricative [β] to represent

[16] Ladefoged & Maddieson (1996: 328) claim that doubly-articulated approximants are not parallel to segments with two closures.

[17] The snapshot made for the bilabial approximant was taken from a video of a bilabial fricative due to the absence of a corresponding video for a bilabial approximant. The degree of approximation between the articulators matches the expected degree for an approximant.

a bilabial approximant which has no designated symbol. The arrows point to the points of maximal approximation.

When no designated symbol exists, it is possible to represent doubly-articulated consonants by a sequence of two consonants under a tie-bar. The most common cases are labial-velar plosives, which involve simultaneous closure at the lips and the velum. Figure 6.2 compares a voiceless labial-velar plosive [k͡p] (watch) with a voiceless bilabial plosive [p]. Both involve contact between the lips, but [k͡p] also involves simultaneous contact between the back of the tongue and the velum that is also released at the same time as the labial closure. Table 6.1 provides examples for labial-velar plosives in the Nigerian language Idoma. To a non-native speaker of Idoma these consonants sound very similar to bilabial plosives, but they have some special quality due to the additional closure at the velum.

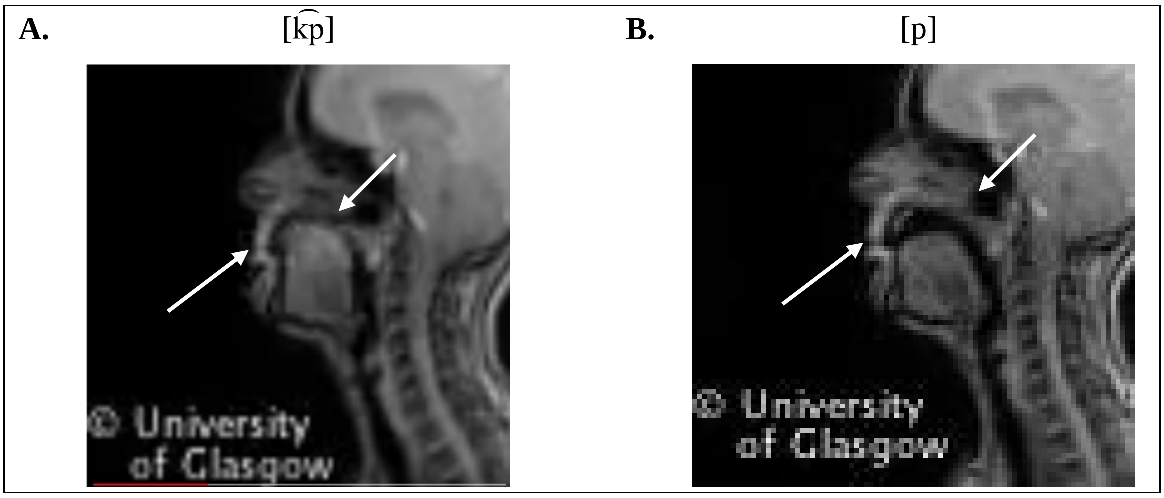


Figure 6.2 Double vs. single articulation. (A) The voiceless labial-velar plosive [k͡p] involves contact between the lips and between the back of the tongue and the velum. (B) The voiceless bilabial plosive [p] involves contact between the lips and no significant raising of the tongue.

Table 6.1 Doubly articulated labial-velar plosives in Idoma (Niger-Congo; Nigeria) (Ladefoged, 2005)

Voiced [g͡b]	ag͡ba	'jaw'	(listen)
Voiceless [k͡p]	ak͡pa	'bridge'	(listen)

As we can see, the tie-bar symbol is used with both affricates and doubly articulated consonants. The phonological interpretation of the tie-bar is the same in both types of consonants. The use of the tie-bar indicates that the combined symbols form a single consonant in the language. However, the phonetic interpretation of the tie-bar depends on the type of consonant. For affricates, the tie-bar does not necessarily indicate any change in the articulation relative to a sequence of two consonants. However, doubly articulated consonants under a tie-bar are produced (and perceived) very differently from a sequence of the two consonants. The role of the tie-bar in affricates and doubly articulated consonants is summarized in Table 6.2.

Table 6.2 Tie-bar symbols in affricates and doubly-articulated consonants

Type	Example	Phonetics: with/without the tie-bar
Affricate	d͡ʒ vs. dʒ	Sound the same
Double articulation	k͡p vs. kp	Sound very differently

6.2. Secondary articulation

Some consonants involve simultaneous interference to airflow of different manners in two different places. In these cases, the higher degree of constriction is referred to as **primary articulation**, and the lower degree of constriction is called **secondary articulation**. According to Ladefoged & Maddieson (1996), secondary articulation is always approximant, and the primary articulation can be plosive, nasal, liquid, or fricative. In IPA, the primary articulation is denoted by a regular consonant symbol, and the secondary articulation is denoted by a superscript. The following sub-section reviews various types of secondary articulation.

6.2.1. Labialization

Labialized consonants are (usually non-labial) consonants produced with extra lip rounding. Labialization is indicated in IPA with a superscript w following a consonant symbol that marks the primary articulation (e.g., a labialized voiceless velar plosive [kʷ]). Labialized velar plosives are probably the most common consonants of this type. Table 6.3 provides examples for labialized velar plosives in the Nigerian language Bura. According to the common reconstruction of Proto-Indo-European, this proto language also had labialized velar plosives, for example, in words that are spelled with ⟨qu⟩ (e.g., *quiet*) in Modern English and Romance languages.

Table 6.3 Plain and labialized velar plosives in Bura (Chadic; Nigeria) (Ladefoged, 2005)

	Plain			Labialized		
Voiced	gàm	‘ram’	(listen)	gʷàr	‘poison’	(listen)
Voiceless	kálá	‘to bite’	(listen)	kʷárá	‘donkey’	(listen)

6.2.2. **Palatalization**

Palatalized consonants are non-palatal consonants (i.e., the primary place of articulation is not at the hard palate) produced with simultaneous approximation of the tongue body to the hard palate. Palatalization is indicated in IPA with a superscript j following a consonant symbol that marks the primary articulation (e.g., a palatalized voiced alveolar nasal [nʲ]). Russian is especially known for its palatalized consonants that contrast with plain consonants. See some examples in Table 6.4.

Table 6.4 Plain and palatalized consonants in Russian (Ladefoged, 2005)

Plain			**Palatalized**		
lof	'catch'	(listen)	lʲɛf	'lion'	(listen)
nos	'nose'	(listen)	nʲos	'he carried'	(listen)
sok	'juice'	(listen)	sʲok	'he lashed'	(listen)

6.2.3. **Velarization**

Velarized consonants are non-velar consonants produced with simultaneous approximation of the tongue back to the velum. Velarization is marked in IPA with a superscript gamma following a consonant symbol that marks the primary articulation. Perhaps the most common type of a velarized consonant is the velarized voiced alveolar lateral-approximant [lˠ], often called "dark-L". It is more commonly transcribed with a superimposed tilde sign [ɫ]. This is the typical pronunciation of English ⟨l⟩ in syllable-final position (in some variants of English it is the default pronunciation of ⟨l⟩). Compare, for example the pronunciation of ⟨l⟩ in *lips* [lɪps] and ⟨ll⟩ in spill [spɪɫ].

6.2.4. **Pharyngealization**

Pharyngealized consonants are non-pharyngeal consonants produced with a slight constriction of the pharynx. Pharyngealization is marked in IPA with a superscript reversed glottal stop following a consonant symbol that marks the primary articulation (e.g., a pharyngealized voiceless alveolar plosive [tˤ]). Arabic is famous for its pharyngealized coronal ("emphatic") consonants that contrast with plain coronal consonants. See some examples in Table 6.5.

Table 6.5 Plain and pharyngealized consonants in Arabic (Ladefoged, 2005)

Plain			Pharyngealized		
tiːn	‘figs’	(listen)	tˤiːn	‘mud’	(listen)
suːs	‘licorice’	(listen)	sˤuːsˤ	‘chick’	(listen)
dalːat	‘she pointed’	(listen)	dˤalˤːat	‘she stayed’	(listen)

6.3. Practice

As in the case of “simple” consonants, describing the articulation of complex consonants can help understanding their nature. When describing consonants with second articulation, we begin by describing the primary articulation as usual, and adding a description for the secondary articulation. It is important to emphasize the simultaneity of the articulations.

Describing the articulation of complex consonants proceeds along similar lines to those described in Section 5.11. In particular, the scheme for describing doubly-articulated consonants:

(i) Find both symbols in the phonetic chart.

(ii) Identify the consonants’ places of articulation (IPA columns) and determine the pairs of active and passive articulators (see Table 4.6).

(iii) Identify the consonants’ manner of articulation (row in IPA chart).

(iv) Describe the articulation according to the manner of articulation (use the schemes outlined in this chapter). Include the relevant articulators in the description. Important! The description should refer to both pairs of articulators and indicate the correct order of actions.

(v) Identify the consonants’ voice (IPA cells: left – voiceless, right – voiced) and describe the state of the vocal cords (vibrating/not vibrating).

The scheme for describing the articulation of consonants with secondary articulation:

(i) Find the primary symbol in the IPA CONSONANTS chart and the secondary symbol in the DIACRITICS chart.

(ii) Identify the consonant’s primary place of articulation (column in the CONSONANTS chart) and secondary place of articulation (according to diacritic name) and determine pairs of active and passive articulators (see Table 4.6).

(iii) Identify the consonants’ manner of articulation (according to primary consonant).

(iv) Describe the articulation according to the manner of the primary consonant (use the schemes outlined in this chapter) and add a description of an approximant for the secondary articulation. Include the relevant articulators in the description. Important! The description should refer to both pairs of articulators and indicate the correct order of actions.

(v) Identify the voice of the primary consonant (IPA cells: left – voiceless, right – voiced) and describe the state of the vocal cords (vibrating/not vibrating).

It is also useful to practice the naming convention of consonants:

- **Double articulation**: [voice] [front-back place of articulation] [manner of articulation].
- **Secondary articulation**: [secondary articulation] [voice] [place of articulation] [manner of articulation].

<u>Exercise</u>

Name and describe the articulation of the following consonants:

1. [k͡p]
2. [ɡʷ]
3. [tʲ]
4. [lˠ]
5. [dˤ]

<u>Solution</u>

1. [k͡p]:

- **Name**: Voiceless labial-velar plosive.
- **Place of articulation**: labial-velar.
- **Articulators**: <u>front</u>: passive: upper lip, active: lower lip; <u>back</u>: passive: velum, active: tongue back.
- **Manner of airflow**: plosive.
- **State of vocal cords**: voiceless.
- **Description**: the lower lip touches the upper lip. At the same time, the tongue back touches the velum. Airflow through the oral cavity is blocked, and air pressure increases. Then, the lower lip and tongue back are lowered abruptly at the same time, causing the air inside the oral cavity to flow out with an explosive burst. The vocal cords don’t vibrate.

2. [g^w]:

- **Name**: Labialized voiced velar plosive.
- **Place of articulation**: primary: velar, secondary: bilabial.
- **Articulators**: primary: passive: velum, active: back of tongue; secondary: passive: upper lip, active: lower lip.
- **Manner of airflow**: plosive.
- **State of vocal cords**: voiced.
- **Description**: The back of the tongue touches the velum blocking airflow through the oral cavity. At the same time, the lower lip approaches the upper lip. Air pressure increases in the oral cavity. Then, the back of the tongue is lowered abruptly, and the air flows out in an explosive burst. The vocal cords vibrate.

3. [t^j]:

- **Name**: Palatalized voiceless alveolar plosive.
- **Place of articulation**: primary: alveolar, secondary: palatal.
- **Articulators**: primary: passive: alveolar ridge, active: tongue tip/blade; secondary: passive: hard palate, active: tongue body.
- **Manner of airflow**: plosive.
- **State of vocal cords**: voiceless.
- **Description**: The tip/blade of the tongue touches the alveolar ridge blocking airflow through the oral cavity. At the same time, the tongue body slightly approaches the hard palate. Air pressure increases in the oral cavity behind the point of contact. Then, the tip/blade of the tongue is lowered abruptly, causing the air inside the oral cavity to flow out in an "explosive" burst. The vocal cords do not vibrate.

4. [$l^ɣ$]:

- **Name**: Velarized voiced alveolar lateral approximant.
- **Place of articulation**: primary: alveolar, secondary: velar.
- **Articulators**: primary: passive: alveolar ridge, active: tip/blade of tongue; secondary: passive: velum, active: tongue back.
- **Manner of airflow**: lateral approximant.
- **State of vocal cords**: voiced.

- **Description**: The tongue tip/blade touches the alveolar ridge, blocking airflow through the medial part of the oral cavity. At the same time, the back of the tongue approaches the velum. The air flows freely along the sides of the tongue. The vocal cords vibrate.

5. [dˤ]:

- **Name**: Pharyngealized voiced alveolar plosive.
- **Place of articulation**: primary: alveolar, secondary: pharyngeal.
- **Articulators**: primary: passive: alveolar ridge, active: tip/blade of tongue; secondary: passive: pharynx wall, active: pharynx wall.
- **Manner of airflow**: plosive.
- **State of vocal cords**: voiced.
- **Description**: The tip/blade of the tongue touches the alveolar ridge blocking airflow through the oral cavity. At the same time, the pharynx walls constrict. Air pressure increases in the oral cavity. Then, the tip/blade of the tongue is lowered abruptly, causing the air to flow out in an explosive burst. The vocal cords vibrate.

Chapter 7 Vowels

This chapter describes the articulation of vowels. A **vowel** is a segment whose articulation involves no significant obstruction of the airstream (Trask, 1996). Figure 7.1 is an elaborated version of the IPA vowel chart. The vowel chart organizes vowels in three dimensions according to their articulatory features: the vertical (height) and horizontal (backness) axes reflect the position of the tongue during articulation, while the left-right order within vowel pairs reflects the configuration of the lips.

Reading vowels off the chart is done according to the following naming convention: [Height] [Backness] [Roundedness]. Thus, [u] is a close back rounded vowel, and [a] is an open front unrounded vowel. Names of levels indicated in parentheses (e.g., near-back) are not official names (they are not indicated in the official chart).

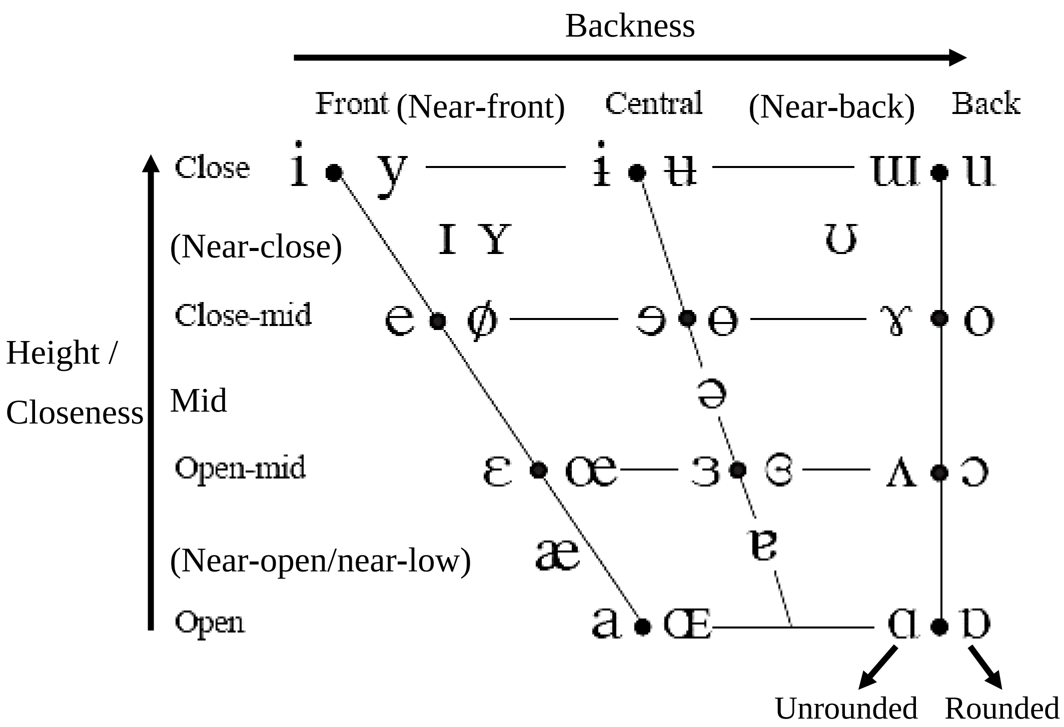


Figure 7.1 The IPA vowel chart (elaborated)

7.1. Primary vowel features

This section describes the primary features of vowels that can be read directly off the vowel chart.

7.1.1. Height ("closeness")

Close, or **high vowels** (e.g., [i]) are produced with the tongue raised relative to its neutral vertical position. Since the tongue is attached to the lower jaw, raising the tongue entails a relatively closed mouth. By contrast, **low**, or **open vowels** (e.g., [a]) are produced with lowered tongue and open mouth. Finally, in **mid vowels** (e.g., [ə]) the tongue is in its neutral vertical position. Figure 7.2 illustrates the difference between a high vowel [i] and a low vowel [a]. Note the vertical position of the tongue in these vowels.

A. [i] **B.** [a]

Figure 7.2 Vowel height. (A) a close/high (front, unrounded) vowel [i]. (B) an open/low (front, unrounded) vowel [a]. In high vowels, the tongue is closer to the roof of the mouth than in low vowels.

Vowel height can be used as a distinctive feature, namely to distinguish vowels that are identical in all other respects. For example, in English, the difference between the vowels of *hid* [h**ɪ**d], *head* [h**ɛ**d], and *had* [h**æ**d] can be said to be a difference in height (all of them are unrounded front vowels). Danish has unrounded front vowels in four height levels:

Table 7.1 Height contrasts in unrounded front vowels in Danish (Ladefoged, 2005)

v**i**:ðə	‘white.DEF’	(listen)	↑ Vowel height
v**e**:ðə	‘wheat’	(listen)	
v**ɛ**:ðə	‘wet’	(listen)	
v**æ**:ðə	‘wade’	(listen)	

7.1.2. **Backness**

Back vowels (e.g., [u]) are articulated with the highest point of the tongue at the back of the mouth, while in **front vowels** (e.g., [y]) the highest point of the tongue is towards the front of the mouth. In **central vowels** (e.g., [ʉ]), the tongue is in a horizontal neutral position.[18] Figure 7.3 illustrates the difference between a front vowel [y] and a back vowel [u]. Note the horizontal position of the tongue in these vowels.

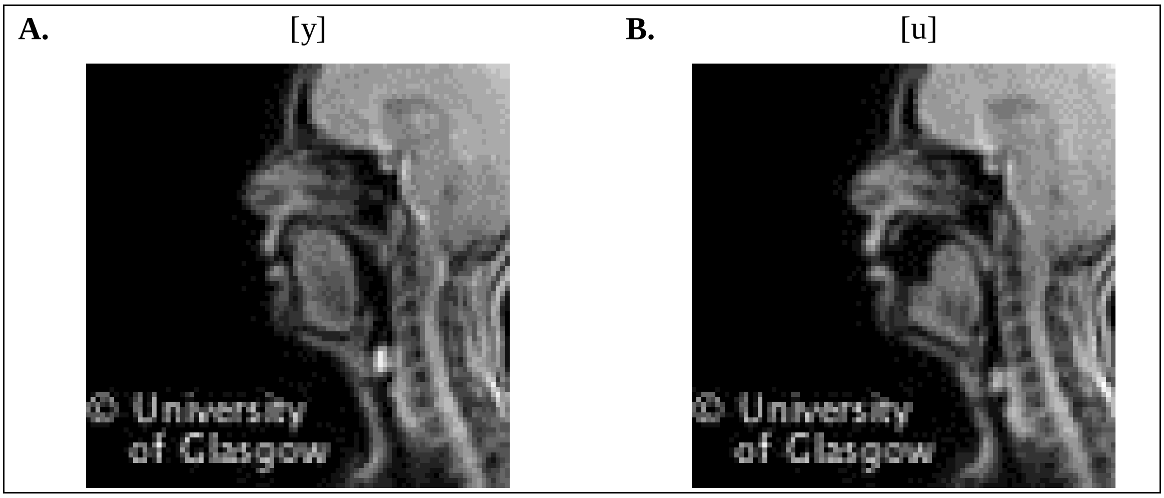


Figure 7.3 Vowel backness. (A) a front (close/high, rounded) vowel [y]. (B) a back (close/high, rounded) vowel [u]. In front vowels, the tongue is stretched forward, while in back vowels, the tongue is retracted.

Vowel backness can also be used as a distinctive feature. For example, the difference between the vowels of *bet* [bɛt] and *but* [bʌt] can be said to be a difference in backness (both are open-mid unrounded vowels). Norwegian has rounded high vowels in three backness levels:

[18] Note that the absolute horizontal position of vowels depends on their height. This is evident from the trapezoid-like arrangement of the vowel chart, which reflects the shape of the vocal tract. Due to the anatomy of the tongue, high front vowels are, in fact, “fronter” than low front vowels.

Table 7.2 Backness contrasts in rounded high vowels in Norwegian (Ladefoged, 2005)

byː 'city' (listen)	bʉː 'shack' (listen)	buː 'live' (listen)
Vowel backness →		

7.1.3. **Roundedness**

Rounded vowels (e.g., [ø]) are articulated with rounded lips, while **unrounded vowels** (e.g., [e]) are articulated with spread lips. In the IPA vowel chart, rounded is indicated by the arrangement of vowels in pairs: the right vowel in each pair is rounded, while the left vowel is unrounded. There are four unpaired vowels on the chart. All of them are unrounded except for [ʊ]. Figure 7.4 illustrates the difference between a rounded vowel [ø] and an unrounded vowel [e]. Note the configuration of the lips in these vowels.

A. [e] **B.** [ø]

Figure 7.4 Vowel roundedness. (A) an unrounded (close-mid, front) vowel [e]. (B) a rounded (close-mid, front) vowel [ø].

English typically does not distinguish vowels based on roundedness alone. French, on the other hand, have several pairs of front vowels that are contrasted by roundedness.

Table 7.3 Unrounded and rounded front vowels in French (Ladefoged, 2005)

	Unrounded				**Rounded**			
Close	*lit*	[li]	'bed'	(listen)	*lu*	[ly]	'read.PAST'	(listen)
Close-mid	*les*	[le]	'the.PL'	(listen)	*le*	[lø]	'the.M.SG'	(listen)

7.2. Additional vowel features

In addition to three main vowel features, languages use several other features to contrast vowels. In IPA these features are indicated by diacritics attached to the vowel letter.

7.2.1. Length

Some languages contrast vowel based on their length (or, duration). Long vowels are indicated by the Length mark 'ː' (e.g., [uː]), which can be found in the SUPRASEGMENTALS section of the IPA sheet (the length mark is not a colon, it is composed of two triangles pointing at each other).[19] Finnish contrasts long and short vowels (and also long and short consonants) as can be seen in Table 7.4.

Table 7.4 Vowel length contrasts in Finnish (Ladefoged, 2005)

Short			Long		
tulen	'I come'	(listen)	tuːlen	'of wind'	(listen)
tullen	'I may come'	(listen)	tulleːn	'of the one who has come'	(listen)

7.2.2. Nasalization

In vowels, air typically flows exclusively through the oral cavity. However, in some vowels, air flows simultaneously through the oral and nasal cavity. We call the former, plain type – **oral vowels**, and the latter – **nasal** (or, **nasalized**) **vowels**. Nasalized vowels are indicated by the superscript tilde (e.g., [ũ]), which can be found in the DIACRITICS section of the IPA sheet.

Frenchcontrasts plain and nasalized vowels as can be seen in Table 7.5. Historically, the nasalized vowels emerged through the interaction of a plain vowel and a nasal consonant. This is still evident in the written form of words containing nasal vowels (e.g., *sans* [sɑ̃] 'without'). Importantly, the nasal consonant is not pronounced in many of these words and, therefore, there is a genuine contrast between oral and nasalized vowels in the spoken language. In other languages, vowels may also acquire a nasal quality in the environment of nasal consonants (e.g., the vowels in *money* may be nasalized for some speakers of English), but not all the languages use nasalization for contrasting meaning.

[19] The length mark is also used to describe long consonants (also called **geminates**), but consonant length is more commonly indicated by repeated consonant letters (e.g., [ll]) rather than by the length mark (e.g., [lː]).

Table 7.5 Nasalized vowels in French (Ladefoged, 2005)

	Oral				Nasal			
Close-mid	*lot*	lo	‘prize’	(listen)	*long*	lõ	‘long’	(listen)
Open	*là*	la	‘there’	(listen)	*lent*	lã	‘slow’	(listen)

As explained in Chapter 5, airflow in nasal consonants is generated by lowering the velum while the oral cavity is blocked. Lowered velum is also responsible for the nasal airflow in nasalized vowels. The difference, however, is that the oral cavity remains open in nasalized vowels, so air flows through both the oral and nasal cavities.

7.2.3. **Rhoticity**

Another feature that vowels can obtain from the interaction of consonants is **rhoticity**. This is common in some forms of English where vowels become **rhotacized**, or “r-colored”, in the presence of a rhotic consonant ([ɻ] or [ɹ]), as in *bird*. In fact, the rhotic consonant is often not pronounced when the vowel is rhotacized (in the case of *bird*, the vowel is both rhotacized and lengthened, i.e., [bɝːd]). Rhotacized vowels are indicated by a right hook diacritic (e.g., [a˞]), which can be found in the DIACRITICS section of the IPA sheet.

The articulation of rhotacized vowels typically involves raising or curling of the tongue, but their common attribute is in their acoustic and perceptual signature, rather than their articulation. Rhoticity as a distinctive feature is extremely rare. The Dravidian language Badaga have (or used to have) a three-way contrast between plain, half rhotacized and fully rhotacized vowels:

Table 7.6 Rhotacized vowels in Badaga (Dravidian; India) (Ladefoged, 2005)

Plain	Half rhotacized	Fully rhotacized
beː ‘mouth’ (listen)	be˞ː ‘bangle’ (listen)	be˞˞ː ‘crop’ (listen)

7.2.4. **Advanced tongue root (ATR)**

The **root of the tongue** is the part of the tongue which lies furthest back and lowest. It can be manipulated somewhat independently than the rest of the tongue. Some languages use the position of the tongue root as another degree of freedom in vowels. In **advanced tongue root vowels ([+ATR])** , the tongue root is stretched forward, increasing the volume of the pharynx. In **retracted tongue root (RTR) vowels ([-ATR])**, the tongue root is retracted, decreasing the volume of the pharynx. ATR vowels are marked with the ‘Advancing sign’ diacritic (e.g., [e̘]), while RTR vowels are marked with the ‘Retracting sign’ diacritic (e.g., [e̙]).

English has an ATR distinction in two pairs of vowels: front: [iː]-[ɪ], back: [uː]-[ʊ]. The contrast is demonstrated in Table 7.7.[20] While the vowels in each pair are said to contrast in ATR, note that the [+ATR] vowels are always long and the [-ATR] vowels are always short. Also, as suggested by Figure 7.5, ATR in English is correlated with tongue height. By contrast, Ladefoged & Maddieson (1996: 304) claim that, in Akan, these dimensions are independent. The difference between the languages is reflected in the transcription: in Akan, ATR contrast is expressed with diacritics (see Table 7.8), while in English it is expressed with different vowel letters.

Table 7.7 ATR contrasts in English

	Front vowels		Back vowels	
[+ATR]	ʃiːp	‘sheep’	kuːd	‘cooed’
[-ATR]	ʃɪp	‘ship’	kʊd	‘could’

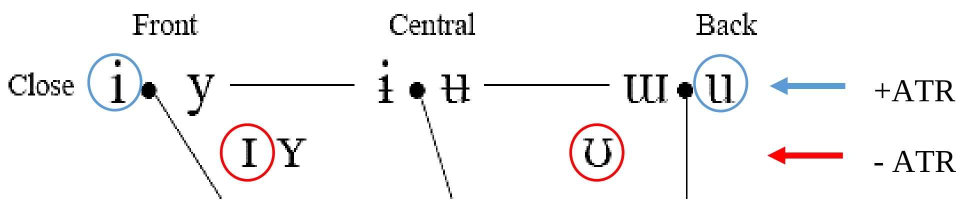


Figure 7.5 ATR contrast in English vowels

Table 7.8 ATR contrasts in Akan (Niger-Congo; Ghana) (Ladefoged, 2005)

	Front vowels			Back vowels		
[+ATR]	si̘	‘wash’	(listen)	bu̘	‘break’	(listen)
[-ATR]	si̙	‘say’	(listen)	bu̙	‘get drunk’	(listen)

7.3. Monophthongs vs. diphthongs

All the vowels we have discussed so far are **monophthongs** – vowels with a single quality. However, many languages also have **diphthongs** – vowels which display a smooth change from one quality to another (within the same syllable). The most common diphthongs begin

[20] The [+ATR] vowels are sometimes called **tense vowels**, and the [-ATR] vowels are sometimes called **lax vowels**, but the ATR distinction is preferred. The tense/lax distinction refers to muscular tension during articulation.

with a lower vowel and end with a higher vowel. Some examples from English are given in Table 7.9. The shifts in vowel quality in these diphthongs are illustrated in Figure 7.6. The [ɪ] and [ʊ] in these diphthongs resemble the **glides** [j] and [w], respectively. Indeed, diphthongs are often transcribed as a vowel-glide sequence (e.g., [aj] and [aw]) rather than a vowel-vowel sequence.

Table 7.9 Diphthongs in English

1.	[aɪ]	aɪs	‘ice’
2.	[eɪ]	beɪk	‘bake’
3.	[aʊ]	aʊt	‘out’

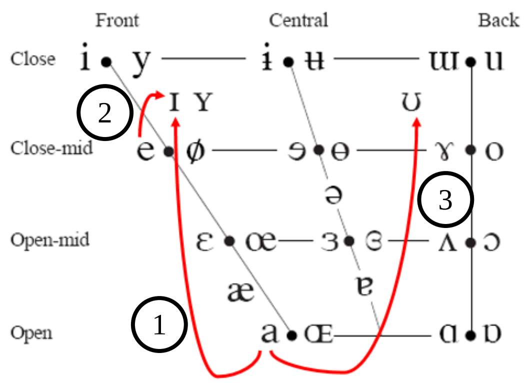


Figure 7.6 vowel quality shifts in diphthongs in English. The numbers correspond to the examples in Table 7.9. The arrow heads indicate the direction of quality shift (e.g., diphthong 1 begins with the quality of the vowel [a] and ends with the quality of [ɪ]).

Despite some similarity, a diphthong is a single vowel (with two qualities), not a sequence of two vowels! As in the case of affricates (see Chapter 5), the difference between a diphthong and a sequence of two vowels is often a matter of representation rather than a matter of articulation, and this is a language-dependent issue. For example, consider the English word *bite* and the Hebrew word *בית* ‘house’. These ae similarly sounding words, which can be roughly transcribed [bait]. However, native speakers of English consider the sequence [ai] to be a single vowel (and *bite* to be a single syllable), while Hebrew speakers consider that same sequence as two vowels (and the word *בית* to consist of two syllables).

Based on the phonetic transcription alone, it is not possible to know, without knowing the particular language, whether a sequence of two vowel letters constitute a diphthong or two independent vowels. If syllable boundaries are marked (see Chapter 18), a sequence of two independent vowels is separated into two syllables, while a diphthong will be contained in a single syllable. Another useful (albeit uncommon) way to disambiguate diphthongs is to use a tie-bar (e.g., [a͡ɪ]) to indicate the sequence as a single vowel.

7.4. Vowels vs. consonants

Generally, vowels and consonants are considered separate classes of segments. This is supported by many types of phonological evidence (e.g., vowels and consonants participate in different phonological processes). However, vowels and consonants can be seen as different groups on one segmental continuum. Accordingly, we can characterize vowels in terms of features we normally use to describe consonants. Under this view, vowels can be regarded as an “extreme” case of approximants in terms of manner of articulation. Also, like approximants, vowels are voiced by default. Finally, although vowels are typically not analyzed according to place of articulation features, some theories propose that vowels do have a place of articulation (e.g., Rice, 1995).

7.5. The cardinal vowels

Being a type of approximant sound, it is very difficult to describe the articulation of vowels accurately. Moreover, the exact pronunciation of vowels varies considerably even across dialects of the same language. Various attempts have been made to establish a conventional system of vowels that have an exact articulatory description which, in turn, can be serve as reference vowels when describing the vowels of various languages and dialects. The most successful reference vowel system, called ***cardinal vowels*** was proposed by the British phonetician Daniel Jones (Jones, 1956). His system is based on 8 primary cardinal vowels and 10 secondary cardinal vowels.

Figure 7.7 illustrates the tongue positions in the 8 primary cardinal vowels. The pillars of the system are cardinal vowel 1 [i], which combines the greatest possible degree of ‘closeness’ and the greatest possible degree of ‘frontness’, and cardinal vowel 5 [ɑ], which combines the greatest possible degree of ‘openness’ and the greatest possible degree of ‘backness’. The idea behind the definition of these two points was to define the anatomical range of what could be considered a vowel. Placing the tongue beyond these positions is either impossible (e.g., the tongue cannot be lowered below point 5) or would result in a consonant rather than a vowel (e.g., raising the tongue farther in cardinal vowel 1 would result in a fricative).

The cardinal vowels 4 ([a]; open front) and 8 ([u]; close back) constitute the other two corners of the diagram. Since this system of contrasts is based on the position of the tongue the diagram has the shape of a right trapezoid since, for example, the horizontal position of the tongue in close front vowels ([i]) is "fronter" than in open front vowels ([a]). Defining the other cardinal vowels is somewhat more subjective and is based on perceived contrasts. For example, to obtain cardinal vowel 2, one starts with cardinal vowel 1 and then lowers one's tongue until the perceived quality of the vowel changes. Other cardinal vowels can be obtained by similarly changing the position of the tongue in a gradual manner until the perceived quality of the vowel changes.[21]

The cardinal vowel system is based primarily on the position of the tongue. The configuration of the lips is modified independently. Cardinal vowels 1-5 are unrounded, while cardinal vowels 6-8 are rounded. This choice follows the distributions of vowels in the languages of the world, where unrounded front vowels are more common than rounded front vowels and rounded back vowels are more common than unrounded back vowels. The only exception is with respect to open back vowels, which are typically unrounded.

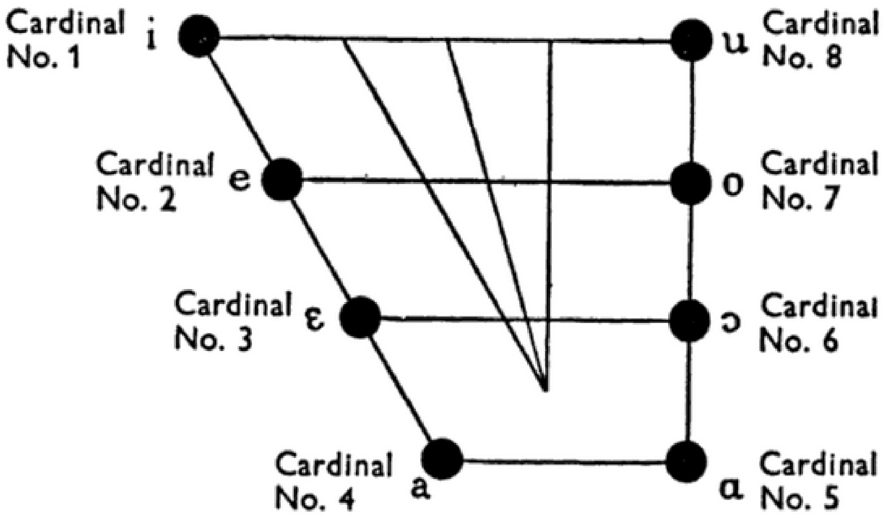


Figure 7.7 A diagram illustrating the tongue positions of the cardinal vowels (Jones, 1956)

The secondary cardinal vowels are defined on the basis of the primary cardinal vowels and typically involve lip rounding. Thus, cardinal vowels 9-13 are the rounded counterparts of cardinal vowels 1-5 (in the same order) and cardinal vowels 14-16 are the unrounded counterparts of cardinal vowels 6-8. Cardinal vowels 17 and 18 are unrounded ([ɨ]) and rounded ([ʉ]) close central vowels, respectively.

[21] In other words, an accurate articulatory definition of the cardinal vowels is possible only if we use the cardinal vowels determined by a specific speaker as reference for all other speakers. In reality, the vowel symbols are used in a rather broad and ambiguous manner.

Chapter 8 Airstream mechanisms

In previous chapters, we described the articulation of consonants listed in the CONSONANTS (PULMONIC) section of the IPA. The term *pulmonic* means ‘produced by the lungs’ (from Latin *pulmo* ‘lung’). That is, the articulation of **pulmonic consonants** uses air that is pushed from the lungs out. The articulatory mechanism that uses air from the lungs is called **egressive pulmonic airstream** (egressive = going or directed outward). In this chapter, we will describe another group of consonants – the **non-pulmonic consonants**. These consonants are listed in the CONSONANTS (NON-PULMONIC) section of the IPA (Table 8.1). As the name suggests, non-pulmonic consonants do not used air that is pushed from the lungs. Instead, they use other airstream mechanisms.

Table 8.1 IPA non-pulmonic consonants

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
ʘ Bilabial	ɓ Bilabial	ʼ Examples:
ǀ Dental	ɗ Dental/alveolar	pʼ Bilabial
ǃ (Post)alveolar	ʄ Palatal	tʼ Dental/alveolar
ǂ Palatoalveolar	ɠ Velar	kʼ Velar
ǁ Alveolar lateral	ʛ Uvular	sʼ Alveolar fricative

8.1. Ejectives

Ejectives are consonants produced by pushing trapped air out via an upward movement of the larynx (glottis). This type of articulatory mechanism is called **egressive glottalic airstream**. Ejectives are found in 18% of the languages of the world (Ladefoged & Maddieson, 1996). They are indicated in IPA by a modifier letter apostrophe following a consonant letter (e.g., [k’] is a velar ejective).

The articulation of ejectives begins with a double closure: complete obstruction at the glottis (glottal closure) and at another place in the oral cavity (oral closure). This is followed by a rapid raising of the larynx, which compresses the air that is trapped between the closures (air pressure is up to twice higher than in pulmonic stops). Then, the oral closure is released, like in pulmonic stops release, only louder. Finally, the glottal closure is released. This sequence of gestures is illustrated in Figure 8.1 for a bilabial ejective [p’] (watch). Figure 8.2 further illustrates the differences between an oral stop and an ejective: while both types of consonants have an oral closure, the ejective also have glottal action which begins with the oral closure and ends after the oral release.

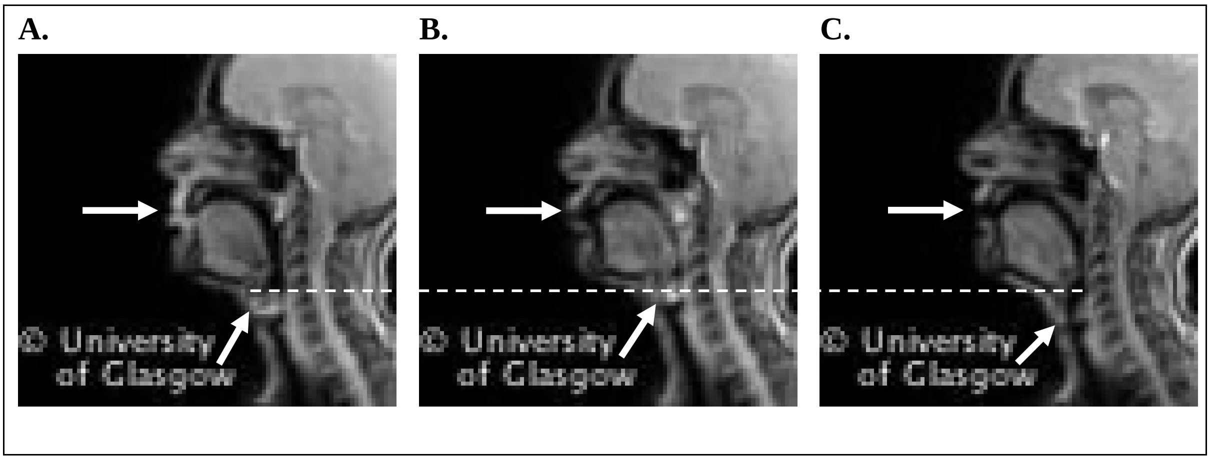


Figure 8.1 Ejectives: from closure to release. The figure illustrates the articulation of a bilabial ejective [p’]: (A) Double closure at the lips and glottis. (B) Laryngeal movement + oral release. (C) Glottal release.

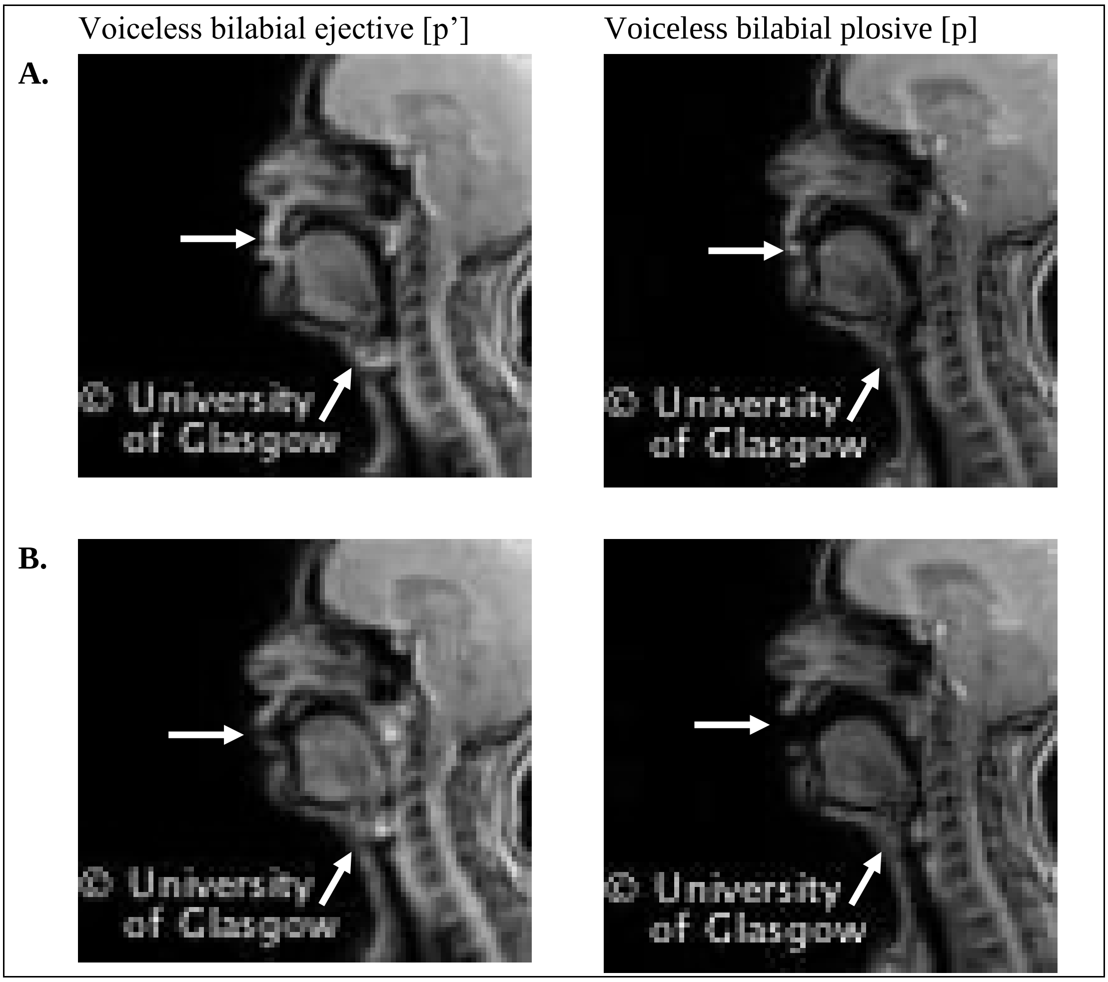


Figure 8.2 Ejectives vs. pulmonic stops. The figure compares a bilabial ejective [p’] and a bilabial stop [p]: (A) closure: both types of consonants have an oral closure (at the lips), but the ejective also have a glottal closure. (B) When the oral closure is released the glottis remains closed in ejectives.

Ejectives are usually stops, but there are also fricative and affricate ejectives. They are always voiceless and can be articulated in various places of articulation, though the most common ejectives are velar. Table 8.2 lists examples for ejectives in Lakhota, a language spoken by the Lakota people in North Dakota and South Dakota.

Table 8.2 Ejectives in Lakhota (Siouan; United States) (Ladefoged, 2005)

Bilabial	p’o	‘foggy’	(listen)
Dental	t̪’uʃe	‘at all costs	(listen)
Velar	k’u	‘to give’	(listen)

8.2. Implosives

Implosives are non-pulmonic consonants which, like ejective, involve glottalic activity, but in the opposite direction. That is, implosives are produced by a downward movement of the larynx (glottis). This type of mechanism is called **ingressive glottalic airstream** (ingressive = going or directed inward). Implosives exist in 10% of the languages of the world. They are represented in IPA by modified consonant letters that have a characteristic “hooktop”.

The articulation of implosives begins with complete obstruction at some place in the oral cavity. Then, the larynx is lowered, which decompresses the air trapped in the oral cavity (i.e., air pressure decreases). Finally, the closure is released. If the larynx is lowered rapidly enough, air flows into the mouth after the release. This sequence of gestures is illustrated in Figure 8.3 for a bilabial implosive [ɓ] (watch).

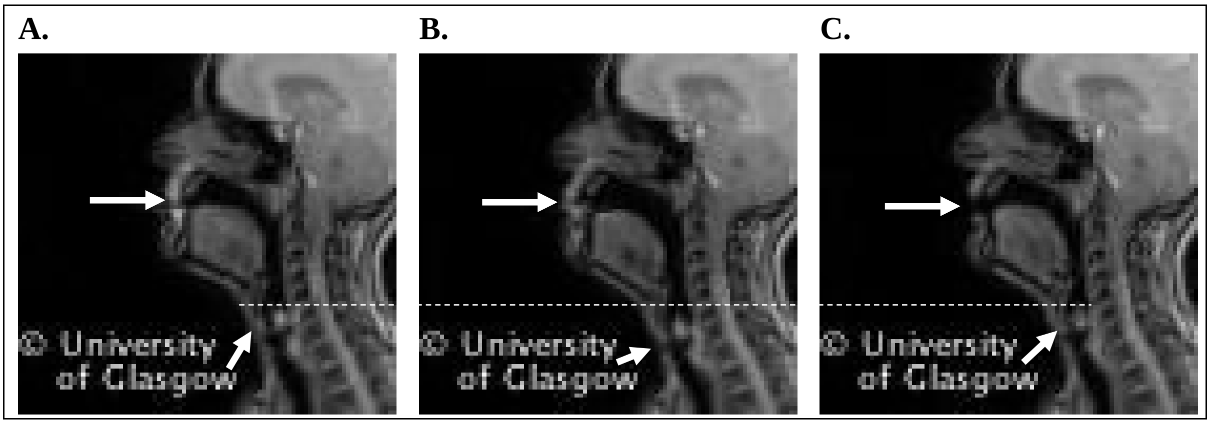


Figure 8.3 Implosives: from closure to release. The figure illustrates the articulation of a bilabial implosive [ɓ]: (A) oral closure (at the lips). (B) Laryngeal movement (lowering) (C) Oral release.

Implosives are always stops and are usually voiced (i.e., the vocal cords vibrate while the larynx is lowered). They are possible in various places of articulation, but the most common implosives are bilabial. Table 8.3 lists some examples for implosives in Sindhi. They are contrasted with plosives, though the contrast is not always clearly audible.

Table 8.3 Implosives in Sindhi (Indo-European; Pakistan) (Ladefoged, 2005)

	Plosive			Implosive		
Bilabial	banu	‘forest’	(listen)	ɓani	‘field’	(listen)
Postalveolar	ɖoːru	‘you run’	(listen)	ᶑinu	‘festival’	(listen)
Velar	guɳu	‘quality’	(listen)	ɠanu	‘handle’	(listen)

The main articulatory differences between egressive (ejectives) and ingressive (implosives) airstreams can be demonstrated using a needle-free syringe, where the syringe body corresponds to the vocal tract and the piston mimics the larynx. To demonstrate egressive airstream, start with retracted piston. Then, seal the syringe opening – this mimics the oral closure. While maintaining the closure, push the piston to compress the air trapped in the syringe. This is equivalent to raising the larynx. Then, remove obstruction from the syringe opening. These steps are demonstrated in Figure 8.4. If done properly, you will hear a tiny burst of air following the release.

To demonstrate ingressive airstream, start with a fully closed piston and a sealed syringe opening. While maintaining the closure, pull piston to decompress the air trapped in the syringe. This is equivalent to lowering the larynx. Finally, remove the obstruction from the syringe opening. These steps are demonstrated in Figure 8.5. If done properly, you will hear a tiny burst of air following the release.

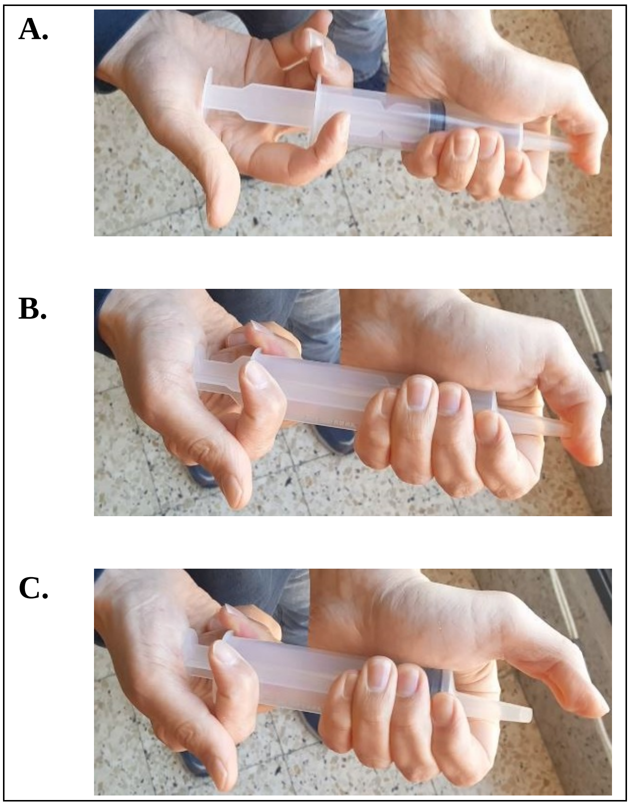


Figure 8.4 Egressive airstream (ejective). (A) frontal closure. (B) compressing trapped air (larynx raising). (C) release of frontal closure.

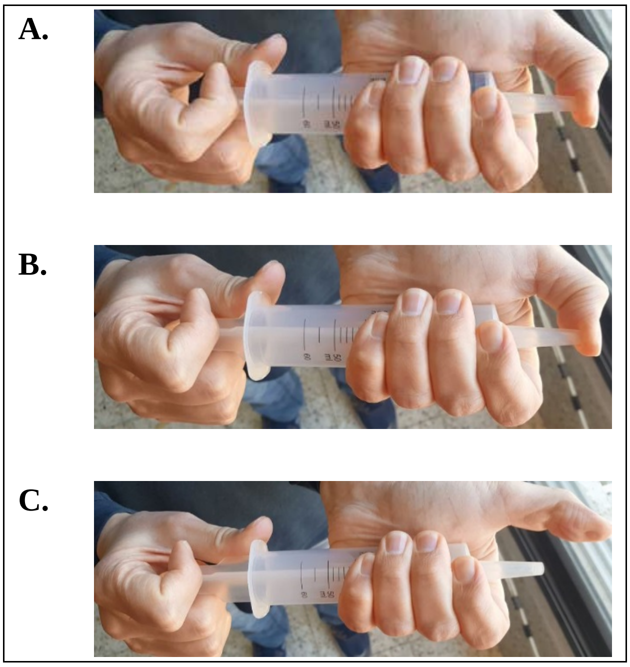


Figure 8.5 Ingressive airstream (implosive). (A) frontal closure. (B) decompressing trapped air (larynx lowering). (C) release of frontal closure.

8.3. Clicks

The last class of non-pulmonic consonants is the class of clicks. Speakers of many languages use clicks in communication (e.g., the *tsk! tsk!* gesture of disapproval used in English is a dental click). However, clicks as linguistic sounds are found only in languages spoken in Southern African. Nevertheless, they are quite common in those languages. In the !Xóõ language, over 70% of the words begin with a click (Ladefoged & Maddieson, 1996).[22]

Clicks are produced by **ingressive velaric airstream**, which involves decompressing air between two closures in the oral cavity. Articulation of clicks begins with a double oral closure: complete obstruction at a front location plus a rear location in the oral cavity. Next, the body of the tongue is lowered, so the air trapped in the oral cavity is decompressed (air pressure decreases). Then, the front closure is released, producing a loud click, as air flows into the mouth. Finally, the rear closure is released. Figure 8.6 demonstrates this articulatory sequence for bilabial, dental, and alveolar clicks.

Each of the click closures has its own place and manner of articulation. The frontal closure can be anywhere from the lips to the hard palate, and its manner is either stop or affricate. The

[22] Click here to hear a story in !Xóõ (the transcription is found here). The story is told first in English and then in !Xóõ. The endonym of the language (the name of the language used by the speakers of the language) is repeated several times in the recording from 0’16” to 0’24”.

rear closure is at the velum or the uvula, and its manner is either stop or nasal (that is, the velum is lowered in nasal clicks, and remains raised in stop clicks). Clicks can be voiced or voiceless.

Clicks have designated IPA letters (see Figure 8.6), which describe the frontal closure of the click. However, a full transcription of clicks also indicates the rear closure. Table 8.4 lists symbols used to represent the rear and front closure of clicks. Conventionally, the transcription of clicks is a concatenation of a rear closure symbol plus a frontal closure symbol, as demonstrated in Table 8.5. Table 8.6 lists some examples for clicks in !Xóõ language.

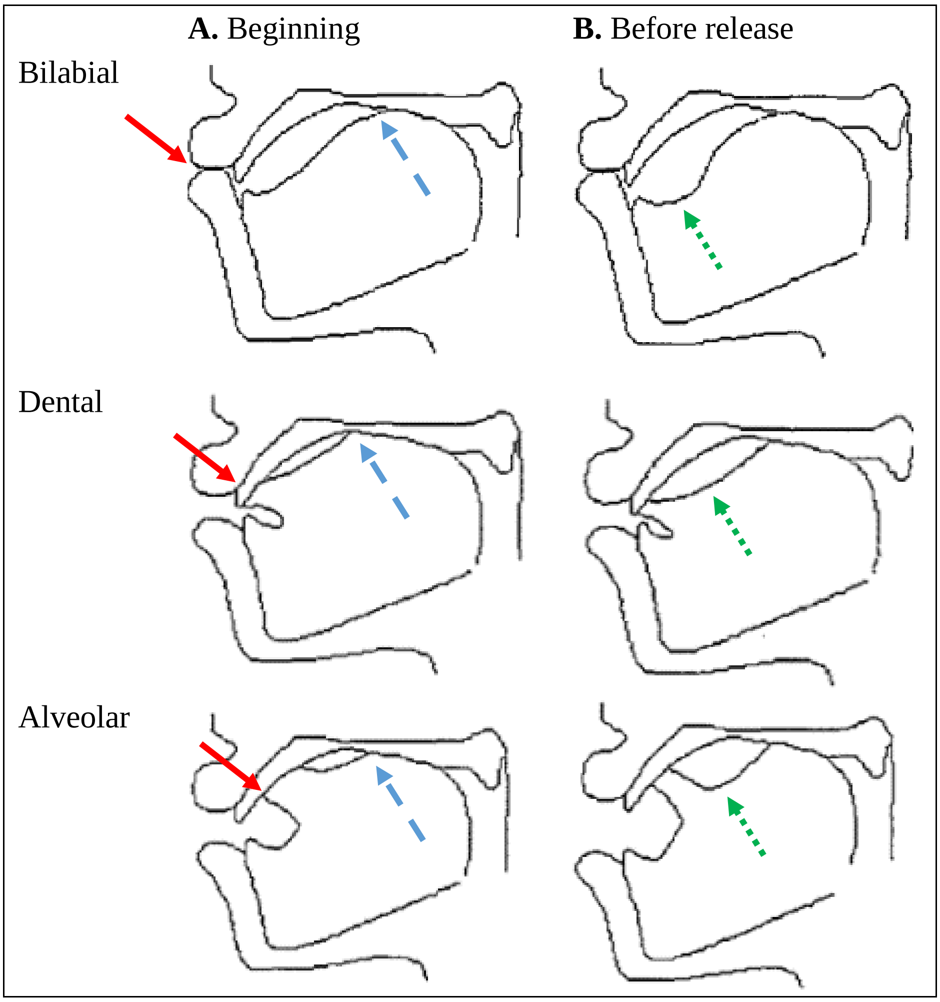


Figure 8.6 Clicks. The figure illustrates the articulation of clicks at three places of articulation. (A) double clousre. (B) lowered body of tongue. (adapted from Ladefoged & Maddieson, 1996)

Front closure   Rear closure   Tongue body lowered

Table 8.4 IPA symbols for clicks

Rear closure		Front closure	
Type	**Symbol**	**Type**	**Symbol**
Velar voiceless stop	k	Bilabial	ʘ
Velar voiced stop	ɡ	Dental	ǀ
Velar nasal	ŋ	(Post)alveolar	!
Uvular voiceless stop	q	Palatoalveolar	ǂ
Uvular voiced stop	ɢ	Alveolar lateral	ǁ
Uvular nasal	ɴ		

Table 8.5 Examples for transcription of clicks

Description	**IPA**
Voiceless bilabial velar click	kʘ
Voiced dental velar click	ɡǀ
Voiced alveolar velar nasal click	ŋ!
Voiced palatoalveolar uvular click	ɢǂ

Table 8.6 Clicks in !Xóõ (Khoisan; Botswana) (Ladefoged, 2005)

Bilabial	kʘʔôo	‘get stuck’	(listen)
Dental	kǀʔâɑ	‘die’	(listen)
Alveolar	k!ʔáã	‘be seated.PL’	(listen)
Palatal	kǂʔāa	‘shoot you’	(listen)
Alveolar lateral	kǁʔàa	‘not to be’	(listen)

8.4. Summary

Type	**Mechanism**	**Front closure**	**Rear closure**
Pulmonic			
Single articulation	**Egressive pulmonic**: Pushing air by the lungs (outward air flow)	Various places (single closure)	
Double / secondary articulation		Various places	Various places
Ejectives	**Egressive glottalic**: Compressing air by raising the larynx (outward air flow)	Various places	Glottis
Implosives	**Ingressive glottalic**: Decompressing air by lowering the larynx (inward air flow)	Various places (single closure)	
Clicks	**Ingressive velaric**: Decompressing air by lowering the tongue body (inward air flow)	Various places (from labial to palatal)	Velum/uvula

8.5. Practice

Describing the articulation of non-pulmonic consonants can be based on the description of pulmonic consonants with adjustments that reflect the different airstream mechanism.

<u>Exercise</u>

Name and describe the articulation of the following consonants:

1. [t’]
2. [ɓ]
3. [k!]
4. [ɢǁ]
5. [ŋǂ]

<u>Solution</u>

1. [t’]:

- Name: alveolar ejective plosive.
- Place of articulation: alveolar.
- Articulators: passive: alveolar ridge, active: tongue tip/blade + larynx.

- Description: the tongue tip/blade touches the alveolar ridge, blocking airflow through the oral cavity. At the same time, the glottis is closed and the vocal cords don't vibrate. The larynx is then raised rapidly, compressing the air trapped in the oral cavity. This is followed by abrupt lowering of the tongue tip/blade, causing the trapped air to flow out in an explosive burst. Finally, the glottis is opened.

2. [ɓ]:
- Name: voiced bilabial implosive.
- Place of articulation: bilabial.
- Articulators: passive: upper lip, active: lower lip + larynx.
- Description: the lower lip touches the upper lip, blocking airflow through the oral cavity. Then, the larynx is lowered, decompressing the air trapped in the oral cavity. This is followed by abrupt lowering of the lower lip, causing an explosive flow of air in an inward direction down the oral tract. The vocal cords vibrate.

3. [k!]:
- Name: Voiceless alveolar velar click
- Place of articulation: front: alveolar, rear: velar.
- Articulators: front: passive: alveolar ridge, active: tongue tip/blade; rear: passive: velum, active: tongue back.
- Description: the tongue tip/blade touches the alveolar ridge. At the same time, the tongue back touches the velum, trapping air between the velum and the alveolar ridge. Then, the body of the tongue is lowered, decompressing the air trapped in the oral cavity. This is followed by abrupt lowering of the tongue tip/blade, producing a loud click, as air flows into the mouth. Finally, the back of the tongue is lowered. The vocal cords don't vibrate.

4. [ɢǁ]:
- Name: Voiced lateral uvular click
- Place of articulation: front: alveolar, rear: uvular.
- Articulators: front: passive: alveolar ridge, active: tongue tip/blade; rear: passive: uvula, active: tongue back.
- Description: the tongue tip/blade touches the alveolar ridge. At the same time, the tongue back touches the uvula, trapping air between the uvula and the alveolar ridge. Then, the body of the tongue is lowered, decompressing the air trapped in the oral cavity. This is

followed by abrupt lowering of the tongue tip/blade, producing a loud click, as air flows into the mouth along the sides of the tongue. Finally, the back of the tongue is lowered. The vocal cords vibrate.

5. [ŋǂ]:

- Name: Voiced palatoalveolar (palatal) velar nasal click
- Place of articulation: front: palatal, rear: velar.
- Articulators: front: passive: hard palate, active: tongue body; rear: passive: velum, active: tongue back.
- Description: the tongue body touches the hard palate, at the same time, the tongue back touches the velum, trapping air between the velum and the alveolar ridge. In addition, the velum is lowered, allowing airflow through the nasal cavity. Then, the body of the tongue is lowered, decompressing the air trapped in the oral cavity. This is followed by abrupt lowering of the tongue tip/blade, producing a loud click, as air flows into the mouth. Finally, the back of the tongue is lowered. The vocal cords vibrate.

Chapter 9 Phonation and voice quality

9.1. Phonation

In previous chapters, we used the term “voicing” as one the features of segments. “Voicing” refers to whether the vocal cords vibrate or not during articulation. Thus, when the vocal cords vibrate, the segment is voiced, and when they do not vibrate, it is voiceless (or, unvoiced). However, the term “voicing” describes only one aspect of vocal cord activity. It doesn’t describe the quality of the vibrations or the interaction between the vocal cords. For example, vibrations can be regular or irregular. In addition, in the opening phase of the vibrations, the vocal cords may open fully or partially. Similarly, in the closing phase of the vibrations, the vocal cords may close completely or remain slightly open. These and other characteristics affect the quality of the produced sounds.

Thus, we may say that voicing is a sub-type of a broader phenomenon called **phonation**. Generally, phonation is defined as the use of the larynx, aided by an airstream, to generate an audible source of acoustic energy. The larynx (“voice box”) is a complex, mostly cartilaginous, structure connecting the trachea to the pharynx and acting as a valve through which air must pass during breathing and speech (Trask, 1996). Figure 9.1 illustrates the structure of the larynx, including the position of the (true) vocal cords.

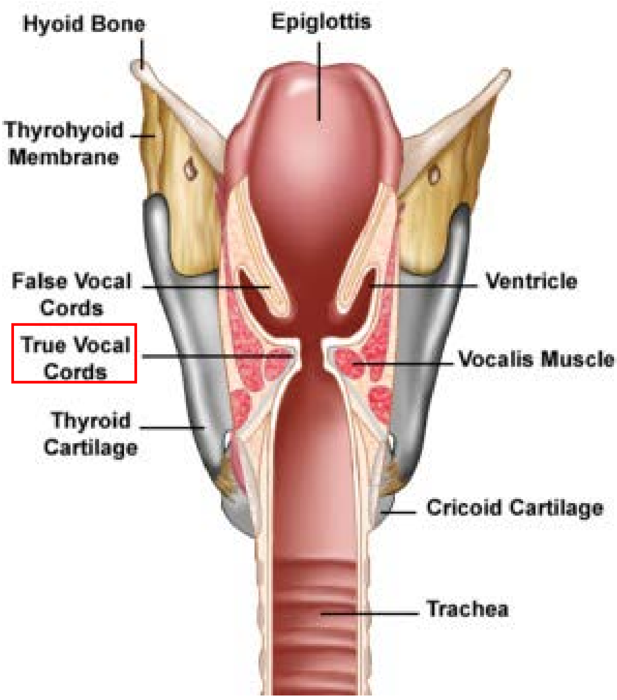


Figure 9.1 The larynx

Several types of simple phonation can be defined according to the degree of glottis constriction (glottis is the space between the vocal cords). They are illustrated in Figure 9.2. The first three categories (from the left) do not involve vibrations of the vocal cords (i.e., they describe voiceless segments). **Nil phonation** is the “classic” voiceless sound. It occurs when the glottis is held open, allowing low volume of air flow through silently. **Breath** is turbulent, noisy (voiceless) flow of air through the open glottis. Typically, the sound of [h]. The difference between nil phonation and breath is the volume of airflow, which is higher in breath.[23]

The last category of voiceless phonation is **whisper**: noisy flow of air through the glottis. Unlike in the first two categories, the glottis is not fully open in whisper. Rather, part of the glottis is tightly closed, while another part is widely open. The volume of airflow in whisper is higher than in breath and nil phonation.

The final two categories on the phonation scale are voiced. **Modal voice** is the “classic” voiced sound. The vocal cords vibrate regularly with complete closure of the glottis on each vibration. In **creak** (also, **vocal fry**), the vocal cords are closed tightly except for a small opening, which vibrates very slowly (typically 20–60 Hz), producing a series of clearly audible taps (giving the voice a harsh quality).

Most open	⟵		⟶			Most closed
	Nil	Breath	Whisper	Modal	Creak	
		Voiceless		**Voiced**		

Figure 9.2 Simple phonation according to degree of glottis constriction

Simple types of phonation can be combined to produce **compound (complex) phonation**. Notable examples include breathy voice and creaky voice. In **breathy voice** (or, **murmur**), the vocal cords vibrate but do not meet completely. The high rate of airflow produces ‘breathy’ or ‘sighing’ voice quality.[24] In **creaky voice**, part of the glottis produces a creak (vocal cords vibrate slowly and irregularly) and another part produces regular voicing.

In phonetic transcription, phonation is treated as a feature modifying a primary segment (consonant or vowel). In IPA, phonation indicated by a diacritic placed below the segment as shown in Table 9.1.

[23] Aspiration of voiceless consonants can be seen as an instance of breath phonation (see Chapter 10).

[24] Aspiration of voiced plosives can be seen as an instance of breathy voice (see Chapter 10).

Table 9.1 IPA diacritics for different phonation categories

DIACRITICS

̥	Voiceless	n̥	d̥	̤	Breathy voiced	b̤	a̤
̬	Voiced	s̬	t̬	̰	Creaky voiced	b̰	a̰

While phonation types other than nil and modal can be associated with certain speech pathologies, some languages use these phonation types to form phonological contrasts. For example, Mazatec, distinguishes modal, breathy, and creaky voice qualities in vowels (see Table 9.2). In addition, phonation can be used as a grammatical marker. For example, creaky voice typically marks phrase ending in American English (Redi & Shattuck-Hufnagel, 2001) .

Table 9.2 Phonation in Mazatec (Oto-Manguean; Mexico) (Ladefoged, 2005)

Modal voice			**Creaky voice**			**Breathy voice**		
t^{h}ǽ	‘seed’	(listen)	ndǽ̰	‘buttocks’	(listen)	ndǽ̤	‘horse’	(listen)

Phonation can also be used as a sociolinguistic marker. For example, Roach (2011) comments on the sociolinguistic role of breathy voice: “it is conventionally thought that breathy voice makes women's voices sound attractive, and it is used by speakers in television advertisements for “soft” products like toilet paper and baby powder.” Similarly, he says about creaky voice: “it has been claimed that among female speakers creak is typical of upper-class English women. “Creaky voice” tends to be associated with products that the advertisers wish to portray as associated with high social class and even snobbery (e.g. expensive sherry and luxury cars).”

Finally, phonation can also be used to express the emotions of the speaker (see more in the *planned* Chapter 26 (“Emotive speech”)). For example, Freese & Maynard (1998) claim that “the use of breathy voice contributes to the hearably “soft” or “soothing” character of many bad news deliveries and responses, while creaky voice is more apt for conveying the "groaning" quality of pain or frustration.”

9.2. Voice quality

A term closely related to phonation is **voice quality**, defined as “the characteristic auditory colouring of an individual speaker’s voice” (Laver, 1980). As the name suggests, voice quality describes the overall quality of the voice (e.g., *soft*, *harsh*). Voice quality can be associated with fixed as well as variable properties of the voice. The fixed anatomy of the vocal tract shapes the speaker’s unique voice. But, voice quality can also be manipulated (intentionally or not) to convey things such as the mood, emotions, attitudes, and social status of the speaker.

Physiologically, voice quality results from the combined effects of laryngeal activity (e.g., phonation) and the configuration of the vocal tract (e.g., smiling while talking affects our quality). Thus, voice quality is closely associated with phonation. In fact, voice quality is often described by labels of phonation (e.g., breathy, creaky). However, the two terms are not identical. First, phonation is only one factor affecting voice quality. Second, phonation is an articulatory mechanism, while voice quality refers mainly to the perceived effect.

Finally, voice quality is often compared to musical **timbre** (also, **tone color** or **tone quality**), defined as the perceived sound quality of a musical note, sound or tone. Timbre distinguishes the same note played on different instruments. In that, it is similar to voice quality which distinguishes the voices of different speakers. On the other hand, the timbre of musical instruments is usually fixed, while the human voice quality is flexible and can be changed more easily.

Chapter 10 Speech planning

Speech is a complex process involving long sequences of articulatory gestures. For example, to pronounce the word *kiss*, the speaker needs to perform the following sequence: first, raise the back of the tongue until it touches the velum to produce [k]. Then, slightly lower the tongue and stretch it forward and, at the same time, start vibrating the vocal cords, to produce the vowel [ɪ]. Finally, the vocal cords should stop vibrating and, at the same time, the tip of the tongue should approach the alveolar ridge to leave just enough room for air to flow with frication and produce [s]. To communicate effectively, speakers need to execute such sequences fast and accurately. In other words, speech requires constant planning.

This chapter reviews several phenomena related to speech planning. In particular, we will discuss the transition between segments, which is one of the most challenging parts of articulation. In fact, some of the most cross-linguistically common phonological processes are a direct result of the challenges posed by segmental transitions (see Part IV).

10.1. Voice onset time (VOT)

During the closure phase of plosives there is complete obstruction in the oral cavity. The obstruction is released in the transition to the following segment. In the transition from a voiceless plosive to a voiced segment (e.g., a vowel) there is also a change in the state of the vocal cords: from no vibrations during the closure phase to vibrations during the vowel production. The time difference (in milliseconds) between the start of vocal cord vibrations and the stop release is called **Voice Onset Time (VOT)**.

$$\boxed{\text{VOT} = (\text{time of vocal cord vibration onset}) - (\text{time of stop release})}$$

VOT defines three qualitatively different types of plosives: in **unaspirated voiceless plosives** (e.g., [p]), the beginning of the vowel is synchronized with the stop release. In other words, the vibrations of the vocal cords starts with the stop release. Thus, the VOT is 0. In **voiced plosives** (e.g., [b]), the vocal cords vibrate during the closure phase of the plosive. That is, the vibrations begin before the stop release, and the VOT is negative. Finally, in **aspirated voiceless plosives**, the vibrations of the vocal cords start somewhat after the stop release. In other words, VOT is positive. During the short period of time between the two events, the vocal cords remain slightly open, and exhaled air flows between them with friction. This process is called **aspiration**. It is essentially an instance of **breath phonation** (see Chapter 9) – the production of an *h*-like sound.[25] Aspirated consonants are denoted by a superscript h (e.g., [pʰ]).[26] Table 10.1 summarizes the three types of plosives according to VOT.

Table 10.1 Classes of plosives according to VOT

Type	**Example**	**VOT**	**Description**
Unaspirated voiceless plosive	[p]	Zero (VOT = 0)	Vibrations of vocal cords begin <u>when</u> the obstruction is released.
Voiced plosive	[b]	Negative (VOT < 0)	Vibrations of vocal cords begin <u>before</u> the obstruction is released.
Aspirated voiceless plosive	[pʰ]	Positive (VOT > 0)	Vibrations of vocal cords begin <u>after</u> the obstruction is released.

In English, aspirated voiceless stops are variants of plain voiceless stops at the beginning of the word before a vowel. For example, the ⟨k⟩ of *key* is typically aspirated (i.e., [kʰi]), but the ⟨k⟩ of *ski* is typically not (i.e., [ski]). If you put your hand in front of your mouth, you might feel an excessive puff of air following the ⟨k⟩ in *key*, but not in *ski*. In other languages, aspirated and unaspirated voiceless plosive are independent consonants (see also Section 14.1). For example, Korean has words that differ only with respect to the aspiration of a single plosive.

Table 10.2 Unaspirated and aspirated plosives in Korean (Korea) (Ladefoged, 2005)

Unaspirated			**Aspirated**		
pul	'fire'	(listen)	pʰul	'grass'	(listen)
tal	'moon'	(listen)	tʰal	'mask'	(listen)
kɨn	'weight of measure'	(listen)	kʰɨn	'large'	(listen)

Voiced plosives can also be aspirated. Aspiration of voiced plosives is called **breathy voice**, or **murmur** (see Chapter 9). It is denoted by a superscript ɦ (e.g., [dʱ]). Such plosives have negative VOT, like regular voiced plosives, but VOT cannot be used to distinguish them from other types of consonants. As can be seen in Table 10.3, Hindi contrasts plosives both on voice and aspiration.

[25] In fact, two distinct definitions of aspiration appear in the literature: one focusing on the volume of airflow, the other focusing on the relative timing of oral and laryngeal activities (see Ladefoged & Maddieson, 1996: 66-70).

[26] Note: aspiration is NOT secondary articulation.

Table 10.3 Postalveolar (retroflex) plosives in Hindi (India) (Ladefoged, 2005)

	Voiced			Voiceless		
Unaspirated	ɖal	'branch'	(listen)	ʈal	'postpone'	(listen)
Aspirated	ɖʱal	'shield'	(listen)	ʈʰal	'wood shop'	(listen)

According to the scheme described above, classes of plosives are defined categorically with respect to VOT zero. However, this is an idealized case. In reality, VOT is a continuous measure. Moreover, the VOT of unaspirated voiceless plosives, which is theoretically 0, is rarely exactly 0. In practice, unaspirated voiceless plosives have VOT which is relatively close to 0, while the VOT of voiced plosives and aspirated voiceless plosives is farther away from 0. In addition, VOT can be different for different plosives, and also for the same plosive in different languages. For example, Lisker & Abramson (1964) report the following VOT values for consonants in Hindi and Marathi (multiple productions by a single speaker of each language).

Table 10.4 Voice Onset Time (in msec) in Hindi and Marathi (Lisker & Abramson, 1964)

	Hindi		Marathi	
	Average	Range	Average	Range
/b/	-85	(-120, -40)	-117	(-160, -80)
/bʱ/	-61	(-105, 0)	-95	(-160, -65)
/p/	13	(0, 25)	11	(0, 25)
/pʰ/	70	(60, 80)	76	(40, 110)
/d/	-87	(-140,-60)	-111	(-175,-65)

Note: the values represent multiple productions for a single speaker of each language.

10.2. Intrusive consonants

Consider the etymology of some English surnames: *Johnson* is John's son and *Robertson* is Robert's son. But, *Thompson* is not Thomp's son. Rather, *Thompson* is a variant of *Thomson*, which means 'Thom's son'. To understand the origin of ⟨p⟩ in *Thompson*, we need to understand the articulatory gestures required to pronounce the name.

Figure 10.1 illustrates the articulatory configurations of [m] and [s]. These consonants occur in a sequence in *Thomson*. Consider the articulatory changes that occur in the transition from [m] to [s]: the lips are opened, the velum is raised, the vocal cord stop vibrating, and the tongue changes its position. We can see that producing the sequence [ms] requires multiple actions that need to be synchronized. What happens if these actions are not synchronized? In particular, what happens if the velum is raised before the lips are opened? As can be seen in

Figure 10.2, the result of such asynchronous actions is a bilabial stop (i.e., [b] or [p])! In addition, if the vocal cords stop vibrating before the lips are opened, we get of a voiceless stop (e.g., [p]).

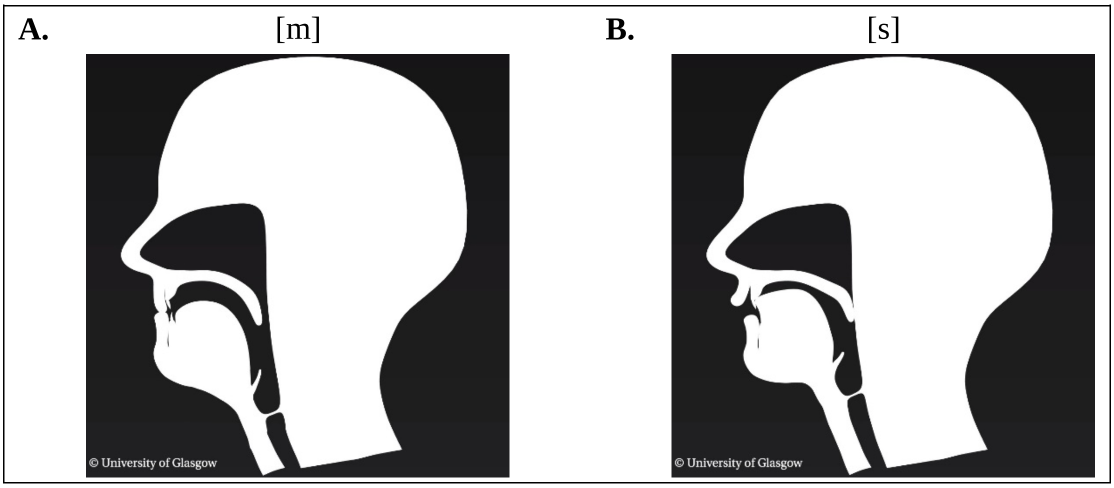


Figure 10.1 The articulatory configurations of [m] and [s]

Thus, when the transition between [m] and [s] is not perfectly synchronized, there is a short period of time during which the vocal tract is configured to produce a bilabial stop. Such a consonant that results from asynchronous transition between segments is called an **intrusive consonant**. It is an unintended consonant because there is no semantic justification for its production (there is no man named Thomp), but when speakers produce it consistently enough it becomes perceivable. The written ⟨p⟩ in Thompson is evidence for the psychological reality of the intrusive [p].

Other historical examples for intrusive stops include the French *chambre* [ʃɑ̃bʁ] ‘room’ and the English *thunder* [θʌndɚ]. *Chambre* is derived from the Latin *camera* [kamera]. The [b] in *chambre* is an intrusive consonant resulting from an asynchronous transition between [m] and the following vowel, similarly to [p] in Thompson. Because the vowel following the [m] is also voiced, the intrusive consonant in *chambre* is [b] and not [p].[27] Similarly, *thunder* which is derived from Old English *þunor* [θunor], contains an intrusive [d], resulting from an asynchronous transition between [n] and the following vowel.

[27] An alternative spelling of the Latin word, *cambra*, suggests that the intrusive [b] already existed in Medieval Latin.

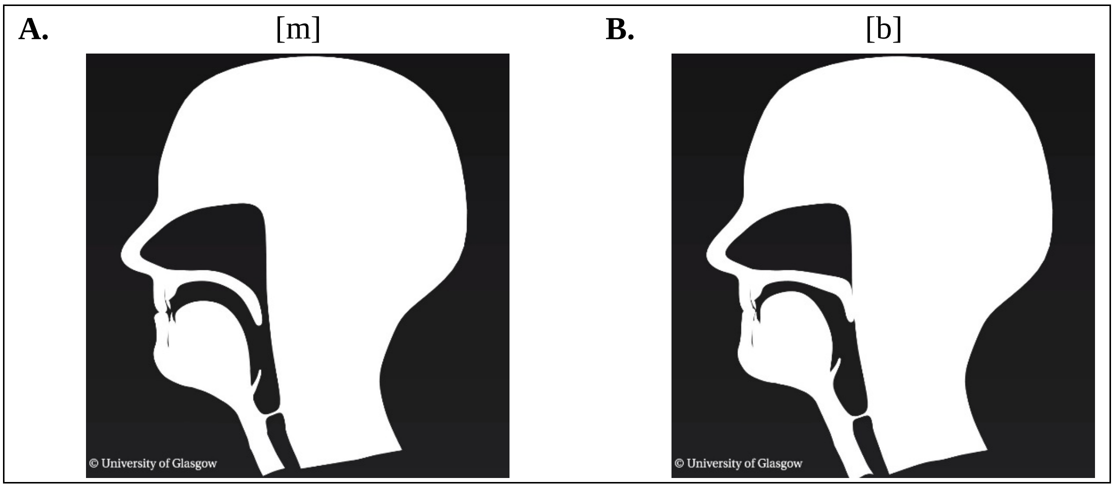


Figure 10.2 The articulatory configurations of [m] and [b]

An interesting synchronic case of an intrusive glide is found the verb system of colloquial Modern Hebrew. Hebrew verbs conjugated in the 1st person, singular, future have a prefixed ⟨א⟩ in their written form (see examples in Table 10.5). This letter was historically pronounced as a glottal stop [ʔ], but in Modern Hebrew it is usually not pronounced. As a result, when such verbs follow the first person singular pronoun *אני* /ani/ ‘I’, the combination of words contains a sequence of [i] and another vowel (e.g., *אני אביא* /ani avi/ ‘I will bring’, *אני אלך* /ani ɛlɛx/ ‘I will go’).

Table 10.5 Intrusive [j] in colloquial Modern Hebrew

Gloss	**Normative**		**Colloquial**	
	Written	**IPA**	**Written**	**IPA**
‘bring.1SG.FUT’	אביא	(ʔ)avi	יביא	javi
‘go.1SG.FUT’	אלך	(ʔ)ɛlɛx	ילך	jɛlɛx

Note that [i] is a high vowel and also the frontest of all vowels. Thus, a transition from [i] to any other vowel involves some retraction of the tongue and, usually, also some lowering. Now, if the tongue is retracted before lowering, the peak of the tongue approaches the hard palate. The resulting configuration is that of a palatal approximant [j] (see Figure 10.3). Indeed, a palatal approximant is often clearly audible in rapidly spoken sequences like [ani javi] ‘I will bring’ and [ani jɛlɛx] ‘I will go’. As in the previous examples discussed in this chapter, the intrusive [j] is reflected in the writings of many Hebrew speakers, who use a prefixed ⟨י⟩ instead of a prefixed ⟨א⟩ in these verbs (see Table 10.5), since ⟨י⟩ reflects pronunciation better ([j]). Note that written forms with a prefixed ⟨י⟩ in first person future verbs are (still) considered substandard, but they become increasingly frequent. The change is likely to be supported by the fact that ⟨י⟩ also serves as a third person future prefix (e.g., הוא ילך [hu jɛlɛx] ‘he will go’).

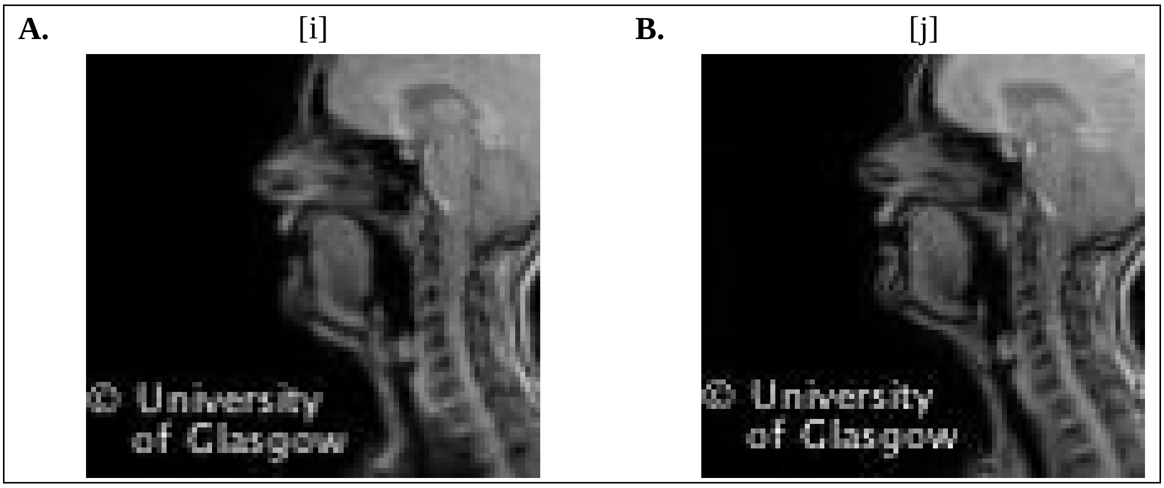


Figure 10.3 The articulatory configurations of [i] and [j]

10.3. Practice

Exercise

Name the following consonants and explain how each consonant is articulated. Relate to the active and passive articulators, the actions of the active articulator, how the airflow is affected by the interaction of the articulators, and the action of the vocal cords.

1. [t^h]
2. [$d^{\text{ɦ}}$]

Solution

1. [t^h]:

- Name: aspirated voiceless alveolar plosive.
- Articulators: passive: alveolar ridge, active: tongue tip/blade.
- Description: The tip/blade of the tongue touches the alveolar ridge blocking airflow through the oral cavity. The vocal cords do not vibrate. Air pressure increases in the oral cavity. Then, the tip/blade of the tongue is lowered abruptly, causing the air inside the oral cavity to flow out in an “explosive” burst. This is followed by turbulent, noisy (voiceless) flow of air through the open glottis.

2. [dʱ]:

- Name: aspirated voiced alveolar plosive.
- Articulators: passive: alveolar ridge, active: tongue tip/blade.
- Description: The tongue tip/blade touches the alveolar ridge blocking airflow through the oral cavity. Air pressure increases in the oral cavity. Then, the tongue tip/blade is lowered abruptly, causing the air inside the oral cavity to flow out in an “explosive” burst. The vocal cords vibrate but do not meet completely, causing turbulent, noisy flow of air through the glottis.

Part III: Acoustic phonetics and perceptual phonetics

This book is devoted to the study of speech sounds. But so far, we have not said much about the nature of sound – what it is, how it is generated, what properties it has. Chapter 11, which is dedicated to **acoustic phonetics**, the study of the physical properties of speech sounds, will provide some answers to these questions. As already mentioned in the Overview (Sections 1.3.3-1.3.4), acoustic phonetics has various theoretical and practical implications. These include, among other things: studying speech disorders, characterization of dialectal differences, and development of applications such as speech recognition and text-to-speech synthesis.

Understanding the properties of speech sounds also helps us understand how we perceive speech, another topic discussed in this part of the book. Chapter 12 describes the general properties of the **auditory system**, which is responsible for the sense of hearing and explains our ability to process the basic acoustic properties of sound. Finally, Chapter 13 discusses more complex phenomena related to **speech perception**, including categorical perception of speech sounds and "slips of the ear".

Chapter 11 Acoustic phonetics

11.1. What is sound?

To understand what sound is, let us examine what happens when our vocal cords vibrate during speech, as illustrated in Figure 11.1. When the vocal cords are at rest (represented by the parallel vertical lines in Figure 11.1A), air particles surrounding them are also at rest. When the vocal cords begin vibrating (zigzagged lines in Figure 11.1B), they move air particles in their immediate environment (represented by the circles at location 1). Those air particles are deflected from their neutral position (1) towards other air particles (location 2). Thus, the air becomes rarefied at location 1 (fewer air particles at that location) and condensed at location 2 (more air particles at that location). The collision between air particles at location 2 sends particles that were at rest before towards other particles at location 3, while particles that came from location 1 are pushed back following to the *action-reaction law* (i.e., Newton's third law of motion), as demonstrated in Figure 11.1C. Thus, air becomes condensed at locations 1 and 3, and rarefied at location 2.

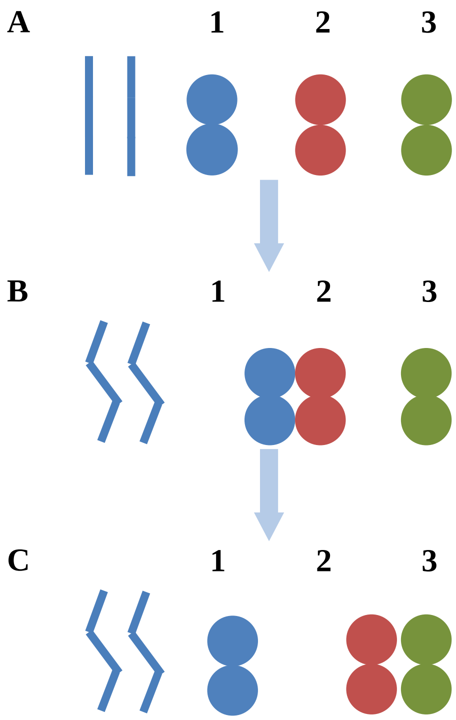


Figure 11.1 Generation of sound by the vibrating vocal cords. (A) System at rest. (B) Air particles deflected by the vibrating vocal cords, resulting in rarefication (reduced air pressure) at location 1 and compression (increased air pressure) at location 2. (C) Interactions among air particles lead to the propagation of pressure changes, which carry the sound away from the source.

The process described above continues similarly with air particles being pushed back and forth from their neutral position. This repeated sequence of collisions between air particles is what carries sound to our ears. [28] We can describe the process of sound generation in terms of air pressure changes – when air becomes rarefied at some location, air pressure decreases around that location, and when air becomes condensed – air pressure increases. As we saw in the illustration above, the collisions among air particles cause the air pressure at a given location to increase and decrease alternately. Or, if we follow the initial motion of the vocal cords, we may say that the increased air pressure resulting from that motion travels away from the vocal cords. From this description we can also conclude that our ears are sensors that detect pressure changes in the air around us.

Our description of sound generation focused on the action of the vocal cords. However, every moving object generates sound via the same mechanism, i.e., by sending vibrations through the surrounding medium (i.e., material). This is true for the buzzing sound of flying insects, loud music that can be heard through the walls, and dolphin squeals that can be heard under water from miles away. This leads us to the definition of sound:

Sound: propagation of pressure changes through a medium.

As can be seen in Figure 11.2, sound propagates in the form of a **wave**. The peaks (or, **crests**) of the wave represent high air pressure and valleys (or, **troughs**) represent low air pressure. These alternating crests and troughs travel through the air and reach our ears. In fact, sound is very similar to other wave phenomena, including water waves, "stadium" waves (where a mass of people stand up and sit down in a synchronized manner), and even light. All these phenomena involve a "disturbance" that travels through a medium.

But what is that "disturbance" that travels in a wave? It is not the medium itself, which usually doesn't move a lot. People participating in a stadium wave usually remain near their seats. Similarly, air particles that carry sound don't travel with it (when someone calls you from another room your ears do not dip in their breath. Instead, what travels in a wave is the motion itself (or energy).

[28] Every time an air particle is being pushed, it gains energy which it delivers to the next particle it collides with. However, some energy is lost on each collision, so the last particle in the chain of collision has less energy than the first particle. This is one of the reasons why sound becomes faint at large distances from the source.

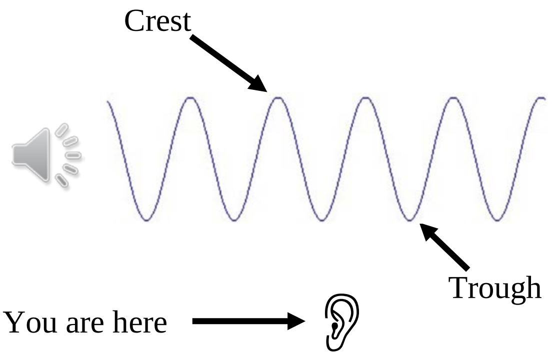


Figure 11.2 Sound propagation. The speaker icon represents the sound source, which sends vibrations through the air. The sound propagates in a wave-like form. The crests of the wave represent momentary high air pressure and the troughs represent momentary low air pressure.

11.2. Basic sound properties

When describing the properties of sound, we refer to two main dimensions: the strength of vibrations of the sound source and the distribution of vibrations in time. The strength of vibrations (or, the magnitude of pressure changes) is represented by the height of the wave crests, a measure known as **amplitude.**[29] It also roughly indicates the loudness of the sound (see more in the next chapter). Figure 11.3 compares two sound waves with different amplitudes. Higher wave amplitude indicates stronger vibrations of the sound source, greater pressure changes, and louder sound (listen to sample tones of high and low amplitudes).

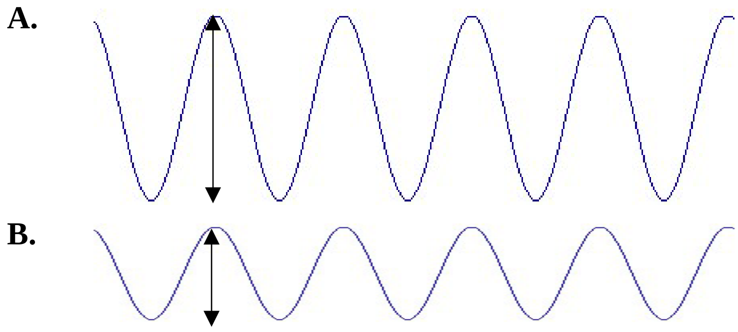


Figure 11.3 Sound amplitude. The height of the wave peak (represented by the two-sided arrow) indicates the strength of vibrations of the sound source. (A) High-amplitude sound wave. (B) Low-amplitude sound wave.

[29] There are several definitions of *amplitude*. The arrows in Figure 11.3 mark *peak-to-peak amplitude*, i.e., the difference between the crest and the trough. Another common measure is the *peak amplitude* – the difference between a crest and the baseline, which is half-way between the crest and the trough.

The amplitude of air pressure changes is an absolute quantity, measured in micro-pascals [μPa]. However, acoustic studies often report the **intensity** of sounds rather than the pressure level. Like pressure level, intensity also indicates the strength of vibrations of the sound source. However, intensity is a relative quantity rather than an absolute quantity, i.e., it is measured relative to some reference level.[30] The scale of acoustic intensity, also called **Sound Pressure Level (SPL)** is measured in decibels [dB]. A common reference level of the intensity scale, i.e., 0 dB, is equal to a pressure amplitude of 20 μPa, corresponding to the faintest sound a young healthy human can hear. Table 11.1 lists the pressure and intensity levels of common sounds.

Table 11.1 Pressure and intensity level of common sounds (from K. Johnson, 2012)

Typical experience	**Pressure [μPa]**	**Intensity [dB]**
Hearing threshold	20	0
Whisper	200	20
Quiet office	2,000	40
Conversation	20,000	60
Bus	200,000	80
Subway train	2,000,000	100
Loud thunder	20,000,000	120
Pain and damage to ears	200,000,000	140

The second property of sound is the distribution of vibrations in time. The time difference between consecutive wave crests (or troughs) is called **wavelength** (marked by λ in Figure 11.4). A more common way to refer to the temporal distribution of peaks is by **frequency** – the number of wave crests (or troughs) in a time period. Wavelength is measured in units of length (e.g., meter), while frequency is measured in **Hertz** ([Hz]) – the number of vibration cycles per second.

The wavelength and frequency of a sound represent the speed of vibrations of the sound source, but they have an inverse relation: faster vibrations of the source produce a sound with shorter wavelength (i.e., a shorter time difference between consecutive peaks) and higher frequency (more cycles per second). In addition, wavelength and frequency are roughly perceived as pitch (see more in the next chapter): a high frequency (short wavelength) sound is perceived as a high-pitched sound, while a low frequency (long wavelength) sound is

[30] The formal definition of acoustic intensity is: $L_p = 20\log_{10}\left(p/p_0\right)$, where p is the root mean square sound pressure and p_0 is the reference sound pressure (e.g., 20 μPa).

perceived as a low-pitched sound. The relation between wavelength and frequency is demonstrated in Figure 11.4 (listen to sample tones of high and low frequencies).

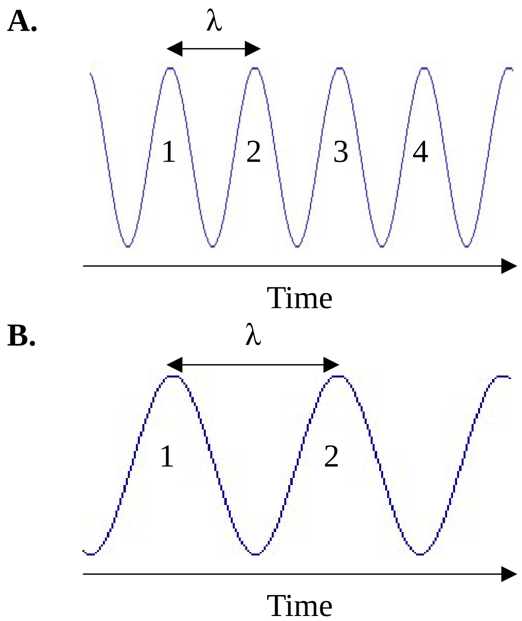


Figure 11.4 Sound wavelength and frequency. λ represents wavelength – the time difference between consecutive crests (or troughs). The numbers represents the number of vibration cycles of the source per time period. (A) High-frequency sound has more cycles per time period and a shorter wavelength, and is perceived as a high-pitched sound. (B) Low-frequency sound has fewer cycles per time period and a longer wavelength, and is perceived as a low-pitched sound.

11.3. Basic acoustic properties of speech sounds

The properties of the sound wave are also important when analyzing speech sounds. The **waveform**, i.e., the shape of the sound wave as a function of time, shows the changes in air pressure (in time) picked up by the microphone. The waveform of pure tones ("beeps") has a simple, smooth shape, as demonstrated in Figure 11.3 and Figure 11.4. The waveform of speech sounds is more complex, and there are several typical patterns, according to the manner of articulation. Figure 11.5 shows typical waveform of several types of segments.

Sonorants (i.e., vowels, glides, liquids, and nasals) are continuous, periodic sounds. Accordingly, their waveform has a relatively clear and repetitive structure (Figure 11.5A). Fricatives are continuous, aperiodic sounds. They have a sharp and dense waveform with no clear structure (Figure 11.5B). Finally, plosives are abrupt sounds. Their waveform contains a distinctive sharp spike, representing the release of the stop. In addition, we can distinguish between voiced and voiceless plosives. During the closure phase of voiced plosives, the vocal cords vibrate, producing a "humming" sound. Accordingly, their waveform has an initial periodic phase (Figure 11.5C). By contrast, no sound is produced during the closure phase of voiceless plosives, and the initial phase of their waveform is relatively flat (Figure 11.5D).

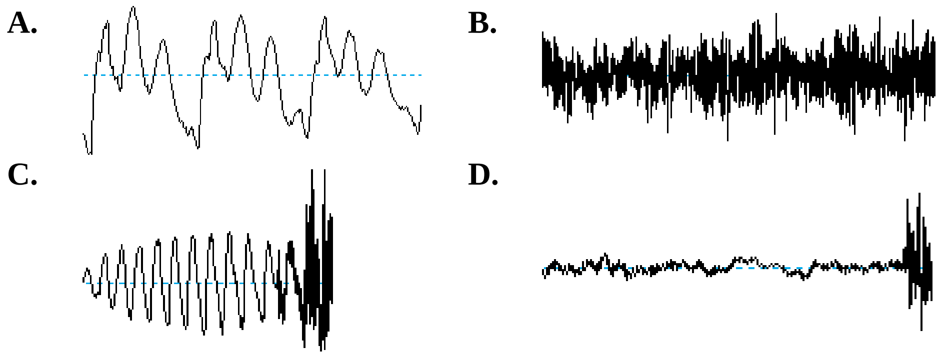


Figure 11.5 Typical waveforms of various sounds: (A) sonorant, (B) fricative, (C) voiced plosive, (D) voiceless plosive.

In addition to the waveform, acoustic software like Praat (Boersma, 2001) can typically calculate the intensity and frequency of recorded speech. In Figure 11.6, the yellow curve represents the intensity of sound in dB. The blue curve represents the **fundamental frequency (F_0)** – the vibration frequency of the vocal cords (commonly and somewhat inaccurately referred to as *pitch*). As can be seen in the figure, the F_0 curve is discontinuous. Since F_0 represents the activity of the vocal cords it can be measured only when they vibrate. Thus, a visible F_0 curve is associated with voiced segments (e.g., vowels, nasals), while in voiceless segments (e.g., [t], [s]) F_0 is typically undefined.

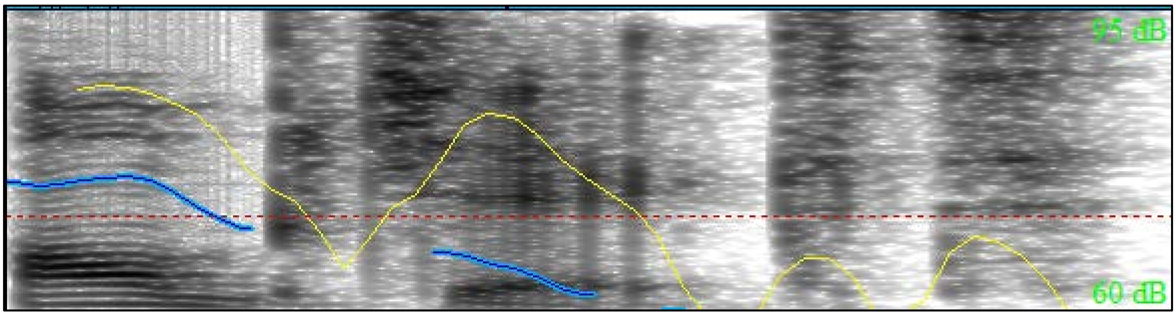


Figure 11.6 Speech acoustics on Praat. The yellow and blue curves represent the intensity and fundamental frequency, respectively. The grey-shaded background represents the spectrogram (see next section).

While the acoustic intensity of segments typically correlates with things such as stress (see Chapter 19) and the emotional intensity of the speaker (see *planned* Chapter 26 (“Emotive speech”)), it is usually not a very reliable parameter for linguistic analysis. One of the reasons for this is the fact that intensity is strongly dependent on the distance of the recording device from the sound source (more specifically, intensity is inversely proportional to the square of the distance from the source). Figure 11.7 shows how intensity (green curve) drops when the microphone moves away from the sound source (Listen).

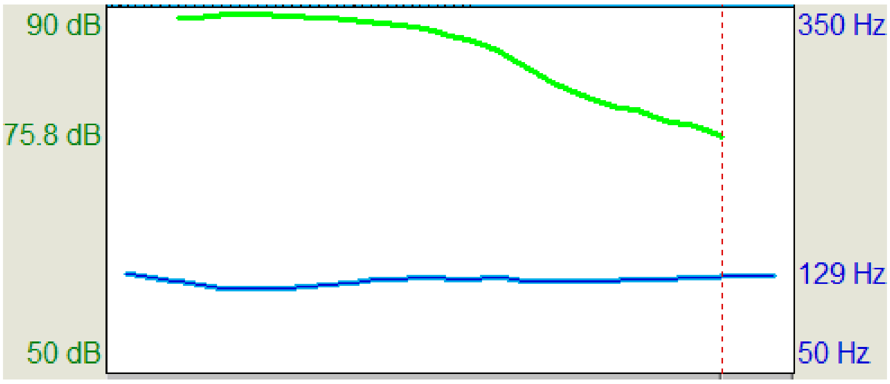


Figure 11.7 Acoustic intensity is inversely proportional to the square of distance from the sound source. Here, the author of this book pronounces a long [a] sound while holding a microphone near his mouth, then gradually moving it away by stretching his arm about 60 cm (~24″). This results in intensity (green curve) dropping from about 90 dB to 75 dB.

F_0, on the other hand, is a very useful but, by no means simple, acoustic measure. First, all else being equal, F_0 is inversely proportional to the mass or length of the sound source. That is, larger, heavier bodies have naturally lower voices than small, lighter bodies. Table 11.2 lists F_0 values of typical human voices. As can be seen, the human F_0 range is quite wide. In addition, as we shall see in Part V of this book, as well as in the *planned* Chapter 26 (“Emotive speech”), F_0 can be manipulated for various linguistic and extralinguistic purposes. In particular, F_0 is the main acoustic feature underlying intonation.

Table 11.2 Typical F_0 values of human voices (see Levrero, Mathevon, Pisanski, Gustafsson, & Reby, 2018)

Typical voice	Hz
Deep male voice	~60
Typical male voice	80-175
Typical female voice	160-270
Baby	350-550
High soprano	~1000

11.4. Spectrum

Another important acoustic property of sound is its **spectrum** (plural: spectra). To better understand what we mean by the spectrum of sound, it is easier to consider first the spectrum of another wave-like phenomenon – light. Sunlight is a mixture of lights of different colors. When these lights travel together they appear as a uniform yellowish-white light. However, when sunlight passes through water drops or a piece of glass called *prism*, the different colors are separated (see Figure 11.8). This is how we get a rainbow. The separation of light into different colors is the **spectrum of light**.

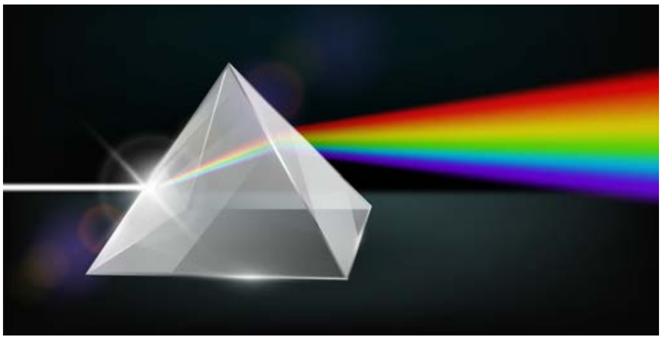


Figure 11.8 The spectrum of light can be seen when light passes through a prism

11.4.1. The spectrum of sound

Sound also has a spectrum, which can be obtained by separating the sound into tones of different frequencies. These tones are the "colors" of the sound spectrum. Pure tones have a simple spectrum with a single peak corresponding to the frequency of the tone. Figure 11.9 shows the spectra of (A) a 500-Hz pure tone (listen) and (B) a 1000-Hz pure tone (listen). The spectrum plot shows the intensity of each frequency component of the sound. As expected, the

500 Hz tone has a single narrow peak at 500 Hz, while the 1000 Hz tone a single narrow peak at 1000 Hz.[31]

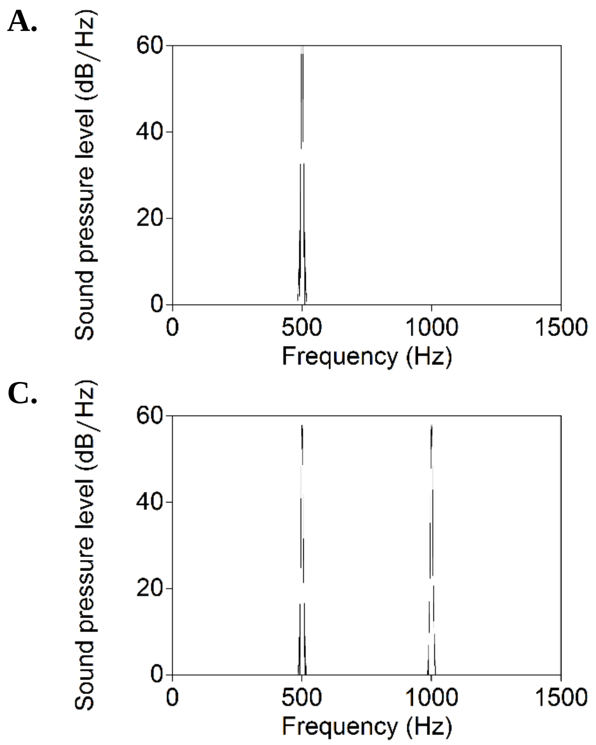


Figure 11.9 The spectra of simple and complex tones. (A) a 500-Hz pure tone, (B) a 1000-Hz pure tone, and (C) a complex of 500-Hz + 1000-Hz pure tones.

When several pure tones of different frequencies are played together, the result is a complex tone. The spectrum of complex tones has multiple peaks, each corresponding to the frequency of a single pure tone component. For example, Figure 11.9C shows the spectrum of a 500-Hz + 1000-Hz complex tone (listen).

11.4.2. **The spectrum of speech sounds**

Speech sounds have a complex spectrum with many components. These components result from the interaction of the sound wave with the vocal tract. When a sound wave is generated somewhere in the vocal tract, it travels and hits different parts of the vocal tract (the tongue, hard palate, etc.). Each such interaction generates an additional sound wave with slightly

[31] Theoretically, the spectrum of a pure tone is a sharp, infinitely narrow spike. In practice, the spike has some width to it due to audio fading effects that "guarantee that no clicks can be heard at the start and the end of the sound" (Weenink, 2018). I created these tones with Praat using fade-in and fade-out values of 0.2 sec.

different properties. All the sound waves generated from a single articulatory gesture are combined in a single segment with multiple sound "colors".

In periodic voiced sounds (vowels, glides, liquids, nasals), there are two major types of "colors" in the spectrum: **harmonics** and **formants**. Harmonics are integer multiples of the fundamental frequency (the fundamental frequency is considered the first harmonic). They are directly related to the activity of the vocal cords. Formants are additional frequencies determined by the configuration of the vocal tract. The formants together with all but the first harmonic are called **overtones** – frequencies greater than the fundamental frequency.

Figure 11.10 shows the spectrum of the vowel [o], as uttered by the author (listen). The harmonics are marked with "H". Thus, H1 represents the fundamental frequency (F_0), i.e., the frequency of the vocal cord vibrations, which in this case is 120 Hz. H2, H3… are integer multiples of H1 (i.e., $2F_0$, $3F_0$ …). Formants are marked with "F". Thus, the first formant, F1, has the lowest frequency, F2 has a higher frequency than F1, and so on. In this example, F1 is 440 Hz and F2 is 930 Hz. The spectrum shows us that the first four harmonics are the most dominant (have higher intensities). In principle, each segment has an infinite number of harmonics and formant. In practice, the first few harmonics and formants contribute the most to the perception of the segment (see next section).

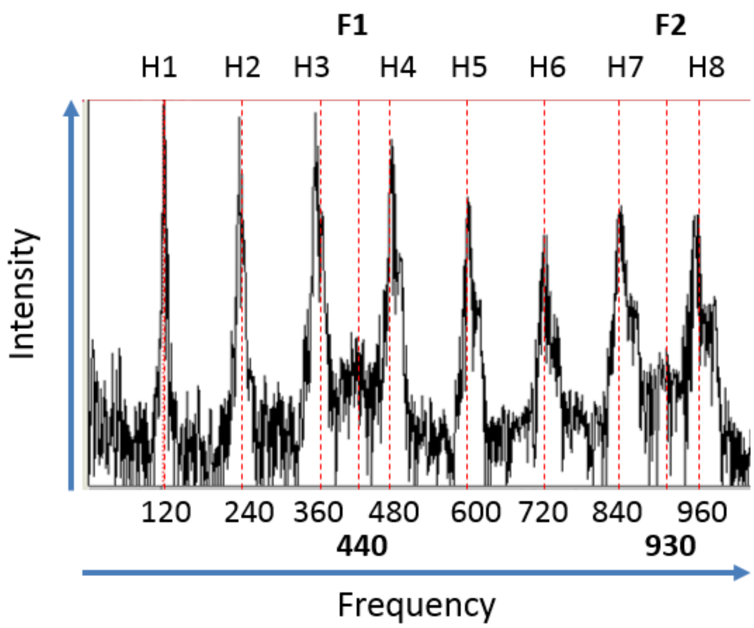


Figure 11.10 The spectrum of [o]. H1, H2, … mark harmonics (integer multiples of the fundamental frequency). F1 (440 Hz) and F2 (930 Hz) represent the first two formants.

11.4.3. **Spectrogram and formants**

Formants have an important role in speech analysis, especially of vowels. In Praat, formants can be displayed as pink and red bands over a **spectrogram**. The spectrogram is a time-dependent spectrum – it shows how the intensity of each frequency changes in time: darker areas represent frequencies with higher intensity. Figure 11.1 shows the formants of the five vowels of Modern Hebrew (Table 11.3), as uttered by the author (listen). Pink bands mark odd formants (F1, F3, and F5), red bands mark even formants (F2 and F4). In addition, the average values of the first two formants of these vowels are shown in Table 11.4. As can be seen, the vowels have rather different formant values. Indeed, it has been shown that the frequencies of the first two (or three) formants are closely linked to vowel type – the formant frequencies reflect the position of the tongue during articulation.

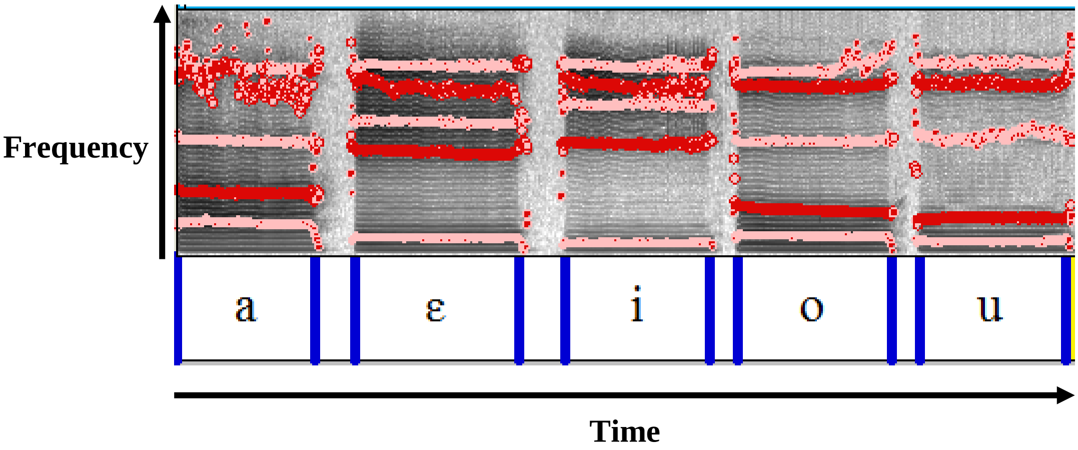


Figure 11.11 The formants of the five Hebrew vowels (uttered by the author). Pink bands mark odd formants (F1, F3, and F5), red bands mark even formants (F2 and F4).

Table 11.3 Vowel chart for Modern Hebrew

i		u
ɛ		o
	a	

Table 11.4 Average formant values of the five Hebrew vowels as uttered by a male (the author)

	[a]	[ɛ]	[i]	[o]	[u]
F2 [Hz]	1273	2085	2252	919	763
F1 [Hz]	659	365	259	399	291

The lowest pink band represents the first formant, F1. As can be seen, its frequency is proportional to vowel openness (or, inversely proportional to vowel height): open vowels (e.g.,

[a]) have higher F1 than closed vowels (e.g., [i], [u]). The lowest red band represents the second formant, F2. F2 (or rather, the difference between the first two formants, F2 - F1) is directly proportional to vowel frontness (or, inversely proportional to vowel backness): front vowels (e.g., [i], [ɛ]) have a higher F2 - F1 than back vowels (e.g., [o], [u]).

In addition to their role in distinguishing vowel qualities, formants also correlate with sex and age. Similarly to F_0, formant frequencies are inversely proportional to the size of the vocal tract, such that men tend to have lower formants than women and children (see above). However, the values of formants are less variable than F_0 and, therefore, they are more reliable indicators of sex and age. As can be seen in Table 11.5, the average formant values of Hebrew vowels are higher for women compared to men.

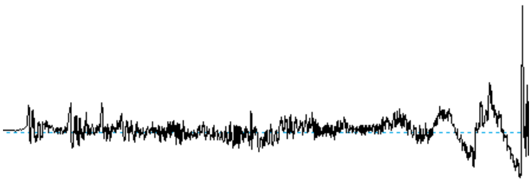
Table 11.5 Average formant values of the five Hebrew vowels as uttered by a female

	[a]	[ɛ]	[i]	[o]	[u]
F2 [Hz]	1475	2350	2710	1087	952
F1 [Hz]	1004	663	330	603	397

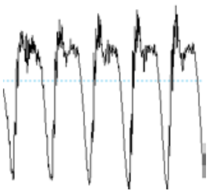
11.5. Practice

1. The figures below illustrate waveforms typical of different types of consonants. For each waveform, select the most likely corresponding consonant among the following:
   (i) [n], (ii) [g], (iii) [k], (iv) [f]. Explain your choices.

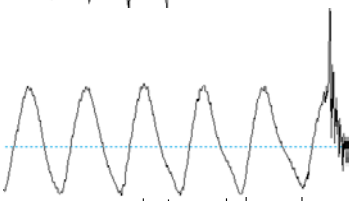
a.



b.



c.



d.

Answers:

a. The most likely consonant is [k]. The initial part of the waveform is relatively flat and it ends with a sharp spike. This is typical of voiceless plosives. The initial part reflects the closure phase (in the case of [k], when the tongue touches the velum). During that time, no sound is produced. The spike reflects the release of the plosive, when the air trapped in the oral cavity escapes in an explosive burst.

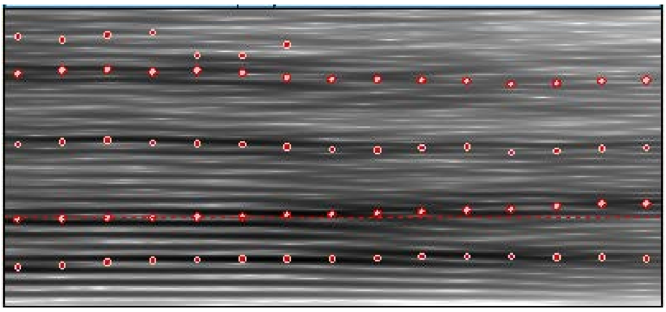
b. The most likely consonant is [n]. The waveform has a clear periodic structure that is typical of periodic sounds: vowels, glides, liquids, and nasals.

c. The most likely consonant is [g]. The waveform begins with a very smooth periodic structure that reflects the vibrating vocal cords during the closure phase of the plosive. The final part of the waveform is a sharp spike that reflects the release of the plosive.

d. The most likely consonant is [f]. The waveform is very dense and sharp with no clear structure. This is typical of fricatives.

2. The spectrograms below represent two different vowels. The values of the first two formants of each vowel are indicated near the spectrograms. Which vowel is more open (lower) and which vowel is fronter? Explain.

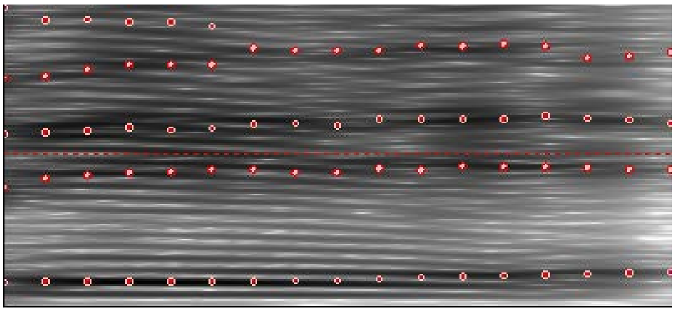
Vowel 1:



F2 = 1550 Hz

F1 = 780 Hz

Vowel 2:



F2 = 2200 Hz

F1 = 450 Hz

Answer:

Vowel 1 is more open because it has a higher F1 than vowel 2. Vowel 2 is fronter because the difference between F2 and F1 is higher than in vowel 1. Vowel 1 is [ʌ], as in *but* [bʌt], and vowel 2 is [ɪ], as in *bit* [bɪt].

Chapter 12 Perceptual phonetics: audition

This chapter discusses the **auditory system** and the basic properties of sound sensation. The auditory system is a body system responsible for the sense of hearing. It contains two sub-systems: the peripheral and the central auditory systems. The ears host the **peripheral auditory system**, which translates acoustic signals (i.e., sound) into neural signals. The ears are responsible for the perception of acoustic intensity and frequency. The **central auditory system** is located in the brain. It is responsible for the interpretation of intensity and frequency, as well as for localizing the sound source. In addition, the central auditory system sends information to the linguistic areas in the brain, which are responsible for speech perception.

This chapter describes some basic properties of the ears and briefly explains how perceive intensity and frequency of sounds. Other properties of the ears and the central auditory system are beyond the scope of this book. For more details, please refer to neuroanatomical resources (e.g., Bear, Connors, & Paradiso, 2007).

12.1. The anatomy of the ear and basic audition

The ear has three parts. The **outer ear** includes the **auricle (**or, **pinna)**, which is the only visible part of the ear, and the **external auditory canal (**or, **ear canal)** . The **middle ear** includes the **tympanic membrane** (also known as **eardrum**) and a chain of three bones called **ossicles** (in order from outer to inner: the **malleus**, the **incus**, and the **stapes**). The middle ear ends at the **oval window**. The parts beyond the oval window are part of the **inner ear**.

Basic sensation of sound begins at the pinna which collects sounds from the outside world. The sound travels down the auditory canal, and moves the eardrum. Vibrations of the eardrum are transferred by the ossicles (the smallest bones in the body) to the oval window (a membrane covering a hole in the skull), and vibrations of the oval window move the fluid inside a part of the inner ear, called the **cochlea**. A structure inside the cochlea, the **basilar membrane**, is responsible for the perception of intensity and frequency. When the fluid inside the cochlea moves, it sends vibrations that propagate through the basilar membrane according to the frequency and intensity of the sound: high frequency vibrations travel a shorter distance than low frequency vibrations, and loud sounds cause stronger vibrations in the basilar membrane. The information about the vibrations of the basilar membrane is sent to the brain via the **cochlear (**or, **auditory) nerve**.

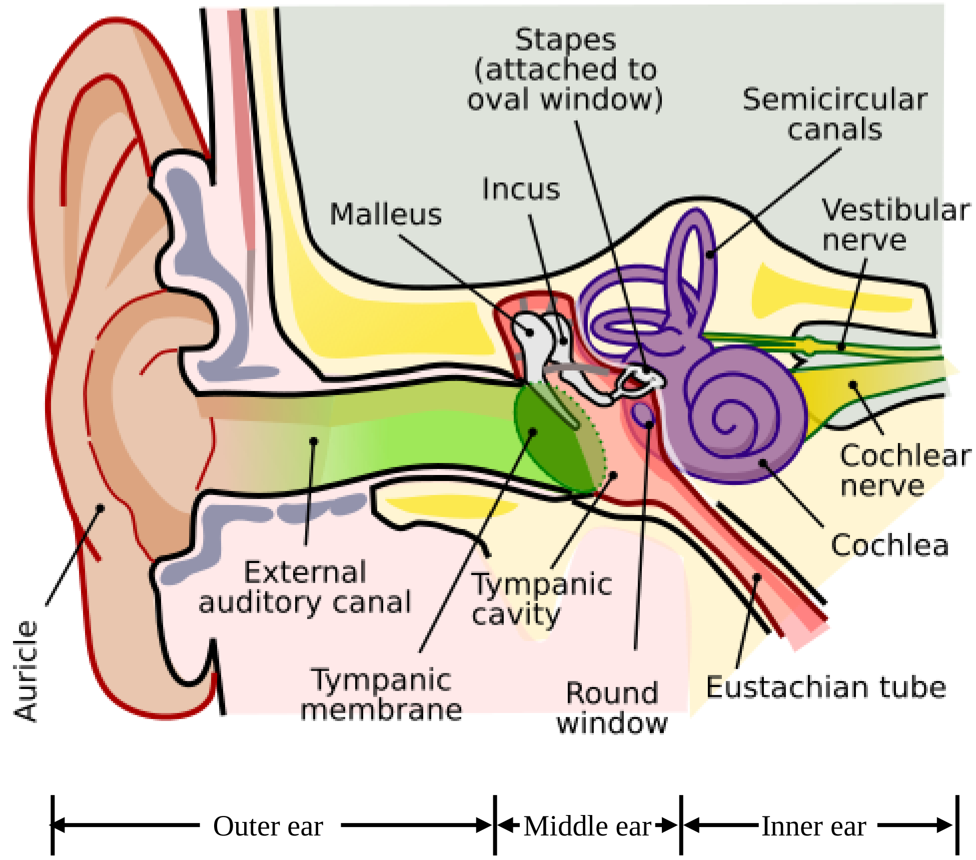


Figure 12.1 The ear (source: commons.wikimedia.org/wiki/File:Anatomy_of_the_Human_Ear.svg)

12.2. Perception of intensity and pressure

As mentioned in Chapter 11, pressure and intensity are physical, objective properties of sound (pressure is an absolute property, while intensity is a relative property). The subjective perception of pressure (or intensity) is called **loudness**: sounds of high pressure and intensity are perceived as louder than sounds of low pressure and intensity. As we shall see, despite the clear correlation between the objective intensity and the subjective perception of intensity, perception is not always accurate.

Loudness is measured in **sones**, a subjective perceptual unit. Loudness of 1 sone is defined as the perceived noise level in a quiet office. Loudness of 2 sones is defined as noise level perceived twice as loud as a quiet office, and so on. Table 12.1 lists the objective pressure and intensity levels and the corresponding subjective loudness levels of some common sounds.

Table 12.1 Pressure, intensity, and loudness of common sounds (from K. Johnson, 2012)

Typical experience	Pressure [µPa]	Intensity [dB]	Loudness [sone]
Hearing threshold	20	0	
Whisper	200	20	
Quiet office	2,000	40	1
Conversation	20,000	60	4
Bus	200,000	80	16
Subway train	2,000,000	100	64
Loud thunder	20,000,000	120	256
Pain and damage to ears	200,000,000	140	1024

The numbers in Table 12.1 suggest that loudness is a non-linear perception: the perceived loudness increases by a factor of 4 when pressure increases by a factor of 10. This means that the same absolute change in sound pressure does not always have the same effect. To see why, consider the objective and perceptual differences between pairs of experiences.

The objective difference between the pressure level in a quiet office and the pressure level of normal conversation is 18,000 µPa. Perceptually, this difference equals to 3 sones. Assuming a linear relation between pressure and perception, each sone would be equal to a pressure change of 6,000 µPa. Now, consider the difference between normal conversation and bus noise levels. The pressure differences is 180,000 µPa, but the loudness difference is only 12 sones. Again assuming a linear relation between pressure and perception, each sone would be equal to a pressure change of 15,000 µPa.

This means that in order to create equal perceptual differences, air pressure needs to increase 2.5 times more when the baseline is normal conversation than when the baseline is a quiet office. Consequently, a pressure increase of 6,000 µPa would have a weaker perceptual effect when the baseline is normal conversation than when the baseline is a quiet office. In other words, the same objective pressure change has a larger perceptual effect for soft sounds compared to loud sounds. This is illustrated in Figure 12.2, where the perceptual effect increase is much steeper for a pressure change from 0 to 500,000 µPa than for a pressure change from 500,000 to 1,000,000 µPa. In both cases, the absolute difference is 500,000 µPa, but the perceptual effect is half the size for the increase from 500,000 to 1,000,000 µPa.[32]

To experience the nonlinearity of loudness, listen to three 220-Hz pure tones of equal pressure differences: 0.1 Pa, 0.3 Pa, and 0.5 Pa (it is recommended to do this in a quiet room and use earphones to minimize external acoustic influence). If done under proper conditions, you will

[32] The sone scale (for intensities above 40dB SPL) can be calculated with the following formula: $N = 2^{(dB-40)/10}$, where dB, stands for "above 40dB SPL" (Johnson, 2012). For example, the intensity of a 500,000 µPa sound is 87.96 dB (see formula in footnote 30) and the perceptual loudness is 27.78 sones.

probably feel that the second (0.3 Pa) tone is closer in loudness to the third (0.5 Pa) tone than to the first tone (0.1 Pa). This impression is reflected in the sone scale. The loudness levels of the three tones are: 10.54, 20.42, and 27.78 sones, respectively. Thus, the perceptual difference between the first and the second tones is 9.88 sones, while the perceptual difference between the second and the third tones is only 7.36 sones.

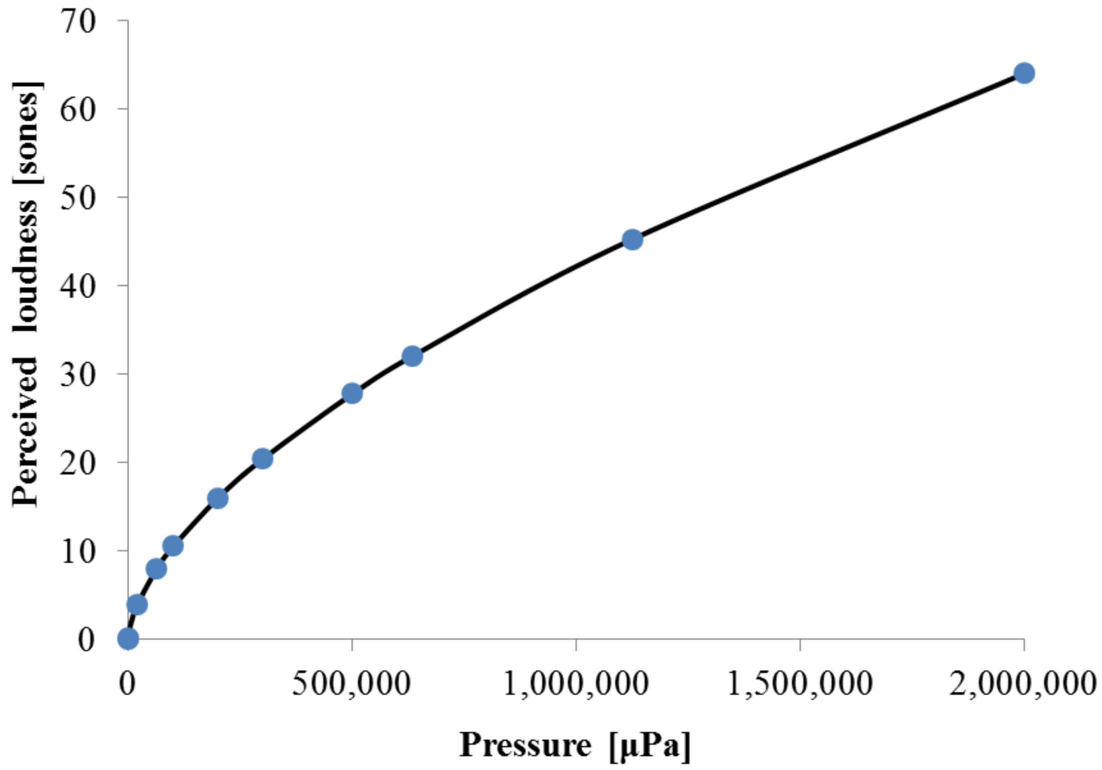


Figure 12.2 Non-linearity of loudness perception. The figure was generated based on the formulas in footnotes 30 and 32.

In addition to being nonlinear, loudness also depends on frequency. Figure 12.3 portrays the intensity of sounds of different frequencies perceived to be equally loud.[33] Four points are marked on the graph, corresponding to frequencies of 100, 1600, 3150, and 10000 Hz. Listen to these pure tones and try to order them by loudness (use earphones and click on the links). If done under proper conditions, you will probably feel that the 3150 Hz tone is the loudest, followed by the 1600 Hz, while the other two tones are considerably quieter.[34] This is despite

[33] This figure can be reproduced for various loudness levels at: https://williamssoundstudio.com/tools/iso-226-equal-loudness-calculator-fletcher-munson.php.

[34] The results are strongly dependent on the listening conditions, including the type of earphones used.

the fact that all four tones were generated with equal pressure level of 0.2 Pa. In other words, sounds of low and high frequencies need to be amplified in order to be perceived as loud as sounds of intermediate frequencies. In fact, studies have shown that the human ears are most sensitive to sounds in the range of 2000-4000 Hz (Heffner & Heffner, 2007).

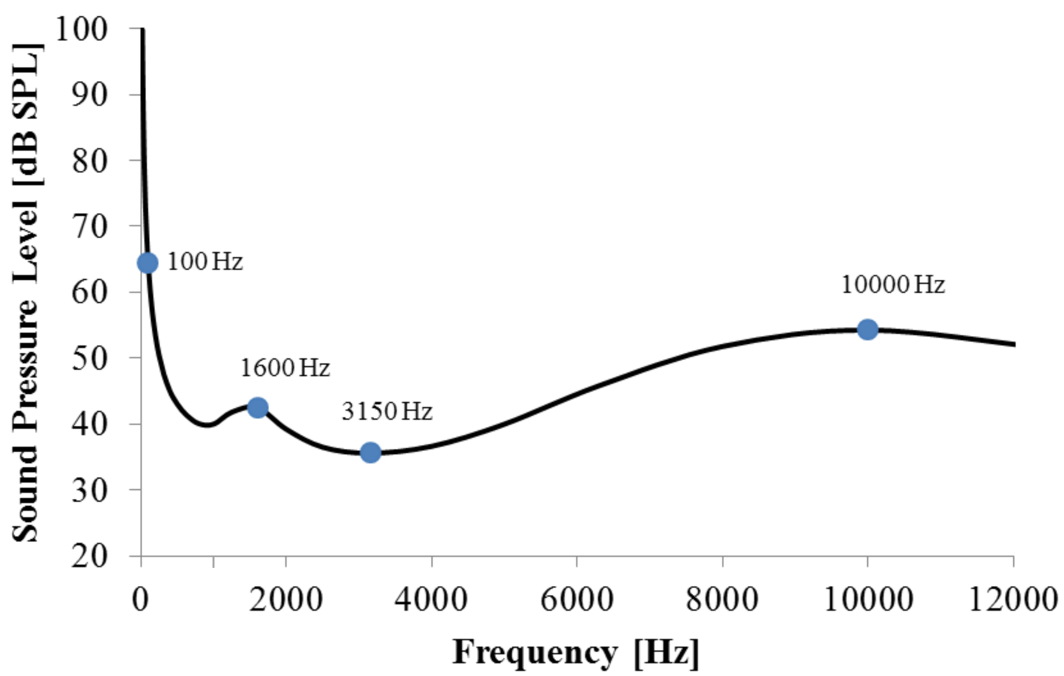


Figure 12.3 Loudness perception depends on frequency. The figure shows the intensity levels required to make sounds of different frequencies to be perceived equally loud at 40 Phon

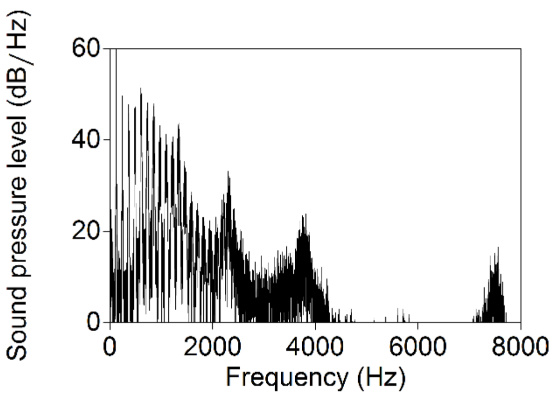
The fact that our ears are most sensitive to sounds in the range of 2000-4000 Hz has an interesting implication. Think about the sound you would make in order to silence someone. You would probably use “sh” or “s” in some combination or another. To be sure, I asked members of the Linguistics Facebook group how they would silence another person in their native language. The answers I received are listed in Table 12.2. Although this query was quite informal and the transcriptions may not be entirely consistent with each other (presumably ⟨sh⟩ stands for [ʃ] and ⟨ch⟩ for [t͡ʃ]), it is evident that speakers of different languages who come from different cultural backgrounds use either [s] or [ʃ] for the purpose of shushing. What these two sounds have in common is that they are both voiceless stridents.

Table 12.2 Shushing in different languages

Language	Silencing sounds
Arabic: Classic	sahh
Arabic: Libyan	ooos
Cantonese	sh
Castilian	tch
Croatian	psst
Dutch	shhh
French	sh
German	pssst, (p)scht
Hebrew	shhh
Ilongo	tssk
Indonesian	sh, sst
Malay	s
Persian	hissss, hishh, ssss, shshsh
Scottish	wheesht
Spanish: Guatemala	sh, shush
Spanish: Southern Spain	shush
Spanish: Venezuela	ch
Thai	shhuuu, shoooo, sh

In order to understand this seemingly universal tendency to use voiceless stridents for shushing, let us examine their spectra. Figure 12.4 shows the typical spectra of the vowel [a] and the consonants [s] and [ʃ] (as uttered by the author). The differences among the spectra are striking. In the vowel, most energy is concentrated in narrow bands (overtones) below 1000 Hz, while the spectra of the stridents in that range are virtually empty. The situation is reversed in the higher frequencies. The spectrum of [s] shows an increase in intensity from about 3000 Hz and up to a peak at around 7000 Hz. In [ʃ], the greatest concentration of energy is between 2000 and 4000 Hz. By contrast, the spectrum of [a] drops exactly in that range.

A.



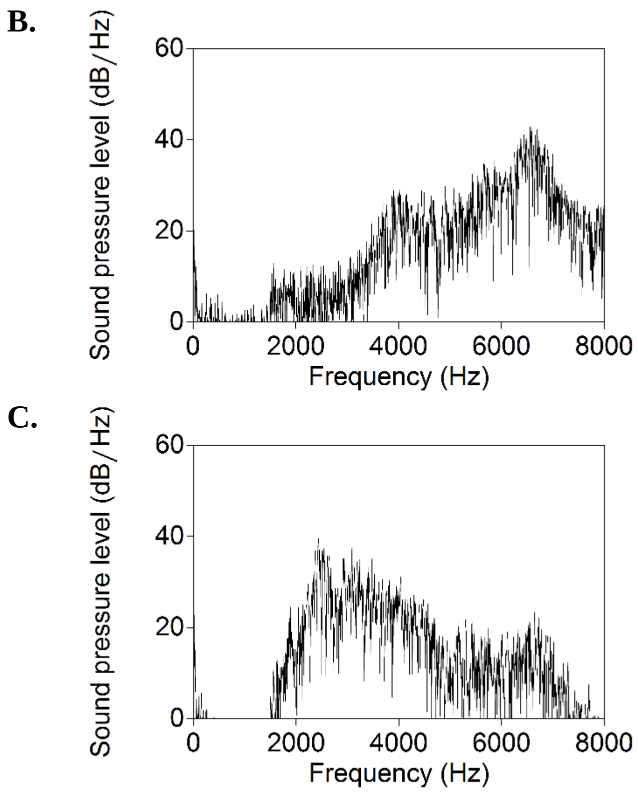


Figure 12.4 spectra of some segments: (A) [a], (B) [s], (C) [ʃ]

To make a more precise, less impressionistic comparison, we can calculate the **spectral centroid** (or, **spectral center of gravity**, or **SCoG**). This is a weighted average of the spectrum that yields a single representative value.[35] The SCoG is a measure for how high the frequencies in a spectrum are on average (it is similar to calculating the final degree grade based on the grades of all the courses a student has taken). Table 12.3 lists the SCoGs of some segments, as uttered by the author. We can see that of all segments, [ʃ] is the only one whose SCoG falls in the 2000-4000 Hz range. In other words, [ʃ] is the most perceptible segment and the most effective sound for getting someone’s attention (without shouting)! Although the SCoG of [s] is somewhat above that range, its spectrum still contains concentration of energy in the 2000-4000 Hz range, making it a noticeable sound too.

[35] Formally, the spectral centroid is given by the formula: $\mathrm{SCoG} = \int_0^\infty f\left|S(f)\right|^2 df \Big/ \int_0^\infty \left|S(f)\right|^2 df$, where f is the frequency and $S(f)$ is the value of the spectrum in frequency f.

Table 12.3 Spectral centroids of some segments (uttered by the author)

Segment	Spectral centroid [Hz]
[m]	258
[j]	278
[l]	301
[a]	619
[i]	647
[f]	1391
[ʃ]	2967
[s]	6340
[z]	7667

12.3. Perception of frequency

Frequency is an objective (physical) property of sound and the subjective perception of frequency is called **pitch** (though people often use the terms *pitch* and *fundamental frequency* interchangeably). In general, higher frequency sound are perceived as a high-pitched sounds.

It is often stated that the human hearing range is between 20 Hz and 20 kHz. However, as shown in the previous section, our ears are differentially sensitive to different frequencies. On the one hand, we can hear frequencies lower than 20 Hz if the intensity is high enough. On the other hand, the hearing range shrinks with age, especially at the higher end. Thus, only a young, healthy ear can hear sounds around 20 kHz, at any intensity (Heffner & Heffner, 2007).

Perception of frequency is measured on the **Bark** scale constructed to reflect the sensitivity of the human ear to frequency. Similarly to loudness, perception of frequency is nonlinear. This is illustrated in Figure 12.5, which shows how the perceived frequency (in Bark) changes as the function of true frequency (in Hz). Four points are marked on the graph, corresponding to frequencies of: 500, 1000, 4500, and 5000 Hz. Listen to these pure tones in pairs: first, the 500 and 1000 Hz tones, then the 4500 and 5000 Hz tones (use headphones and click on the links), and try to judge which two tones are more similar to each other.

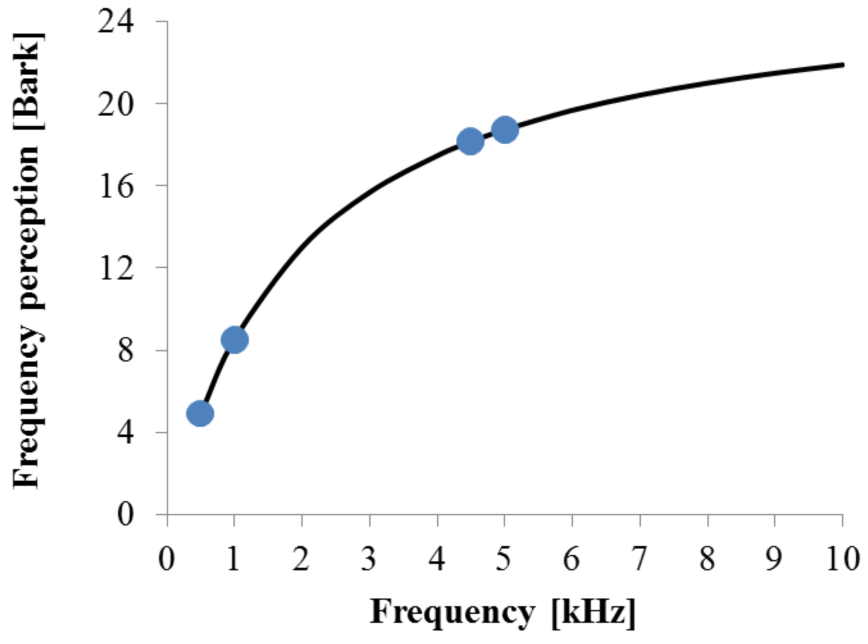


Figure 12.5 Nonlinearity of frequency perception. The figure shows how the perceived frequency (in Bark) changes as the function of true frequency (in kHz). Four points are marked on the scale: 500, 1000, 4500, and 5000 Hz. The objective difference in each pair of frequencies is 500 Hz, but the perceptual difference is larger in the lower frequencies. All tones have the same amplitude of 0.2 Pa. The figure was generated based on the formula in footnote 36.

If done under proper conditions, you will probably feel that the 4500 and 5000 Hz tones are more similar to each other than the 500 and 1000 Hz tones. This impression is reflected in the Bark values of these tones, as indicated in Table 12.4.[36] The objective difference in each pair of frequencies is 500 Hz, but the perceptual difference is six times larger in the lower frequency pair (a 3.61 Bark difference between the lower tones compared to a 0.58 Bark difference between the higher tones).

Table 12.4 Objective and perceived frequencies of four tones

Objective frequency [Hz]	**Perceived frequency [Bark]**
500	4.92
1000	8.53
4500	18.15
5000	18.73

[36] The relation between the Bark and the Hertz scales is given by: $z = \left[26.81 \big/ \left(1 + 1960/f\right)\right] - 0.53$ (Traunmüller, 1990), where *f* is the frequency in Hz and *z* is the frequency in Bark.

In other words, the same absolute change in frequency has a larger perceptual effect for low-pitched sounds compared to high-pitched sounds. This is similar to the observation we made regarding intensity. In fact, these observations are special cases of a more general fact about human perception, known as **Weber–Fechner law**: the perceived change in stimuli is proportional to the initial value of the stimuli. This law applies not only to hearing but to other senses as well. For example, it is relatively easy to perceive the difference in quantity between a group of three and a group of four people. But the same one-person difference would be virtually impossible to perceive (without counting) when comparing a group of 50 and a group of 51 people.

12.4. Practice

12.4.1. Perception of intensity & pressure

Which of the following statements is true according to Figure 12.6?

(a) The human ears are more sensitive to a frequency of 1,000 Hz than to a frequency of 40 Hz.

(b) Sounds of 400 Hz need to have higher intensity than 100 Hz sounds to be heard equally well.

(c) People can’t hear sounds of 4,000 Hz.

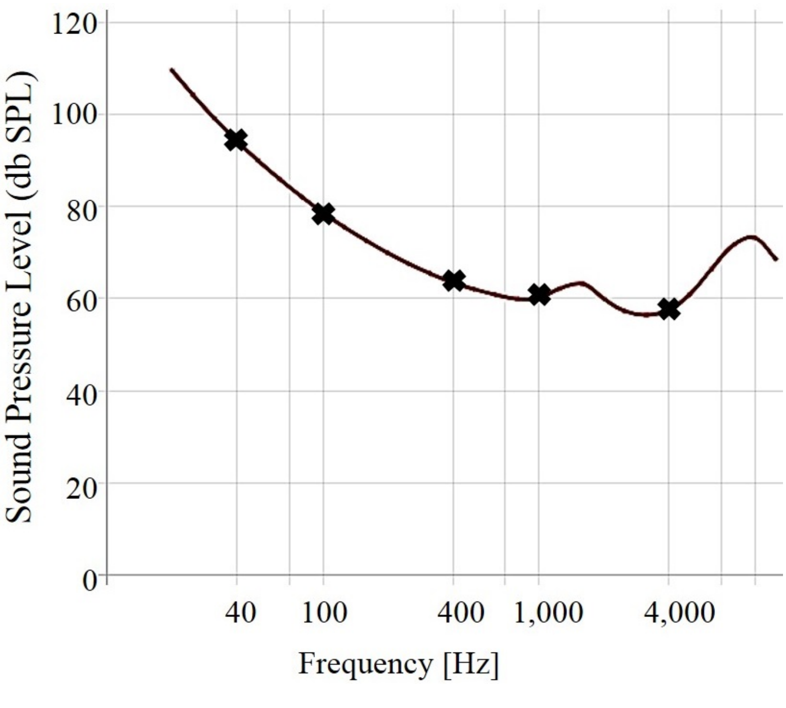


Figure 12.6 Dependence of the perception of intensity on frequency

Solution: the correct answer is **(a)**.

Explanation:

- The graph suggests that a sound of 40 Hz needs to have higher intensity than a sound of 1,000 Hz in order to be perceived as equally loud. Therefore, the human ears are naturally more sensitive to sounds of 1,000 Hz.
- Answer **(b)** is incorrect because: The graph suggests that a sound of 100 Hz needs to have higher intensity than a sound of 400 Hz in order to be perceived as equally loud.
- Answer **(c)** is incorrect because: The graph suggests that humans hear sounds of 4,000 Hz easily because they need to have relatively low intensity to be perceived as equally loud as sounds in other frequencies.

12.4.2. **Perception of frequency**

Which pair of tones is more easily distinguishable (i.e., in which pair, the perceived difference between the tones is larger): (a) 300 Hz – 500 Hz or (b) 2300 Hz – 2500 Hz? Explain.

Solution: the correct answer is **(a)**.

Explanation: pair (a) is more distinguishable since the relative difference between the tones is larger than in pair (b). Although the absolute difference is the same in both pairs (200 Hz), it is larger compared to the magnitudes of the tones in pair (a) than compared to the magnitudes of the tones in pair (b).

Chapter 13 Perceptual phonetics: speech perception

Although the ability to perceive the intensity and frequency of sounds is crucial to speech perception, most of the perceptual work in the context of speech perception is done in “higher” linguistic areas of the brain. These areas integrate information from multiple sources to facilitate comprehension. This information includes not only the acoustic signal of the uttered speech, but also our previous linguistic experience, our world knowledge, situational information about the context, and so on. In this chapter, we will review some phenomena that will serve as evidence for the incredible complexity of speech perception.

13.1. Multimodal perception

If you ever watched a video of a talking person in which the voice and movements of the lips were not synchronized, you probably found this to be distracting. Also, if you are near-sighted and use eyeglasses (or contact lenses), you might have experienced difficulty understanding what people say when you don’t have your glasses on. These observations indicate that when we are engaged in a conversation we don’t only listen to our interlocutors, but we also look at their mouths. In other words, speech perception relies on auditory information as well as on visual information. Such a perceptual process that involves information from multiple senses is called **multimodal perception**.

Another nice demonstration of the role of vision in speech perception is the **McGurk effect** (McGurk & MacDonald, 1976). This is a type of perceptual illusion by which a mismatch between audio and video recordings of speech result in a perception that doesn’t necessarily match either of the inputs. For example, if you dub an audio recording of [ba] onto a video mouthing [ga] it will sound like [da]. The effect shows that listeners combine information from the ears and eyes to make coherent phonetic judgments. The effect disappears when not looking at the video, or if the voice doesn’t match the person (e.g., a familiar face with the wrong voice).

13.2. Categorical perception

Each language has a set of sounds that native speakers identify as belonging to the language. Each sound in that set is called a **phoneme** (see more in Chapter 14). For example, native speakers of English would identify the following consonants as English phonemes: /b/, /k/, /l/, /p/, /t/, and /v/. But they would not consider /ħ/ and /ʕ/ to be English phonemes. By contrast, native speakers of Arabic would identify /ħ/ and /ʕ/, but not /p/ as Arabic phonemes.

Phonemes form the basis for **categorical perception** of speech sounds in a given language. The term "categorical perception" describes two aspects of the same phenomenon. In some contexts, categorical perception refers to the ability of speakers to perceive different pronunciations as representing a single intended categorical sound (i.e., a single phoneme). For example, voiceless plosives in English are aspirated at the beginning of a word before a vowel, but not after a consonant (see Section 10.1). Thus, /k/, /t/, and /p/ are aspirated in the following words: *key* [kʰiː], *tea* [tʰiː], and *pea* [pʰiː], but not in: *ski* [skiː], *stay* [steɪ], and *spy* [spaɪ]. Yet, native speakers of English are able to identify aspirated and non-aspirated plosives (e.g., [kʰ] and [k]) as representing the same categorical phoneme (e.g., /k/).

Categorical perception also describes the inability of speakers to discriminate different sounds in certain contexts. For example, consider the following words in Malayalam:

Table 13.1 Nasal consonants in Malayalam (Dravidian; India) (Ladefoged, 2005)

Alveolar [n]	kʌnni	'virgin'	(listen)
Retroflex [ɳ]	kʌɳɳi	'link in chain'	(listen)
Palatal [ɲ]	kʌɲɲi	'boiled rice & water'	(listen)

Each of these words contains a nasal consonant, and native speakers of Malayalam likely perceive these consonants as different. In other words, they have distinct mental categories for the three nasal consonants. By contrast, speakers whose native language does not contrast all of these consonants would probably struggle to hear a difference between them, especially when the words are uttered within a spontaneous stream of speech, rather than in isolation. If your native language has a single category for a non-labial nasal, such as /n/, you most likely hear all these consonants as different "versions" of the categorical sound /n/.

Categorical perception is an automatic, involuntary, and subconscious process. In fact, it requires conscious effort and a lot of training to notice variations in the pronunciation of phonemes and to learn new distinctions that do not exist in one's native language. In addition, categorical perception is an acquired skill. At an early age, infants are able to distinguish sounds that are not contrastive in the ambient language (such as [p] and [b] in Arabic). Yet, they show increased sensitivity to native compared to non-native sounds and, by the age of one year, they are more-or-less tuned to the phonemes of their native language (Kuhl et al., 1992).

The automaticity of categorical perception facilitates communication: it helps us understand the speech of different people, with different accents, in noisy environments, etc. However, this automaticity can lead to errors in perception, e.g., when learning a foreign language. Anyone who has ever tried to learn a foreign language or heard a person speaking a foreign language could easily discern a difference between a native and a non-native accent. To a "naïve" listener, such a difference of accent may seem as a mere nuisance – a different way to pronounce the same sounds. However, an "accent" often reflects different phonemic contrasts. That is, what may sound as "different pronunciations of the same sounds" to a non-native speaker could be perceived as entirely separate sounds by native speakers of the other language (e.g., different nasal consonants in Malayalam). One of the challenges in learning a foreign language is acquiring such new contrasts.

13.3. Perception errors

As demonstrated in the previous sections, our perceptual system is prone to make errors. This could result from lack of experience with the input (e.g., a foreign language) or from incoherent inputs (mismatch between auditory and visual inputs). However, perceptual errors can also occur in "normal" everyday situations. Auditory misperceptions (i.e., hearing something different than what was actually said) are often called **slips of the ear** (in analogy to *slips of the tongue*, which describe accidental speech errors).

A very productive source for slips of the ear is lyrics of songs. In that context, slips of the ear are also known as **Mondegreens**. A famous example of Mondegreen is due to the line "'scuse me while I kiss <u>the sky</u>" from Jimi Hendrix's *Purple Haze*, which is often claimed to be perceived as "'scuse me while I kiss <u>this guy</u>" (*KissThisGuy: The Archive of Misheard Lyrics*, 2021). The term *Mondegreen* itself is the result of misperceived words. It was coined by the American writer Sylvia Wright who described a childhood experience of misperceiving a line in one of the ballads in Percy's *Reliques*: "... and laid him on the green", which she mistakenly heard as "... and Lady Mondegreen" (Wright, 1954).

Mondegreens typically involve both phonological and semantic factors. The loud music which may mask the lyrics, the singing rhythm which may be different from the rhythm of spoken language, and slack articulation, all contribute to the ambiguity of the acoustic signal. In addition, the lyrics themselves don't always make sense (e.g., people normally kiss other people and not the sky…). The misperceived lyrics often substitute similar segments, such as [k] and [g] (see next section). This is often mediated by semantics (as in *the sky* → *this guy*), though there is no prerequisite for the perceived lyrics to be more sensible than the true lyrics.

13.4. Perceptual similarity

Examining sound replacements in slips of the ear and perception of foreign language, we may ask why sound replacements go in one direction and not another. One possible mechanism underlying these phenomena is **perceptual similarity**. Slips of the ear are more likely to replace a word/phrase with a similarly sounding word/phrase. In the "Purple Haze" example above, *the sky* [ðə skaɪ] was misperceived as *this guy* [ðɪs gaɪ], which involved two changes: /ə/ → [ɪ] and /k/ → [g]. Without going into details, we can feel intuitively that both substitutions involved relatively similar segments (e.g., it is less likely that anyone would perceive *the sky* as *that man*, which is semantically similar to *this guy* and equally plausible in the context [kiss _]). Similarly, listeners are more likely to replace a foreign segment with a similar segment from their native language. Thus, listeners with no knowledge of Malayalam are more likely to perceive [ɳ] and [ɲ] as /n/ than as /s/ or /k/.

But, what do we mean when we say that two segments are similar (or dissimilar)? How can we determine which segments are similar? And, is there a quantitative way to distinguish between more and less similar segments? Theoretically, there are a number of ways to predict the degree of similarity between any two segments, such as the number of features or natural classes they have in common (Frisch et al., 2004) (see more about features and natural classes in Chapter 15).

In practice, to determine the degree of similarity between two segments, we can measure how often listeners confuse these segments. A typical experimental paradigm for testing perceptual similarity goes as follows (see K. Johnson, 2012): participants hear a series of segments and are asked to identify each of them (e.g., via multiple choice questions). Usually, the experimenter adds noise to the recorded segments to force participants to make some errors. Then, we count the number of times a segment was mistakenly identified as a different segment. The more times segment 1 was mistakenly identified as segment 2, the more similar they are.

I conducted a short "experiment" in my class and found that consonants of the same manner of articulation and voice are more likely to be confused with each other. For example, 32% of the participants mistakenly identified /m/ as [n], but only 18% perceived it as [b]. Similarly, /k/ was identified as [t] by 68% of the participants, and only 9% identified it as [g]. Such measurements of confusion can serve as a quantitative metric for the perceptual similarity of segments.

13.5. Modes of auditory perception

Most of the time we listen to speech for comprehension. That is, we focus on sounds in order to identify words. This mode of listening is called the **speech** (or, **phonetic**) **mode** of auditory perception (Liberman & Mattingly, 1985). When operating in the speech mode, we employ categorical perception to identify spoken words. Moreover, we don’t pay attention to the properties of speech sounds for purposes other than recognizing words.

One time, I presented my students with two words in Ewe (a language spoken in Ghana): [g͡bàg͡bàg͡bà] and [kítsíkítsí] (Ohala, 1984). One of the words means ‘small’ and the other means ‘large’. When I asked my students to guess which was which, they all guessed correctly that [g͡bàg͡bàg͡bà] means ‘large’ and that [kítsíkítsí] means ‘small’. Without any semantic knowledge of these words, my students had to rely on the properties of the sounds to infer their potential meaning. In this case, it seems that they followed the common intuition that [i] “sounds” smaller than [a], and probably that voiceless consonants like [k] “sound” smaller than voiced consonants like [b] (see more on this in the *planned* Chapter 27 (“Sound symbolism”)).

As I said above, “probing” into the properties of sounds like that is something we don’t normally do. Paying attention to properties of speech sounds for purposes other than identifying words is more typical of creative use of language, such as writing poetry and slogans, or making puns. For example, in one of the episodes of the sitcom “Friends”, Phoebe and Rachel argue about the potential name for a new enterprise “a taxi that people take when they need to relax”. Rachel proposed the name “relaxi cab”, which Phoebe dismissed as “not good”, proposing “relaxi taxi” instead. The two proposed names are semantically equivalent. Thus, judging their suitability involves paying attention to properties of sounds beyond their contribution to the meaning of the words. Reuven Tsur refers to this attention to the properties of sounds as the **“poetic mode”** of auditory perception (Tsur, 1992).

13.6. Practice

13.6.1. Perceptual similarity

The following table summarizes the results of a perceptual similarity experiment. The rows of the table refer to the consonants used in the experiment, and the columns refer to the participants' responses. For example (first row), out of 28 participants, 24 identified the consonant [g] correctly and 2 mistakenly identified it as [d].

	b	d	g	k	l	m	n	p	r	t	x	Total
g		2	24								2	28
k				15						7		22
l		2			17		9					28
m	5				1	6	11					23
n					7		12		3			22

According to the above table which statement is more likely to be correct?

(a) [m] is perceptually more similar to [n] than to [b].

(b) [n] is perceptually more similar to [r] than to [l].

(c) [l] is perceptually more similar to [d] than to [n].

<u>Solution</u>: **(a)** is correct.

<u>Explanation</u>: More participants misidentified [m] as [n] (11) than as [b] (5).

Part IV: Segmental phonology

This part of the book is dedicated to segmental phonology: the study of the representation and behavior of speech sounds within a given language. Every language (and dialect) has an inventory of **phonemes** – a set of sounds that native speakers perceive as the basic sounds of their language. This set can be (and usually is) different from language to language.

Phonemes are abstract, mental representations of speech sounds. Moreover, they are not unitary, but rather, categorical representations. The set of phonemes forms the basis for categorical perception of sounds (see Section 13.2). Thus, speakers of different languages hearing the same spoken sound might perceive it differently, depending on how that sound maps onto their mental speech sound categories (i.e., phonemes).

Similarly, the same phoneme can be realized in spoken words as different sounds via **phonological processes** due to interactions of the phoneme with neighboring sounds, among other reasons. Often, there is a phonetic basis for phonological processes. The most common case is that of **assimilation** between sounds, which results from speech planning operations that reduce the number and extent of articulatory movements during speech. For example, the plural suffix of English nouns is realized as a voiceless [s] when following a voiceless sound (e.g., [kæts] 'cats') or as a voiced [z] when following a voiced sound (e.g., [dɔgz] 'dogs').

The phonetic motivation is the reason why certain phonological processes are common cross-linguistically. However, while phonetics provides the initial trigger for many phonological processes, it is not sufficient on its own. "Phonetic" processes can occur spontaneously and sporadically in the speech of an individual. Only when a community of speakers use it consistently, it becomes a phonological process – a rule in the mental grammar "book" of the language. From that point and on, the original phonetic motivation is no longer necessary to trigger the process. In fact, one might encounter curious phonological processes whose original motivation is now completely obscure. In addition, we often find well-motivated processes that apply productively in a certain part of a given language but not in other parts of the same language. Proving again that grammar is not always grounded in general cognition and physiology.

The goals of this part of the book are three-fold: (i) introducing the notions of phonological representations and phonological processes; (ii) establishing the formal principles for theoretical phonological analysis; and (iii) reviewing common phonological processes and understanding their underlying motivation. The learning process will take us on a linguistic journey across the globe, as we review data from various languages exhibiting various phonological processes.

Chapter 14 Phonological representations and phonological rules

14.1. Phonemes and allophones

Section 13.2 introduced the concept of phonemes as perceptual categories of speech sounds. The current section elaborates more on phonemes and focuses on their role as units of grammar in a given language. Consider the following list of English words given alongside their phonetic transcription.

(14-1) English phonemes

bad	[bæ**d**]
bag	[bæ**g**]
ban	[bæ**n**]
back	[bæ**k**]
bat	[bæ**t**]

It is easy to see that all these words begin with the same sequence [bæ]. The only thing differentiating the words is the final consonant. Thus, we can say that the difference in meaning across the words is due to a different final consonant. Segments that can differentiate the meanings of words (e.g., [g] and [n]) are said to be **distinct phonemes** in the given language. In formal terms (according to some theories), a **phoneme** is the minimal linguistic unit that can make a difference in meaning (see more on the notion of "phoneme" in Trask, 1996).

According to Section 13.2, the set of sounds that speakers perceive as belonging to their native language is the set of phonemes of that language, and each language/dialect has its own set of phonemes. Thus, native speakers of English would probably consider the final consonants in the words in (14-1) as basic, native consonants of the English language, because they can differentiate the meaning of words.

It is noteworthy that the examples in (14-1) demonstrated the notion of phonemes using contrasts between consonants. However, this principle applies to vowels equally. For example, the words *man* /mæn/ and *men* /mɛn/ are perceived as different words by native speakers of English. The difference in meaning can be attributed to the change of vowels: /æ/ and /ɛ/. Thus, /æ/ and /ɛ/ are considered distinct phonemes in English (the transcription in this example follows General American English).

In light of the discussion above, we can summarize the main defining properties of phonemes as follows: (i) they are perceived as basic sounds of the language by native speakers, and (ii) replacing one phoneme with another can change meaning. In phonetic transcription, phonemes are enclosed between slashes (e.g., /n/).

It should be noted that a phoneme is not an actual sound but, rather, a theoretical, abstract speech sound unit. It is the **mental representation** of a speech sound in the minds of speakers. Furthermore, a phoneme corresponds to a single sound in perception, but not in pronunciation. There is no one correct way to pronounce a phoneme. Rather, the same mental phoneme can be pronounced in different ways, e.g. in different words, by different speakers, and so on (see also Sections 2.2 and 3.5 regarding the accuracy of phonetic transcriptions). I illustrate the last claim with the following words:

(14-2) Allophones of /n/ in English

A			**B**		
1.	now	[**n**aʊ]	6.	hungry	[hʌ**ŋ**gɹi]
2.	queen	[kwiː**n**]	7.	finger	[fɪ**ŋ**gɚ]
3.	band	[bæ**n**d]	8.	bank	[bæ**ŋ**k]
4.	changes	[t͡ʃeɪ**n**d͡ʒəz]	9.	hangs	[hæ**ŋ**gz]
5.	since	[sɪ**n**s]	10.	sink	[sɪ**ŋ**k]

All the words in (14-2) contain the letter ⟨n⟩. Now, try pronouncing these words while paying attention to the position of the tongue. You might notice that when pronouncing ⟨n⟩ in the words in (A), your tongue makes contact at the front part of mouth (at the alveolar ridge or, perhaps, the upper teeth). By contrast, when pronouncing ⟨n⟩ in the words in (B), your tongue makes contact at the back part of mouth (at the velum). The given phonetic transcriptions reflect this difference: ⟨n⟩ is transcribed as [n] in (A), and as [ŋ] in (B).

Despite the different pronunciations of ⟨n⟩ in these words, “naïve” native speakers of English might consider these ⟨n⟩s as representing the same sound. They might not even notice any difference if the word *bank* is pronounced with an alveolar [n] ([bænk]) or with a velar [ŋ] ([bæŋk]). By contrast, if someone pronounces *bank* with an [m] ([bæmk]), English speakers will easily perceive this as a wrong pronunciation.

These observations demonstrate the idea of categorical perception, as discussed in Section 13.2. English speakers perceive a difference between [n] and [m] because these consonants belong to different sound categories in their minds. That is, /n/ and /m/ are distinct phonemes in English. By contrast, English speakers do not normally perceive a difference between [n] and [ŋ] because there is no “category” for a nasal velar consonant [ŋ] in English (but see footnote 41). So, when native speakers of English hear [ŋ], they subconsciously assign it to the most similar English sound, which happens to be /n/. In other words, for native speakers of English [n] and [ŋ] are two alternative pronunciations of the same intended sound. Different pronunciations of the same sound unit are said to be **allophones of the same phoneme**.

An **allophone** is the phonetic realization of a phoneme – the way the phoneme is pronounced in a given instance. Unlike phonemes, allophones are actual sounds. Roughly, we can say that a phoneme is the sound the speaker "had in mind", and an allophone is how the sound was actually pronounced.[37] In phonetic transcription, allophones are enclosed between square brackets. For example, [n] and [ŋ] are both allophones of the English phoneme /n/.

As noted above, phonemes are single sound units in perception. In order to correctly identify spoken words, listeners need to assign each pronounced sound to the intended phoneme. Thus, both [n] in [bænd] and [ŋ] in [bæŋk] should be assigned to the phoneme /n/.

In theory, each phoneme has an infinite number of possible realizations, i.e., an infinite number of allophones. For example, /n/ is pronounced by touching the alveolar ridge with the tip/blade of the tongue for a certain period of time, while the velum is lowered and the vocal cords vibrate. Any change in the exact point of contact, the duration of contact, or even the shape of the lips, creates a different allophone. In reality, when talking about allophones, we refer only to changes in pronunciation that are consistent, predictable, and large enough to be noticed (by anyone). For example, a change of 1 mm in the point of contact is probably not enough to be considered a change of allophone. On the other hand, moving the point of contact from the alveolar ridge to the velum is a major change. But, we talk about a change of allophone only when: the change is made consistently by a group of native speakers (of the same language/dialect), and, we can define the conditions under which this change is likely to occur.

Let us demonstrate the process of identifying the allophones of the same phoneme. As, we saw earlier there is a good reason to assume that, in English, [n] and [ŋ] are two allophones (realizations) of the same phoneme /n/. By examining the examples in (14-2) we can see that [ŋ] appears before [g] and [k] (words 6-10), while [n] can appear in various phonetic environments: at the beginning/end of the word and before various consonants, except for [g] and [k]. Thus, the appearance of [ŋ] is predictable (i.e., it appears only before [g] and [k]), while the appearance of [n] is less predictable – there is no simple generalization that predicts where [n] occurs. Moreover, we do not find [n] and [ŋ] in the same phonetic environment: [n]

[37] When discussing allophones we refer only to cases of "successful" production (i.e., the word was produced as the speaker intended), as opposed to cases of slips of the tongue like *take my bike* pronounced as *bake my bike* (in this case, [b] is not an allophone of /t/ because the speaker probably did not intend to say *take* with a [b]).

does not appear before [g] and [k], while [ŋ] does not appear at the beginning/end of the word, before [s], etc.

The **phonetic environment** (in short, **environment**) of a segment is the phonetic material surrounding the segment (namely, other segments and word boundaries).

Based on the observations regarding the appearance of [n] and [ŋ] we can formulate the typical characteristics of allophones of the same phoneme:

Two segments are **allophones of the same phoneme**, in a given language, if:

1. There is evidence that the two segments are linked to each other (i.e., native speakers of the language perceive them as the same sound).
2. One allophone appears in a restricted, predictable environment that can be given a simple description.
3. The other allophone is much less predictable – it appears in a more general environment.
4. The allophones do not appear in the same environment.

Every phoneme has at least one allophone – the so-called **general allophone**, which appears in the most general phonetic environment. Its features (e.g., place of articulation, voicing) are the same as those of the phoneme. In other words, every phoneme is an allophone of itself. The general allophone is represented by the same symbol as the phoneme. For example, the general allophone of /n/ is [n]. Under most circumstances, a phoneme is realized as its general allophone (this reflects the intention of the speaker). However, when a phoneme is placed in specific environments, it can have alternative realizations (for reasons to be discussed). We term such alternative realizations: **restricted allophones**. Each restricted allophone appears in a restricted and predictable phonetic environment. For example, [ŋ] is an allophone of /n/, realized before [g] and [k].

The observation that allophones of the same phoneme do not appear in the same phonetic environment is referred to as **complementary distribution**: each restricted allophone appears only in its restricted environment, while the general allophone appears everywhere except for in the environments where the restricted allophones appear. *Complementary* in this context means that the allophones complement each other: they divide all possible phonetic environments among them, such that the general allophone occupies most possible environments, and the restricted allophones occupy environments that cannot be occupied by the general allophone. For example, [n] and [ŋ] are allophones of /n/ in English. [ŋ], the restricted allophone, appears only before [ɡ] and [k]. [n], the general allophone, appears elsewhere.

A nice analogy that might help clarifying the idea of allophones and complementary distribution is the “Superman-Clark analogy”.[38] The fictional superhero named *Superman* has an alter ego, a journalist named *Clark Kent*. Under normal circumstances, the character appears in the form of Clark Kent. However, when he hears a distress call he changes his appearance and transforms into Superman. Superman and Clark Kent are obviously linked to each other – they are essentially the same person and, for that reason, they cannot appear in the same place at the same time. Thus, we might say that Clark and Superman are “allophones” of the same character. In this case, Clark Kent is analogous to the general allophone, since it is the form in which the character appears normally. Similarly, Superman is analogous to the restricted allophone, since it appears only specific circumstances (i.e., following a cry for help).

14.2. Phonemic analysis

Some of the challenges we face when analyzing phonological data from an unfamiliar language are to determine which segments are distinct phonemes, which segments are allophones of each other, and to determine the conditions under which a phoneme is realized as each of its allophones. To determine whether a set of segments are distinct phonemes or allophones of the same phoneme we follow a three-step procedure, referred to as **phonemic analysis**:

<u>Phonemic analysis</u>

Look for –

1. A minimal pair: if found, the segments are distinct phonemes.
2. A near minimal pair: if found, the segments are distinct phonemes
3. Evidence for “linkage” + complementary distribution: if found, the segments are allophones of the same phoneme

[38] I would like to thank Outi Bat-El for this analogy.

14.2.1. **Identifying distinct phonemes**

We start by looking for a **minimal pair** – two words of distinct meaning which exhibit different segments in one position but identical segments in all other positions. The segments differentiating the two words are distinct phonemes in the language. For example, consider the pair of English words: [hæ**t**] 'hat' – [hæ**d**] 'had'. The words are identical in everything except for the last consonant: [hæt] ends with a [t] and [hæd] ends with a [d]. Therefore, /t/ and /d/ are distinct phonemes in English, since they are responsible for the difference in meaning. We say that [hæt] and [hæd] are a minimal pair with respect to [t] and [d]. Similarly, /æ/ and /ʌ/ are distinct phonemes in English, since we can find a minimal pair like [hæt] 'hat' – [hʌt] 'hut'. These words are identical in everything except for the vowel ([æ] versus [ʌ]).[39]

Occasionally, we are faced with sparse data that does not contain minimal pairs. In such cases, in order to determine whether two segments are distinct phoneme we need to look for a **near-minimal pair** – two words containing different segments in the same phonetic environment, plus additional differences, which are assumed to be irrelevant. The contrasted segments in the identical environments are distinct phonemes in the language.

When analyzing environments it is important to remember that each environment has two sides: left and right. If segment *A* appears after segment *X* and before segment *Y*, the environment of *A* is: **X_Y**, where *X* is the left side of the environment, *Y* is the right side of the environment, and the underscore marks the position of *A*. For example, the phonetic environment of [æ] in [hæt] 'hat' is: h_t. The left side of the environment is: [h], and the right side of the environment is: [t]. When a segment appears at the beginning or at the end of a word, the word boundary is part of the phonetic environment of the segment. We mark word boundaries by the number sign (or, hash), #. For example, the phonetic environment of [h] in [hæt] 'hat' is: #_æ. The left side of the environment is the beginning of the word, and the right side is: [æ]. Similarly, the environment of [t] in [hæt] is: æ_# (left: [æ], right: the end of the word).

Note that the analysis of environments is done on phonetic forms rather than on orthographic forms. In other words, environments refer to segments (sounds) and not letters. For example, the letters ⟨a⟩, ⟨n⟩, and ⟨g⟩ are pronounced differently in the words *changes* and

[39] More precisely, minimal (and near-minimal) pairs prove that two segments are **allophones of distinct phonemes**. To see why, consider, for example, the pair [pʰaɪ] 'pie' and [baɪ] 'buy'. These words are a minimal pair with respect to [pʰ] and [b]. However, [pʰ] is not a phoneme in English, but a restricted allophone of /p/. To prove that a segment *X* is a phoneme in a given language, one should, theoretically, find all *X*'s allophones and show that *X* appears in a more general environment than the other allophones.

hangs, so the phonetic environments in these words are not the same, even though the letters are.

To demonstrate how analyzing the phonetic environment can help identifying the phonemes of a given language, consider a pair of words from the native Australian language Wangurri (Ladefoged, 2005): [ɡanaʔ] 'enough' – [naŋa] 'see'. Although these words do not constitute a minimal pair, they can be considered a near-minimal pair with respect to two pairs of consonants: [ɡ]–[n] and [n]–[ŋ]. First, we find both [ɡ] and [n] at the beginning of the word and before the vowel [a] (or, in the #_a environment). Similarly, both [n] and [ŋ] appear between two [a] vowels (i.e., the a_a environment). The fact that two segments can appear in exactly the same phonetic environment indicates that they are distinct phonemes. Thus, we can conclude that /ɡ/ and /n/ are distinct phonemes in Wangurri, and so are /n/ and /ŋ/.[40] In analyzing near-minimal pairs, we rely on the underlying assumption that the phonetic realization of a phoneme is determined by the environment. Therefore, if two segments can appear in the same environment, the alternation between them is not because of the environment.

Note that allophones of one phoneme in some language can be distinct phonemes in another language. For example, [n] and [ŋ] are allophones of /n/ in English – native speakers of English perceive the two consonants as representing the same sound. By contrast, /n/ and /ŋ/ are distinct phonemes in Wangurri – speakers of Wangurri perceive the two consonants as representing different sounds (i.e., sounds that can differentiate words). Similarly, as shown in Section 10.1, aspirated and non-aspirated voiceless plosives (e.g., [k] and [k^h]) are allophones of the same phoneme in English, but distinct phonemes in Korean.

14.2.2. Identifying allophones of the same phoneme

If we cannot find a minimal pair or a near-minimal pair for a pair of segments, we may suspect that they are allophones of the same phoneme. In order to establish that two segments are allophones of the same phoneme, we first need a reason to believe that they are linked to each other (i.e., represent the same phoneme). This may require somewhat more knowledge of the given language than establishing distinct phonemes.

For example, one can detect allophones of the same phoneme in one's native language by noticing that the same conceptual sound is pronounced differently in different words. Often,

[40] Normally, we examine the immediate phonetic environment (i.e., one segment to the left and one segment to the right) of target segments in order to establish near-minimal pairs. In certain cases, this is not sufficient since the realization of segments can also be affected by more distant segments (see Section 17.1.1.6).

this can be done by focusing on the pronunciation of a give letter (e.g., the pronunciation of ⟨n⟩ in *bank* and *band*), although, relying on orthography in phonological analysis is not recommended as a main strategy, as it can lead to incorrect conclusions.

A more reliable way to establish allophonic relations is through morphology, e.g., analyzing the changes in pronunciation of a suffix consonant as the suffix is attached to different bases. For example, the English plural suffix ⟨s⟩ is pronounced differently in *cats* [kæt-s] than in *dogs* [dɔg-z]. We shall discuss this method in depth in Chapter 16.

After establishing that two segments are likely linked to each other, one proceeds to examining the distributions of the segments. First, list all the environments in which each of the segments is found in the data. If one segment is found only in a restricted environment and the other segment is found everywhere except for in that environment, one may conclude that they have complementary distributions and are, thus, allophones of the same phoneme.

Note that the condition of complementary distribution is **necessary but not sufficient** for the purpose of identifying allophones of the same phoneme, since complementary distributions can also be an artificial property of small datasets. Thus, one should first find some evidence that the segments-in-question are linked to each other before attempting to determine that the segments have complementary distributions.

Let us demonstrate the process of allophonic analysis on the segments [n] and [ŋ] in the words in (14-2), repeated here for convenience.

A			**B**		
1.	now	[**n**aʊ]	6.	hungry	[hʌ**ŋ**gɹi]
2.	queen	[kwiː**n**]	7.	finger	[fɪ**ŋ**gɚ]
3.	band	[bæ**n**d]	8.	bank	[bæ**ŋ**k]
4.	changes	[t͡ʃeɪ**n**d͡ʒəz]	9.	hangs	[hæ**ŋ**gz]
5.	since	[sɪ**n**s]	10.	sink	[sɪ**ŋ**k]

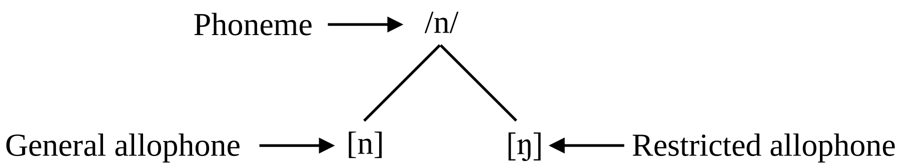
Since there are no minimal or near-minimal pairs for these segments, we make a list of environments in which each segment is found. For ease of tracking, we specify the indices of words containing each environment.

[n]		[ŋ]	
(1)	#_a	(6)	ʌ_g
(2)	iː_#	(7)	ɪ_g
(3)	æ_d	(8)	æ_k
(4)	eɪ_d͡ʒ	(9)	æ_g
(5)	ɪ_s	(10)	ɪ_k

After listing the environments of each segment, we go over them and look for a generalization. Importantly, the left and right sides of the phonetic environment are usually independent, and only one of them is relevant for identifying allophones of the same phoneme (the most notable exception for this “thumb rule” is when the restricted allophone appears only <u>between</u> vowels, while the general allophone appears either before or after vowels, but not between vowels). Therefore, we look at each side of the environment (left, right) separately.

Examining the left side of the environments listed above, we see that both consonants can appear after the vowel [ɪ] (words 5 and 10). Therefore, there is no complementary distribution on the left side environment, and we cannot determine that [n] and [ŋ] are allophones of the same phoneme based on the left side environment. On the other hand, examining the right side of the environments, we see that [ŋ] appears only before [ɡ] and [k], while [n] appears before vowels ([a]), other consonants ([d], [d͡ʒ], and [s]) and word boundary.

We see that: the right side environment of [ŋ] is restricted and can be labeled easily (i.e., before [ɡ] and [k]), while the right side environment of [n] is general and cannot be labelled as easily (it includes vowels, various consonants, and word boundary). Moreover, [n] and [ŋ] have complementary distributions on the right side: [ŋ] appears only before [ɡ] and [k], while [n] appears before elements other than [ɡ] and [k]. Therefore, we may conclude that they are allophones of the same phoneme. [ŋ] is the restricted allophone, because it appears in a more restricted environment (before [ɡ] and [k]), while [n] is the general allophone, because it appears in the more general environment: the **elsewhere** environment. The following diagram depicts the relation between the allophones and their phoneme. It is read: the phoneme /n/ has two allophones: [n] and [ŋ].[41]



[41] In reality, the status of [ŋ] in English is somewhat complicated. Historically, it was indeed only an allophone of /n/ appearing before velar consonants. However, in many English words, [ɡ] is deleted after [ŋ], obscuring the motivation of the n → ŋ rule. Moreover, such deletion can produce minimal pairs like [θɪŋ] ‘thing’ and [θɪn] ‘thin’. The existence of such minimal pairs may serve as a basis for the claim that /ŋ/ is a phoneme in English, though it is noteworthy that the distribution of /ŋ/ is highly restricted (e.g., it does not appear in word-initial position), which is atypical for ordinary phonemes. Despite the controversial status of /ŋ/ in English, I treat it as merely an allophone of /n/ for pedagogical reasons: I found that, during teaching in class, it is rather easy to explain the articulatory change between n and ŋ. By contrast, other, more clear-cut allophonic relations in English, such as *t flapping* (e.g., /wɔtəɹ/ → [wɔɾɚ] ‘water’) are harder to perceive.

When analyzing distributions of segments it is not always easy to determine whether one allophone appears in a more general environment than another. The following approach may serve as a basic guideline: phonetic entities can be divided into three groups: consonants, vowels, and word boundaries. If entities from only one of these groups constitute the phonetic environment of a segment (e.g., the segment appears only at the end of the word), it is likely a restricted allophone. On the other hand, if entities belonging to more than one of these groups can constitute the phonetic environment of a segment, we consider it a "general" environment. If we find such a general environment for both segments, it means that either they are distinct phonemes or that the specific side of the environment is irrelevant for the test. For example, if both segments appear before both consonants and vowels (general right environment), we need to look at their left environment to determine if they are allophones of the same phoneme. On the other hand, if only one of the segments has such a general environment (e.g., one segment appears before consonants and vowels, and the other appears only at the end of the word), we can say that they are likely allophones of the same phoneme. We can define environments of various degrees of "generality" following the concept of *natural classes* (see Chapter 15).

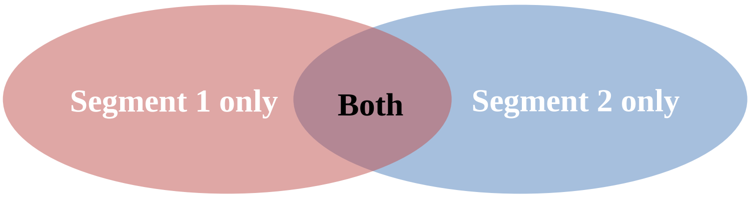
14.2.3. **"Double agents"**

Due to historical reasons, two segments can be distinct phonemes and allophones of the same phoneme in the same language. For example, *s* and *z* are distinct phonemes and also allophones of the same phoneme (/z/) in English. Such cases are typically discovered when we find (near) minimal pairs for the two segments (e.g., [suː] 'sue' vs. [zuː] 'zoo') and also an alternation between the segments around morpheme boundaries (e.g., [kæt-s] 'cats' vs. [dɔg-z] 'dogs') or in specific phonetic environments.

The existence of such "double agents" can cause some confusion and complicate the analysis (this is why I refer to them as "double agents"). I will demonstrate how to handle such cases in later chapters. In the meantime, it is important to remember that the conditions for phonemes and allophones can exist in the same data. If we find a (near) minimal pair, we can conclude with certainty that the contrasted segments are distinct phonemes in the language. Yet, they can still be allophones of each other, if they alternate (i.e., show complementary distributions) in specific environments.[42] The following diagram illustrates the relation between

[42] In fact, it is likely that "double agents" are more common in published phonological data than allophones that are not phonemes in the given language. The reason for this hypothesis is that transcribers are more likely to be sensitive to allophones that are also phonemes in their native language than to allophones that are not phonemes. For example, native speakers of English are more likely to notice and transcribe correctly alternations between [s]

the phonetic environments of “double agents”. The area labelled “Both” marks phonetic environments in which both segments can appear, that is, the environments in which they behave as distinct phonemes. The areas labelled “Segment 1 only” and “Segment 2 only” denote environments in which they behave as allophones of each other. For example, considering [s] and [z] in English, the environment in which they behave as distinct phonemes (the “Both” environment) includes the majority of possible environments (beginning of word, end of word, before and after various vowels and consonants). The exclusive environment for [s] (“Segment 1 only”) includes only cases when it is part of a suffix following voiceless consonants, while the exclusive environment for [z] (“Segment 2 only”) includes cases when it is part of a suffix following voiced segments (consonants and vowels).



14.3. Phonological rules and phonological representations

In the previous sections, we introduced the notions of phonemes and allophones. Phonemes are abstract mental representations of speech sounds and allophones are the phonetic realizations of phonemes. Under most circumstances, a phoneme is realized as its general allophone: a segment whose features are identical to those of the hypothesized phoneme. But, in certain environments the phoneme is realized as one of its alternative, restricted allophones. According to the rule-based phonological theory, the phoneme (or the general allophone) is turned into a restricted allophone via a **phonological process** that changes some of the features of the general allophone. The formulation of a phonological process is called a **phonological rule**.

A phonological rule is an expression of the form:

$$A \rightarrow B / X \_ Y$$

It reads: “*A* becomes *B* after *X* and before *Y*”, where *A* is the “input” to the rule (e.g., the phoneme), *B* is the “output” of the rule (e.g., the allophone), the underscore (_) indicates the

and [z] (both are phonemes in English) than alternations between [m] and [ɱ] (the latter is an allophone of /m/ appearing before the labiodental consonants [f] and [v], as in [ɛɱfəsɪs] ‘emphasis’).

position in which the rule applies, and *X* and *Y* are the left- and right-side phonetic environments of the altered segment, respectively. Note that *X* or *Y* can be omitted if the corresponding side of the environment is irrelevant to the rule (in most cases only one of them is actually needed). The phonological rule is a compact way to describe a situation in which a phoneme *A* is placed between the phonetic elements *X* and *Y*, forming the sequence *XAY*, but then *A* changes into *B* under the influence of *X* and *Y*, forming the sequence *XBY*.

Phonological rules are grammatical rules, just like any syntactic, semantic, or morphological rule. Thus, infants learn the phonological rules of their native language just as they learn to use articles, inflections, etc. Similarly, learning a foreign language requires learning the phonological rules of the relevant language.

To demonstrate the formulation of phonological rules, let us return to the English data in (14-2). We concluded that /n/ is a phoneme with two allophones: the general allophone [n] and the restricted allophone [ŋ], which appears only before [ɡ] and [k]. These facts are represented by two rules, each transforming [n] into [ŋ] in a different environment.

(14-3) i. n → ŋ / _ ɡ
ii. n → ŋ / _ k

Rule (14-3)i. turns /n/ into [ŋ] before [ɡ] as in the case of /fɪnɡɚ/ → [fɪŋɡɚ] 'finger'. Rule(14-3)ii. turns /n/ into [ŋ] before [k] as in the case of /bænk/ → [bæŋk] 'bank'. These formulations suggest that in words where the phoneme /n/ should come before [ɡ] and [k], it changes into [ŋ].[43] Note that these rules do not specify the *X* slot (nothing before the underscore), since the left side of the environment is irrelevant to the present dataset.

As we said, a phonological rule turns a phoneme into one of its allophones in a certain environment. But the phoneme and its phonetic environment are contained in a word. This means that a phonological rule changes the word containing the phoneme and its surrounding environment. In other words, every word has two "versions" (or two "phonological representations") – before and after the application of the rule. The version of the word before the application of a phonological rule is referred to as an **underlying representation (UR)**.

[43] In principle, we could also write a rule to describe when the phoneme /n/ is realized as the general allophone [n] (e.g., n → n / _ elsewhere). However, the convention is that the phoneme is realized as the general allophone in all cases, except for in cases described by the rules for the restricted allophones.

The version of the word after the application of the phonological rule is referred to as a **phonetic representation (PR)**. What are these phonological representations?

The PR is the phonetic realization of a word – the way it is pronounced. The UR is a more theoretical representation. According to some theorists, it is the hypothesized abstract representation of the word in the mental lexicon of the speaker. Another possible interpretation of the UR is the historical form of the word – the way people believe it was pronounced when created. In the context of the current chapter, the UR is the sequence of <u>phonemes</u> constituting the word. For example, the UR of *bank* is /bænk/ (notice the /n/). By contrast, the PR is the sequence of <u>allophones</u> of the phonemes constituting the word. For example, the PR of *bank* is [bæŋk] (notice the [ŋ]).

While the PR simply reflects the pronunciation of the word, the motivation behind the UR is mainly theoretical. By assuming the existence of these representations we can construct a phonological theory that accounts for various phonological phenomena consistently and economically. When we analyze phonological data, we are usually given the PR forms of words and our goal is to reconstruct the UR forms of these words in a way that would capture phonological "variations" we find in the data. The bottom line of the analysis would be to formulate phonological rules that turn the hypothesized URs of words into the observed PRs. The following schemes illustrate the application of the rules in (14-3) to two hypothesized URs, yielding the correct pronunciations of the words *bank* and *finger*.

UR:	/bæ**n**k/
Rule: n → ŋ / _ k	bæ**ŋ**k
PR:	[bæŋk]

UR:	/fɪ**n**gɚ/
Rule: n → ŋ / _ g	fɪ**ŋ**gɚ
PR:	[fɪŋgɚ]

The first row of each scheme specifies the hypothesized UR. The second row applies the phonological rule to the UR, transforming the phoneme /n/ into the restricted allophone [ŋ]. The last row is the final outcome – the PR. It is essentially identical to the second row (in the *planned* Chapter 22 ("Rule application and interaction") we will see that this is not always the case).

14.4. Practice

In the final section of this chapter, I demonstrate the application of the principles of phonemic analysis.

1. Consider pairs of aspirated and non-aspirated plosives in Sindhi (c-cʰ, k-kʰ, p-pʰ, t-tʰ, ʈ-ʈʰ). For the consonants in each pair, determine whether they are distinct phonemes or allophones of each other.

(14-4) **Sindhi** (Indo-European; Pakistan) (Ladefoged, 2006)

Non-aspirated			**Aspirated**		
1.	caʈu	'to destroy'	6.	cʰaʈu	'crown'
2.	taru	'bottom'	7.	tʰaru	(district name)
3.	kanu	'ear'	8.	kʰaɳu	'you lift'
4.	panu	'leaf'	9.	pʰaɳu	'snake hood'
5.	ʈanu	'ton'	10.	ʈʰaɠu	'thug, cheat'

<u>Solution</u>

The data contains two minimal pairs: words 1 and 6 constitute a minimal pair for the consonants [c] and [cʰ], respectively. Similarly, words 2 and 7 constitute a minimal pair for the consonants [t] and [tʰ], respectively. Thus, /c/ and /cʰ/ are distinct phonemes in Sindhi, and so are /t/ and /tʰ/.

While the data does not contain minimal pairs for the remaining consonants of interest, there are near-minimal pairs that allow us to conclude that they are also distinct phonemes in Sindhi: [k] and [kʰ] appear in the environment #_a in words 3 and 8, respectively. Similarly, [p] and [pʰ] appear in the same environment in words 4 and 9, respectively, as do [ʈ] and [ʈʰ] (words 5 and 10, respectively).

In summary, all the consonants in the examined pairs are distinct phonemes in Sindhi. Therefore, it is not surprising that the alphabets used for representing Sindhi contain separate letters for aspirated and non-aspirated consonants.

2. Consider the following pair of words in : [tak] 'under' – [takʰ] 'hot' (Ladefoged & Maddieson, 1996):

   A. Such a pair of words suggests that [k] and [kʰ] are (select the most appropriate answer):

   i. Distinct phonemes in Armenian

   ii. Allophones of the same phoneme in Armenian

iii. Neither of the above

Explain your choice.

Solution: (i). The given words form a minimal pair with respect to [k] and [k^h]. [k] and [k^h] can appear in the same phonetic environment in Armenian, replacing one consonant with the other can change meaning.

B. The pair of words you considered above is called (select the most appropriate answer):

i. Minimal pair

ii. Near-minimal pair

iii. Allophonic pair

Explain your choice.

Solution: (i). The given words are identical except for an alternation between [k] and [k^h].

C. Dutch does not have pairs of words that differ only in the contrast between [k] and [k^h]. Choose the most probable statement:

i. Both native speakers of Dutch and Armenian likely perceive [k] and [k^h] as different consonants.

ii. Both native speakers of Dutch and Armenian likely perceive [k] and [k^h] as the same consonant.

iii. Native speakers of Dutch likely perceive [k] and [k^h] as distinct consonants, while native speakers of Armenian likely perceive [k] and [k^h] as the same consonant.

iv. Native speakers of Armenian likely perceive [k] and [k^h] as distinct consonants, while native speakers of Dutch likely perceive [k] and [k^h] as the same consonant.

Explain your choice.

Solution: (iv). In Armenian, [k] and [k^h] appear in minimal pairs – replacing one of the consonants with the other may change meaning. Therefore, the consonants are likely perceived as distinct consonants by native speakers of Armenian. In Dutch, replacing one of the consonants with the other does not result in a change of meaning. Therefore, native speakers of Dutch probably do not perceive any contrast between the consonants.

Chapter 15 Natural classes

In the previous chapter, we saw that, in English, the phoneme /n/ is realized as [ŋ] before [ɡ] and [k]. This fact is represented by two rules:

(15-1) i. n → ŋ / _ ɡ
ii. n → ŋ / _ k

The reader might wonder why /n/ transforms into [ŋ], and why it happens in these particular cases. Examination of the IPA consonant chart might suggest an explanation: [ɡ] and [k] are velar plosives and [ŋ] is a velar nasal. It seems that whenever /n/ appears before these velar plosives, its place of articulation is shifted to the velum. This is a case of **place assimilation** – /n/ assimilates in place of articulation to the following (velar) plosive. This observation suggests that the two rules in (15-1) are, in fact, special cases of a more general rule:

(15-2) n → ŋ / _ [velar plosive]

Thus, [ɡ] and [k] have a similar effect on /n/ in English. We call such a group of segments that have a similar phonological behavior a **natural class**. More precisely, a natural class is a set of segments in a given language that share certain phonological features, and there are no other segments in that language that share the same phonological features. [ɡ] and [k] form a natural class in English – the class of velar plosives: they share place-of-articulation (velar) and manner-of-articulation (plosive), and there are no other velar plosives in English.

As the introduction to this chapter hints, our motivation for introducing the notion of natural classes, is to group segments that have a similar phonological behavior, e.g., they participate in the same phonological processes (although, similar phonological behavior is not part of the formal definition of natural classes). Such grouping will help us capture general phonological patterns in the language. In fact, as we shall see, natural classes of segments tend to behave similarly across languages. Such cross-linguistic phenomena may offer us a glimpse into cognitive processes underlying language.

Note that the notion of natural class is language-dependent: a certain set of segments that constitutes a natural class in one language, may not form a natural class in another language. For example, /p/-/t/-/k/ is the natural class of voiceless plosives in English. However, this set is not a natural class in Arabic, since Arabic does not have /p/, while it has additional voiceless plosives that do not exist in English (e.g., /ʔ/ and /q/).

15.1. Distinctive features

According to some phonological theories, segments can be seen as bundles of features. The features of consonants represent the place-of-articulation, manner-of-articulation, voicing, and secondary articulation. The features of vowels represent height, backness, roundedness, etc. A **distinctive feature** is any phonological feature that can distinguish two segments in a language (in theories assuming distinctive features, the feature is the minimal phonological unit, rather than a phoneme). For example, [voice] is a distinctive feature in English (and other languages): it distinguishes pairs of segments such as /t/ ([-voice]) and /d/ ([+voice]). The set of phonological features is defined such that no two segments will have exactly the same features. These features form the basis of natural classes. That is, a natural class is the set of all segments in a given language that share certain phonological features.

Distinctive features have one of two forms: **binary features** can have one of two opposite values: "+" (e.g., [+voice]) or "-" (e.g., [-voice]). Voicing, manner-of-articulation, vowel features, and some of the place-of-articulation features are represented by binary features. By contrast, **unary features** are single-valued features. The major place-of-articulation features are unary. For example, labial consonants have a [Labial] place feature (not [+Labial]). Non-labial consonants are not characterized as [-Labial], they simple don't have this feature (though, characterizing features as unary or binary is theory-dependent). The next section describes various phonological features that are useful for classifying various consonants and vowels.

15.2. The features of consonants

15.2.1. Manner-of-articulation features

a. **Continuant = [cont]**: continuity of airflow through the oral cavity.

- [+cont]: continuous airflow through the oral cavity: vowels, approximants, fricatives, taps, and trills.
- [-cont]: discontinuous airflow through the oral cavity: plosives, affricates, nasals.

b. **Sonorant = [son]**: sonority (roughly defined as the degree of airflow or acoustic energy of the segment; see more in Section 18.4).

- [+son]: **sonorants**. High sonority: vowels, approximants, nasals.
- [-son]: **obstruents**. Low sonority: plosives, affricates, fricatives.

c. **Nasal = [nas]**: airflow through the nasal cavity.

- [+nas]: nasal consonants, nasalized vowels.
- [-nas]: oral consonants and vowels.

d. **Lateral = [lat]**: air flows through the medial part of the mouth or along the sides of the tongue.

- [+lat]: lateral consonants – air flows along the sides of the tongue. Examples: [l], [ɮ].
- [-lat]: air flows through the medial part of the mouth. All non-lateral consonants are [-lat].

As can be seen, several groups of consonants share the same value of each feature. For example, plosives, affricates, and nasals are all [-cont]. In order to contrast these groups, we need to combine the values of several features. For example, by combining the [cont] and [son] features, we can define four unique combinations that allow to contrast four groups of consonants. This is demonstrated in Table 15.1. To characterize classes of segments in terms of feature combinations, we specify the features between square brackets, separated by commas. For example, nasals can be described as [-cont, +son], while fricatives can be described as [+cont, -son].

Table 15.1 Contrasting consonants based on the [cont] and [son] features

	[cont]	**[son]**
Plosives, affricates	-	-
Fricatives	+	-
Nasals	-	+
Approximants	+	+

Note: the combination of [cont] and [son] is insufficient for contrasting plosives and affricates. In order to contrast these consonants, we need an additional feature. **Delayed release** (or [d.r.]) is one such potential feature. In [-d.r.] consonants, such as plosives, obstruction in the oral is released abruptly, while in [+d.r.] consonants, such as affricates, obstruction in the oral is released gradually. There is also a debate regarding the features of trill and flap (or tap) consonants (see Chapter 4 in Hall, 1997)

15.2.2. **Major place-of-articulation features**

Consonants are often divided into three major classes that represent their place of articulation, mainly according to the active articulator (see more in Table 4.6). These classes are represented by the following unary features.

a. **Labial = [Lab]**: labial consonants are produced through the action of the lower lip. The term covers bilabial and labiodental consonants.
b. **Coronal = [Cor]**: coronal consonants are produced with the flexible part of the tongue, namely the tip or the blade. Accordingly, the following groups of consonants are coronal: dental, alveolar, postalveolar, and retroflex.
c. **[Dorsal] = [Dor]**: dorsal consonants are articulated with the back of the tongue, namely velar and uvular consonants. Some theorists also include pharyngeal and glottal consonants in this class.

Note that palatal consonants are articulated with the body of the tongue. Thus, strictly speaking, they belong to neither group. In practice, palatals are grouped either with coronals or with dorsals, depending on the language and the theoretical framework.

15.2.3. **Minor place-of-articulation features**

In addition to the major place-of-articulation features, there are additional place features for finer distinctions within each major class.

a. **Anterior = [ant]**: "front" vs. "rear" coronal consonants.

- [+ant]: front coronal consonants: alveolar and dental.
- [-ant]: rear coronal consonants: postalveolar, retroflex, and palatal.

b. **Strident [strd]**: **stridents** or **sibilants** are segments produced with loud friction noise. The feature is specifically useful to distinguish (inter)dental fricatives and affricates from other coronal fricatives and affricates.

- [+strd]: coronal fricatives and affricates, except for (inter)dental. Examples: [s], [ʃ], [z], [t͡s].
- [-strd]: all the other segments, specifically, (inter)dental fricatives and affricates (e.g., [θ]).

15.2.4. **Laryngeal features**

As the name suggests, these features are related to the activity of the larynx and, specifically, the vocal cords.

a. **Voice**

- [+voice]: voiced = vocal cords vibrate. Examples: [b], [v], [n], [a].
- [-voice]: voiceless = vocal cords don't vibrate. Examples: [p], [f], [t͡ʃ].

b. **Spread glottis = [spread]**: segments produced by spreading the vocal cords are accompanied by audible glottal friction.

- [+spread]: aspirated consonants (audible glottal friction) (e.g., [kʰ]).
- [-spread]: unaspirated consonants (no audible glottal friction).

15.3. The features of vowels

In principle, consonants and vowels can be described in terms of the same features. However, some features are more useful when talking about consonants, while others are more useful for describing vowels. In this section, I describe the typical vowel features. The starting point of the discussion is the vowel chart (Figure 15.1), which provides the basis for natural classes of vowels in terms horizontal tongue position, vertical tongue position, and lip rounding (features between parentheses are not official IPA features).

The IPA vowel chart provides the **"phonetic" features** of vowels. These are unary features that reflect the pronunciation of the vowels. For example, [u] is a back vowel, [ɨ] is a central vowel, and [i] is a front vowel. These features represent the horizontal position of the tongue during articulation and, as can be seen in Figure 15.1, there are at least five levels along the front-back axis. Similarly, there are at least seven levels along the open-close axis.

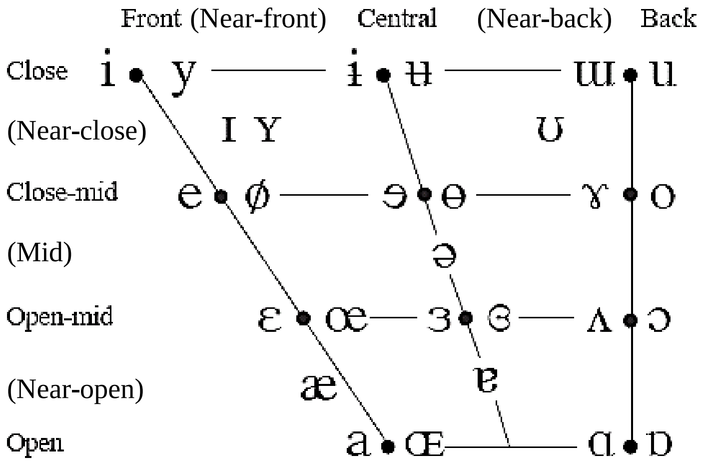


Figure 15.1 The IPA vowel chart

As in the case of consonants, phonologists prefer to represent vowels using binary rather than unary features. Moreover, the focus of phonological analysis lies on the behavior of segments rather than on their exact pronunciation. As a result, the system of binary **phonological features** of vowels is somewhat gross as far as articulation is concerned. This is mainly because there are fewer phonological levels than phonetic levels, as we will see below.

15.3.1. **The vertical dimension**

In the IPA vowel chart, the vertical dimension is represented by the categorical "Open-Close" axis. In phonological analysis, this dimension is represented by two binary features: [low] and [high], corresponding to *Open* and *Close*, respectively.

a. **[high]** :

- [+high]: the body of tongue is raised from a neutral position. This value corresponding to high (close) and near high (near close) vowels (e.g., [i], [ɪ]).
- [-high]: the body of tongue is not raised from a neutral position. This value describes all vowels that are not high or near-high.

b. **[low]** :

- [+low]: the body of tongue is lowered from a neutral position. This class describes low (open) and near low (near open) vowels (e.g., [a], [æ]) .
- [-low]: the body of tongue is not lowered from a neutral position. This class describes all vowels that are not low or near-low.

By combining the binary [high] and [low], we can define a three-way distinction along the vertical axis: high (close) vowels and near-high (near-close) vowels are [+high, -low], low (open) and near-low (near-open) vowels are [-high, +low], and all the other vowels are [-high, -low]. This classification is illustrated in Figure 15.2.

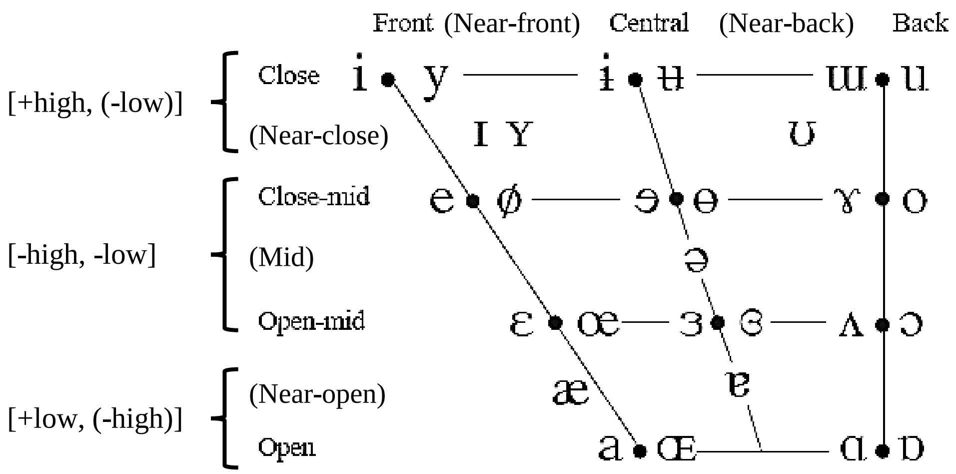


Figure 15.2 Vowel contrasts along the vertical dimension

Note that the fourth theoretical combination [+high, +low] is not used, since a vowel cannot be high and low at the same time. For the same reason, the classes of extreme vowels can be defined using a single feature ([+high] vowels are necessarily [-low], and [+low] vowels are necessarily [-high]). This is why I used parentheses inside the labels of these classes.

It is evident that this system of binary features cannot capture all the possible phonetic heights. In particular, middle vowels are indistinguishable within this system. However, from a phonological perspective, we only need to contrast vowels that show distinct phonological behavior and, as it turns out, defining three phonological vowel heights is usually sufficient for that purpose. In the case of middle vowels, we note that these vowels share [-low] with high vowels and [-high] with low vowels. This allows us to group middle vowels with either high vowels or low vowels, depending on how these groups behave in a given language. In Chapter 17, we will examine phonological processes that are sensitive to vowel height. Such processes demonstrate the idea of similar phonological behavior across different vowel groups.

Finally, in English (and possibly other languages), vowels differentiated only by a small height difference (e.g., high and near-high) can be contrasted by the [ATR] feature, which is correlated with height in some languages. Obviously, this distinction is only needed if these vowels show a distinct phonological behavior, for example, if a certain phonological process affects [+ATR] but not [-ATR] vowels.

15.3.2. The horizontal dimension

In the IPA vowel chart, the horizontal dimension is represented by the categorical "Front-Back" axis. In phonological analysis, this dimension is represented by a single binary feature: [back].

a. **[back]**:

- [+back]: the tongue is pulled to the back of the oral cavity. This class represents back and near-back vowels (e.g., [u], [ʊ]).
- [-back]: the tongue is not pulled to the back of the oral cavity. This class represents front and near-front vowels (e.g., [i], [ɪ]).

Like the vertical axis, the phonological horizontal axis of vowels is also not fully specified with respect to intermediate levels. Here, central vowels are not pre-defined for backness. Instead, their classification is language-dependent and affected by phonological behavior. Thus, in some languages they are grouped with [-back] vowels, and in others with the [+back] vowels (see more in Chapter 17). The phonological classification of vowel by backness is illustrated in Figure 15.3.

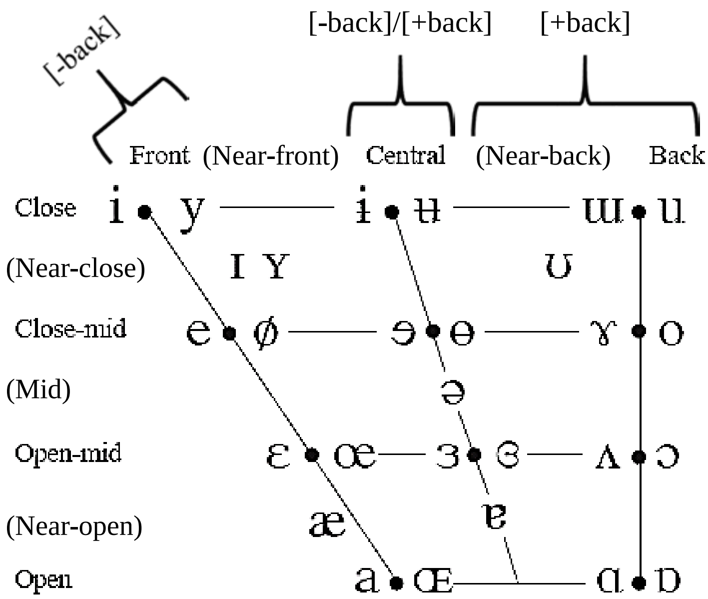


Figure 15.3 Vowel contrasts along the horizontal dimension

15.3.3. **Additional vowel features**

The other phonological vowel features are more intuitive and faithful to the articulatory properties, as defined in Chapter 7.

a. **[round]** :

- [+round]: rounded vowels (i.e., articulated with rounded lips).
- [-round]: non-rounded vowels (i.e., articulated with spread lips).

When vowels appear in pairs in the IPA chart (e.g., [i]-[y]), the vowel on the right (e.g., [y]) is [+round], and the vowel on the left (e.g., [i]) is [-round]. Of the four unpaired vowels, only [ʊ] is [+round], and the others ([ə], [ɐ], and [æ]) are [-round].

b. **[ATR]**:

- [+ATR]: tongue root stretched forward.
- [-ATR]: tongue root not stretched forward.

This feature is useful mainly for languages with vowels in “extreme” and “near-extreme” heights (i.e., close, near-close, open, near-open). The following convention is often used: close and close-mid vowels (e.g., [i], [u], [e], [o]) are classified as [+ATR], while all the other vowels (e.g., [ɪ], [ʊ], [ə], [ɛ]) are classified as [-ATR].

c. **Length**:

- [+long]: long vowels.
- [-long]: short vowels.

15.4. Major classes

We use C to represent any consonant and V to represent any vowel. While the major classes are not considered as features in the usual sense, they are relevant in this context, since they are incorporated with features in phonological notation. For example, voiceless aspirated consonants are represented as $C_{[-\text{voice},\ +\text{spread}]}$, and high back vowels are represented as $V_{[+\text{high},+\text{back}]}$.

15.5. Using distinctive features and natural classes in phonological rules

The purpose of phonological rules is to capture general phonological tendencies in a language. Thus, if we find several rules that make a similar change, we will attempt to unify them to a single rule. Generalization is done by replacing single segments with natural classes (sets of segments), represented by bundles of distinctive features. For example: as we saw, in English, /n/ is realized as [ŋ] before the velar plosives [ɡ] and [k]. Initially, we represented this fact by two specific rules:

(15-3) i. n → ŋ / _ ɡ
ii. n → ŋ / _ k

Later, we combined these rules into a more general rule:

(15-4) n → ŋ / _ [velar plosive]

Now, let us consider the following dataset featuring English adjectives in plain and negated form.

(15-5) English adjectives

	Adjective		Negated adjective	
1.	ækjɚət	‘accurate’	ɪ**n**ækjɚət	‘inaccurate’
2.	əfɛktɪv	‘effective’	ɪ**n**əfɛktɪv	‘ineffective’
3.	dɪpɛndənt	‘dependent’	ɪ**n**dɪpɛndənt	‘independent’
4.	səkjʊɻ	‘secure’	ɪ**n**səkjʊɻ	‘insecure’
5.	vælɪd	‘valid’	ɪ**n**vælɪd	‘invalid’
6.	fɻiːkwənt	‘frequent’	ɪ**n**fɻiːkwənt	‘infrequent’
7.	bælənst	‘balanced’	ɪ**m**bælənst	‘imbalanced’
8.	pɑːsəbəl	‘possible’	ɪ**m**pɑːsəbəl	‘impossible’
9.	pɝːfɛkt	‘perfect’	ɪ**m**pɝːfɛkt	‘imperfect’

The table shows that the negation prefix has two forms: [ɪn-] (words 1-6) and [ɪm-] (words 7-9). Intuitively, these forms are linked to a single underlying prefix, since they have the same meaning (negation) and a rather similar form (Indeed, historically, they originated in the Latin prefix /in-/). This suggests that [n] and [m] are allophones of the same phoneme in English.[44] Since the etymology of the prefix is known, we can determine that the phoneme is /n/. But for pedagogical purposes, I will perform a phonemic analysis to support the diachronic analysis. Thus, we begin by listing the environments in which each of the allophones appear. Note that since [n] and [m] are part of a prefix in this dataset, the left-side environment is identical in all the examples (#ɪ).

[n]		[m]	
(1)	#ɪ_æ	(7)	#ɪ_b
(2)	#ɪ _ə	(8)	#ɪ_p
(3)	#ɪ _d	(9)	#ɪ_p
(4)	#ɪ _s		
(5)	#ɪ _v		
(6)	#ɪ _f		

Examining the right-side environment, we see that [m] appears only before bilabial consonants, while [n] appears before both vowels and consonants, excluding bilabial consonants. In other words, [n] appears in the most general environment of the two allophones and, therefore, the phoneme is /n/, with [n] as the general allophone and [m] as the restricted allophone. This is illustrated in Figure 15.4:

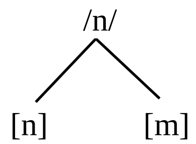


Figure 15.4 Allophones in the context of the negation prefix *in-*

Through this analysis we may conclude that /n/ is realized as [m] before bilabial plosives in English.[45] This observation can be captured by the following rule:

(15-6) n → m / _ [bilabial plosive]

[44] This observation is not in conflict with an independent observation suggesting that /n/ and /m/ are also distinct phonemes in English, a phenomenon, which I termed "double agents" (see Section 14.2.3 and more below).

[45] This conclusion is only partly true, since [n] does surface before [b] and [p] in other English words. Here, I preferred the simplified analysis to demonstrate that plosives in different places of articulation form a natural class as they can affect /n/ in a similar fashion. See a more accurate analysis of the /in-/ prefix in Section 16.2.4.

Recall that earlier we have also established the following rule:

(15-7) n → ŋ / _ [velar plosive]

These rules have a similar structure: both state that [n] is replaced by another consonant before plosives. Moreover, the described processes seem to be of the same kind: both involve a change in the place of articulation of /n/ that is in accordance with the place of articulation of the following plosive. Thus, /n/ changes to a bilabial consonant [m] before bilabial plosives, and to a velar consonant [ŋ] before velar plosives.

A change in the features of a segment in accordance with the features of a neighboring segment is referred to as **assimilation**. In this case, the place of articulation of /n/ is changed according to the place of articulation of the following plosive. This process is known as **place assimilation** (i.e., /n/ assimilates in place of articulation to the following plosive). Place assimilation is a common process cross-linguistically that affects nasal consonants in particular.

After realizing that several distinct phonological rules may, in fact, represent a single underlying mechanism, one would like to obtain a formal notation that captures this observation. In other words, one would need to merge the multiple individual rules into a single rule that will capture their similarity and their differences at the same time. Both rules in (15-7) and (15-6) change the place of articulation of /n/, but each to a different value. Thus, in order to merge these rules, one would need a formal notation that generalizes over different places of articulation. Such type of notation is called an **index** – a variable taking multiple potential values. In the case of place assimilation, we use an index that runs through all possible major places of articulation. As a convention, we use Latin letters (e.g., {i, j}) as indices for unary features like place of articulation. Thus, the indexed feature [place_i] stands for the union of {[Lab], [Cor], [Dor]}:

(15-8) place_i = {[Lab],[Cor],[Dor]}

Consequently, we can use the notion of distinctive features in order to replace the two rules (15-7) and (15-6) by a single rule:

(15-9) n → $[place_i]$ / _$C_{[-son,-cont,placei]}$

(15-9) is an **indexed rule** that is equivalent to the combination of the two rules in (15-7) and (15-6). It reads: "*n* assimilates in place of articulation to the following plosive" (or: "*n* takes the value of the [place] feature of the following plosive). Indexed rules are used mainly with assimilatory processes, i.e., a segment assimilates to another segment in some feature, regardless of the value of the feature. Thus, indexing is a neat formal way to capture common underlying structure and motivation across phonological rules. The representation of the assimilation is further emphasized by the repetition of the indexed term $[place_i]$: it appears once in the output slot (right after the arrow) and once in the environment of the rule. This repetition captures the essence of the assimilation process: a feature of the input segment (e.g., /n/) changes according to the corresponding feature of another segment.

The formal rule (15-9) suggests that the phonological rule does not change the consonant /n/ to another consonant holistically. Rather, the rule changes one specific feature of /n/ – its place of articulation. Thus, if we take /n/, change its place of articulation to dorsal and maintain all the other features, we would get a dorsal nasal (here, a velar nasal). The form of rule (15-9) also demonstrates an important principle of phonological rules: phonological rules should be as economic as possible and exclude any piece of information that can be spared. In this case, the manner of articulation and voice of [n] do not change, so there is no need to specify them in the output (the right side of the arrow).

Finally, note that (15-9) predicts that /n/ takes the place of articulation of the following plosive, regardless of the exact place. Thus, it is a shorthand for the following rules:

(15-10) n → [Lab] / _$C_{[-son, -cont, Lab]}$
n → [Cor] / _$C_{[-son, -cont, Cor]}$
n → [Dor] / _$C_{[-son, -cont, Dor]}$

According to this formulation, /n/ should also assimilate to alveolar plosives ([t, d]). Although not demonstrated explicitly in the data, this prediction can be guaranteed to be true. This is because /n/ itself is alveolar, and remains so when followed by alveolar plosives. The prediction that /n/ also assimilates in place to alveolar plosives demonstrates a "vacuous" application of the place assimilation rule – no surface change is caused by the assimilation of /n/ to alveolar plosives. Though it might seem redundant to include the vacuous case in the formal analysis, doing so is important for capturing generalizations. Excluding the vacuous case from the analysis would hinder our ability to represent place assimilation as a single phenomenon, since it would be harder to write a single formal rule that would include some places of articulation and exclude others.

To demonstrate another instance of generalization with indices, consider the following words in Modern Hebrew. The data provides two alternative pronunciations for each word: the so-called normative pronunciation is typical of "careful" speech, which might be affected by the written form of the words. The "rapid-speech" pronunciation, which is typical of spontaneous speech, is how speakers tend to pronounce the words when they do not pay special attention to articulation. This data demonstrates **free variation** – alternative pronunciations of words by the same speaker or group of speakers.

(15-11) Voicing assimilation in Modern Hebrew (Semitic; Israel)[46]

	Careful pronunciation	**Rapid-speech pronunciation**	
1.	kvut͡sa	gvut͡sa	'group'
2.	misgɛʁɛt	mizgɛʁɛt	'frame'
3.	hɛsbɛʁ	hɛzbɛʁ	'explanation'
4.	χɛʃbon	χɛʒbon	'account'
5.	tguva	dguva	'reaction'
6.	mitbaχ	midbaχ	'kitchen'
7.	zχuχit	sχuχit	'glass'
8.	zkɛna	skɛna	'old.SG.F'
9.	bχina	pχina	'exam'
10.	savta	safta	'grandma'

Assuming the UR of these words is represented by the carful pronunciation, we can identify the following alternations in rapid speech:

	UR (careful speech)		**PR (rapid speech)**	**Environment**	
(1)	k	→	g	#_v	{#,i,ɛ}_{v,g,b}
(2, 3)	s	→	z	i_g, ɛ_b	
(4)	ʃ	→	ʒ	ɛ_b	
(5, 6)	t	→	d	#_g, i_b	
(7, 8)	z	→	s	#_χ, #_k	{#,a} _ {χ,k,t}
(9)	b	→	p	#_χ	
(10)	v	→	f	a_t	

[46] Data compiled by the author.

Examining the data, we can see that the examples involve alternations between voiced and voiceless obstruent counterparts (i.e., [ʒ] is the voiced counter part of [ʃ] – both are postalveolar fricatives and differ only in voice). Moreover, the value change is not the same in all examples: in examples (1-6), a voiceless obstruent in the UR is changed into its voiced counterpart in the PR, while in (7-10), a voiced obstruent is changed into its voiceless counterpart.

To account for these alternations, we need to examine the environments in which they occur. The left-side environment in all the examples is either a vowel or the beginning of the word, while the right-side environment always contains an obstruent. Moreover, on close examination, we can see that the voice alternation is predictable: when the right-side environment contains a voiced obstruent, we get a voiced alternant and, when the right-side environment contains a voiceless obstruent, we get a voiceless alternant. Thus, the voiceless /s/ changes into the voiced [z] when followed by the voiced [b] and [g] (words 2 and 3), while /z/ changes into [s] when followed by the voiceless [χ] and [k] (word 7 and 8).

Examining the evidence, we can conclude that the [voice] feature of realized obstruents tends to match the [voice] feature of the following obstruents. The process of changing the voice of an underlying consonant according to the voice of an adjacent consonant is called **voicing assimilation**, i.e., a consonant assimilates in voice to a neighboring consonant. We can write this formally using the following rules:

(15-12) i. $C_{[-son,-voice]} \rightarrow [+voice]\ /\ \_C_{[-son,+voice]}$

ii. $C_{[-son,+voice]} \rightarrow [-voice]\ /\ \_C_{[-son,-voice]}$

Rule(15-12)i describes the change of an underlying voiceless obstruent into a voiced obstruent before a voiced obstruent, in examples (1-6), while Rule (15-12)ii describes the change of an underlying voiced obstruent into a voiceless obstruent before a voiceless obstruent, in examples 7-10. There are also cases of vacuous application, that is, when the [voice] of two consecutive consonants matches in the UR, it remains unchanged in the PR (e.g., /tfila/ → [tfila] ‘prayer’). Thus, the rules (15-12) can be further revised by excluding the [voice] feature from the input to the rules.

(15-13) i. $C_{[-son]} \rightarrow [+voice]\ /\ \_C_{[-son,+voice]}$

ii. $C_{[-son]} \rightarrow [-voice]\ /\ \_C_{[-son,-voice]}$

According to (15-13)i, an obstruent surfaces as [+voice] when followed by a [+voice] obstruent, regardless of its underlying [voice], while (15-13)ii states that an obstruent surfaces as [-voice] when followed by a [-voice] obstruent, regardless of its underlying [voice].

Thus, we obtained two formal rules that account for the alternations in the data. However, the definition of voicing assimilation provided above describes essentially a single process, whereby a consonant simply “takes” the [voice] of the following consonant, regardless of its valence. Consequently, we opt for a single formal rule to describe the process of voicing assimilation. For that purpose, we need to use an indexed feature, similarly to what we used in the analysis of place assimilation. In particular, we use Greek letters {α, β, γ} in phonological rules to represent binary features when the rule applies equally to both “+” and “-”values. The indexed notation allows us to collapse the two rules in (15-13) to a single rule:

(15-14) $C_{[\text{-son}]} \rightarrow [\alpha\text{voice}]$ / _ $C_{[\text{-son},\alpha\text{voice}]}$

(15-14) can be read as: “an obstruent takes the [voice] of the following obstruent”, where the index α generalizes over all the possible values of the [voice] feature, that is α = {+, -}.

15.6. To generalize or not to generalize: issues in phonological rules

After formulating a phonological rule it is important to check its accuracy and verify that it does not make any wrong predictions. For dataset (15-11), rule (15-14) predicts that all sequences of obstruents should agree in [voice] in the PR. If counterexamples are found, the rule must be revised.

There are several types of possible errors in phonological analysis. The most obvious type of error is arriving at irrelevant conclusions (e.g., concluding that the change observed in the Hebrew data is in manner of articulation, or that the process is word-initial voicing rather than voicing assimilation). To detect such errors one simply needs to go over the data and look for counterexamples for the analysis. Other types of errors might be more subtle and harder to detect. They include: overgeneralization and undergeneralization.

Overgeneralization leads to the prediction that a phonological process should apply to cases in which it does not apply. Consider the following alternative analysis of dataset (15-11). One may propose that every two adjacent consonants agree in [voice], and formulate rule (15-15) as an alternative to (15-14). This is, in fact, consistent with the data.

(15-15) $C \rightarrow [\alpha\text{voice}] / \_ C_{[\alpha\text{voice}]}$

However, dataset (15-16) contains sequences of consonants with conflicting voicing. Words (11-13) contain sequences of voiceless + voiced consonants (i.e., [tl], [sm], and [ʃn]), while words (14-16) contain sequences of voiced + voiceless consonants (i.e., [mt], [nt], and [lχ]).

(15-16) Lack of voicing assimilation in Modern Hebrew (Semitic; Israel)[47]

	Careful pronunciation	**Rapid-speech pronunciation**	
11.	mitlɛ	mitlɛ	‘hanger’
12.	smiχa	smiχa	‘blanket’
13.	ʃnija	ʃnija	‘second’
14.	mamtak	mamtak	‘candy’
15.	psantɛʁ	psantɛʁ	‘piano’
16.	halχana	halχana	‘composition’

These are obviously counterexamples to rule (15-15), which means that the rule is too general and needs to be revised. In order to correct the rule, we need to examine the involved segments more closely. All the alternating consonants in words (1-10) are plosives and fricatives. In addition, the consonants to which the alternating consonants assimilate are also plosives and fricative. On the other hand, the consonants that fail to trigger or undergo voicing assimilation in words (11-16) are nasals and lateral approximants (i.e., [m], [n], and [l]).

Recall that plosives and fricatives are part of the obstruents class, while nasals and lateral approximants are part of the sonorants class (see Section 15.2). Thus, it seems that in the examined data, voicing assimilation is an internal interaction among obstruents (in fact, obstruents tend to be affected by voicing assimilation more than other classes of segments, cross-linguistically). Therefore, rule (15-15) is overgeneralizing and should be rejected in favor of rule (15-14). The overgeneralization discussed here could only be detected by considering additional data. However, one may mistakenly arrive at such incorrect conclusions even when all data is available.

The opposite of overgeneralization is undergeneralization, namely failing to capture the full scope of a phenomenon. Other than missing relevant examples in the data, undergeneralization may result from over-conservative analysis that misses “the bigger

[47] Data compiled by the author.

picture". Consider, again, dataset (15-11) and rule (15-14). The analysis predicts that every two adjacent obstruents should agree in [voice]. However, not all possible examples of this voicing assimilation are demonstrated in the data. The data does not contain examples of voicing of /p/, devoicing of /d/, voicing of /k/ before consonants other than /v/, and so on. In light of this sparse data, one may be reluctant to accept the generalization proposed in (15-14) and, instead, opt for a list of rules representing only the cases demonstrated in the data. While this approach is valid, in principle, it offers very little insight. Even with the original sparse data, there is enough information to conclude that there is some general pattern beyond the individual cases. Yet, finding the most accurate generalization is not always easy.

In general, when attempting to generalize phonological rules, there are two possibilities, each following a different philosophical approach. According to the "conservative" approach, one should make the most general rule that is <u>supported</u> by the data. In other words, we should not generalize the rule if the data does not allow us to test the revised rule directly. Accordingly, rule (15-14) should be rejected, since not all its predictions can be tested with the current data.

By contrast, according to the more "permissive" approach, we should make the most general rule that is <u>not contradicted</u> by the data. In other words, we can make a rule as general as we want, as long as it does not generate incorrect forms (compared to the data we have). This approach allows us to generalize rules for forms that are not observed in the data.

Which approach is better? In principle, both are legitimate, since we want to have the most general rules but, at the same time, we don't want to over-generalize. If we have a large enough dataset, we should be able to capture the correct generalization without much speculation. The problem of over- vs. under-generalization arises mainly in small datasets that force us to some conjecture. There is no fool-proof way to decide which approach is better in a given situation, but there are some points worth considering.

First, remember that rules can apply only to individual segments and natural classes. In other words, it is not possible to write a rule that applies to a set of segments that do not form a natural class. In the present case, we cannot write a voicing assimilation rule that applies to {k, s, ʃ, t, z, b, v}, but not to other obstruents, since these seven consonants do not form a natural class (there are no phonological features that are shared only by these consonants in Hebrew). Thus, the only way to write a less-general form of rule (15-14) is to write a list of specific rules for each individual sequence of segments. However, while this would be technically accurate, as said above, we have good reasons to assume it does not capture the whole picture and, therefore, should not be maintained.

Second, it is safe to take the permissive approach when dealing with phonological processes that are common cross-linguistically. Voicing assimilation is a well-known phenomenon, so we can assume that it applies to the entire class of obstruents and not just to an "arbitrary" subset of a class. Of course, one should gain some experience to be able to predict the probable form of phonological processes.

The conservative approach is preferable when we can find a generalization that is directly supported by the data, while we lack direct evidence for a more general approach. As we saw, rules applying to obstruents do not necessarily apply to sonorous consonants as well. Thus, assuming the highest possible level of generalization is not always accurate.

Either way, phonological rules should not be expected to apply to phonemes that are not part of the language (e.g., Hebrew does not have palatal plosives like [c]), so there is no need to consider them when writing up the rule. And remember, whatever approach you take to analyzing phonological data, make sure the rules you formulate do not generate forms that contradict the given data.

15.7. Practice

1. For each group of segments, determine if it forms a natural class in English. If yes – characterize the class; if no – explain or give an example for a missing member of a potential class.
   i. b, d, g
   ii. p, k
   iii. s, j, æ
   iv. t͡ʃ, d͡ʒ

Solution:

   i. Yes. Voiced plosives.
   ii. No. They are voiceless plosives, but so is [t].
   iii. No. They don't share any features.
   iv. Yes. Affricates.

2. Describe the following formal rules in words using the notation of distinctive features:
   i. V → [+high] / _ #

Solution: “A vowel is raised (=becomes high) at the end of the word.”

This process is *final vowel raising*. Consider, for example, the following nouns in the singular nominative and partitive cases in Finnish (Odden, 2013):

	NOM.SG	**PART.SG**	
1.	englanti	englanti-a	‘England’
2.	joki	joke-a	‘river’

The data shows that the partitive case is -a. In addition, there is an alternation in the second vowel in *river*. Analyzing data suggests that /e/ is raised to [i] in word-final position. Thus, the UR of [joki] ‘river.NOM.SG’ is: /joke/.

ii. $V \rightarrow [+nas] / C_{[+nas]} \_ C_{[+nas]}$

Solution: “a vowel becomes nasal (nasalized) between nasal consonants.”

An example for this vowel nasalization can be found in spontaneous pronunciation of Hebrew words like /ʁ̞imon/ ‘pomegranate’, in which the vowel [o] enclosed between nasal consonants can become nasal: [ʁ̞imõn].

3. Write a formal rule for the following descriptions of phonological processes using distinctive features:
   i. “An unrounded, non-low vowel becomes high and non-back at the end of the word.” (=final vowel fronting/raising)

Solution: $V_{[-round,-low]} \rightarrow [+high,-back] / \_ \#$

An example for final vowel fronting/raising is found in (Uralic; Russia) (Odden, 2013):

	NOM	**PART.SG**	
1.	jałka	jałka-a	‘foot’
2.	jarvi	jarvə-a	‘lake’

The data shows that the partitive case is -a. In addition, there is an alternation in the second vowel in *lake*. Analyzing data suggests that /ə/ is raised and fronted to [i] in word-final position. Thus, the UR of [jarvi] ‘lake.NOM’ is: /jarvə/.

ii. “A consonant becomes voiced after a nasal consonant.”

<u>Solution</u>: Rule: $C \rightarrow [+\text{voice}] / C_{[+\text{nas}]}\ \_$

An example for this voicing is found in (Kenya) (Odden, 2013). Consider three conjugations of the verb ‘buy’: [kula] ‘IMP’ – [βa-kula] ‘3PL.PRS’ – [ŋ-gula] ‘1SG.PRS’. The UR of [ŋ-gula] ‘I buy’ is /ŋ-kula/, where the UR of the base form of ‘buy’ is /kula/, and the first consonant /k/ becomes voiced under the influence of the 1SG.PRS prefix /ŋ/.

4. **Tiberian Hebrew** (Semitic)

Each row in the table below specifies two verb conjugations (change of tense, person, etc.) derived from the same consonantal root in Tiberian Hebrew. Each conjugation involves a morphological change (vowel change, prefixation). In addition, there are also phonological changes in the boldfaced root consonants. In this exercise we will analyze the process responsible for the phonological changes. The exercise focuses on six pairs of consonants: [p]-[ɸ], [b]–[β], [t̪]–[θ], [d̪]-[ð], [k]–[x], and [g]-[ɣ].[48]

		A		B	
1.	‘turn’	**p**ɔnɔ	(3SG.M.PST)	ji**ɸ**nu	(3PL.M.FUT)
2.	‘count’	lis**p**or	(INF)	so**ɸ**er	(3SG.M.PRS)
3.	‘dig’	laħ**p**or	(INF)	ħɔ**ɸ**ar	(3SG.M.PST)
4.	‘fly’	ʕu**ɸ**	(2SG.M.IMP)	ʕɔ**ɸ**oθ	(3PL.F.PRS)
5.	‘hang’	**t̪**ɔlɔ	(3SG.M.PST)	ji**θ**lu	(3PL.M.FUT)
6.	‘seal’	lis**t̪**om	(INF)	so**θ**em	(3SG.M.PRS)
7.	‘sign’	jaħ**t̪**om	(3SG.M.FUT)	ħo**θ**moθ	(3PL.F.PRS)
8.	‘drink’	liʃ**t̪**oθ	(INF)	ʃo**θ**oθ	(3PL.F.PRS)
9.	‘mine’	**k**ɔrɔ	(3SG.M.PST)	ji**x**ru	(3PL.M.FUT)
10.	‘hire’	lis**k**or	(INF)	so**x**er	(3SG.M.PRS)
11.	‘sell’	jim**k**or	(3SG.M.FUT)	mo**x**er	(3SG.M.PRS)
12.	‘win’	liz**k**oθ	(INF)	zo**x**oθ	(3PL.F.PRS)
13.	‘create’	**b**ɔrɔ	(3SG.M.PST)	ji**β**rɔ	(3SG.FUT)
14.	‘break’	niʃ**b**or	(1PL.FUT)	ʃɔ**β**ar	(3SG.M.PST)
15.	‘demand’	**d̪**ɔraʃ	(3SG.M.PST)	ji**ð**roʃ	(3SG.M.FUT)
16.	‘tyrannize’	t̪ir**d̪**ɛ	(3SG.F.FUT /.2SG.M.FUT)	ro**ð**ɛ	(M.PRS)
17.	‘wean’	**g**ɔmal	(3SG.M.PST)	t̪i**ɣ**mol	(3SG.F.FUT/2SG.M.FUT)
18.	‘err’	jiʃ**g**ɛ	(3SG.M.FUT)	ʃɔ**ɣ**a	(3SG.M.PST)

[48] Due to the absence of data from living native speakers of Biblical Hebrew, there is some uncertainty regarding the exact pronunciation of some consonants (see for example Blau, 2010). In such cases, I preferred to use reconstructed forms that would make the analysis neater (e.g., using bilabial fricatives instead of labiodental fricatives and dental plosives instead of alveolar plosives).

A. Fill in the consonant and vowel charts according to the data. Make sure to correctly position pairs of consonants with respect to voice (Note: the charts do not represent the full phonetic inventory of Tiberian Hebrew).

	Bilabial	Dental	Alveolar	Postalveolar	Palatal	Velar	Pharyngeal	
Plosive								
Nasal								
Trill								
Fricative								
Approximant								
Lateral approximant								

	Front	Back
Close		
Close-mid		
Mid		
Open-mid		
Open		

B. List all the segments (from the data) that belong to the following natural classes:

i. $C_{[+son]}$:

ii. $C_{[+son,-cont]}$:

iii. $C_{[-son,-cont]}$:

iv. $C_{[-son,-cont,-voice]}$:

v. $C_{[-son,+cont]}$:

vi. Velar consonants:

vii. $V_{[+back,+round]}$:

C. For each of the boldfaced consonants in the data, list the phonetic environments it appears in. Indicate the indices of the words containing each environment. If a consonant is found in the same environment in several words, list the environment once.

D. Examine the following pairs of consonants: [p]-[ɸ], [b]–[β], [t̪]–[θ], [d̪]-[ð], [k]–[x], and [g]-[ɣ]. Are the consonants in each pair distinct phonemes or allophones of the same phoneme in Tiberian Hebrew? Provide evidence to support your decision:
   (i) If they are distinct phonemes: provide minimal pairs or near minimal pairs for them, and specify a phonetic environment in which both consonants appear.
   (ii) If they are allophones of the same phoneme: (a) give a motivation to justify your decision (why you decided the consonants are linked to each other), (b) determine which of them is the general allophone and which is the restricted allophone, and (c) describe the phonetic environment in which each of the allophone appears.

E. If you concluded that the pairs of consonants discussed above are allophones:
   (i) Write a formal rule to account for the realization of each allophone.
   (ii) Generalize the rules you wrote above to a single rule, using natural classes and distinctive features. Write a formal rule and describe it in words.

F. Find the underlying representations (UR) of (1B) [jiɸnu] 'turn.3PL.M.FUT', (6B) [soθem] 'seal.3SG.M.PRS', and (11B) [moxer] 'sell.3SG.M.PRS'. Show how the phonetic representations (PR) are derived from the UR.

<u>Solution</u>

A. Consonant chart:

	Bilabial	Dental	Alveolar	Postalveolar	Palatal	Velar	Pharyngeal
Plosive	p b	t̪ d̪				k g	
Nasal	m		n				
Trill			r				
Fricative	ɸ β	θ ð	s z	ʃ		x ɣ	ħ ʕ
Approximant					j		
Lateral approximant			l				

Vowel chart:

	Front	Back
Close	i	u
Close-mid	e	o
Open-mid	ɛ	ɔ
Open	a	

B. Natural classes:

i. $C_{[+son]}$: m, n, r, l, j

ii. $C_{[+son,-cont]}$: m, n

iii. $C_{[-son,-cont]}$: p, b, t̪, d̪, k, g

iv. $C_{[-son,-cont,-voice]}$: p, t̪, k

v. $C_{[-son,+cont]}$: ɸ, β, θ, ð, s, z, ʃ, x, ɣ, ħ, ʕ

vi. Velar consonants: k, x, g, ɣ

vii. $V_{[+back,+round]}$: u, o, ɔ

C. Consonants of interest and their phonetic environments.

p	
(1A)	#_ɔ
(2A)	s_o
(3A)	ħ_o

ɸ	
(1B)	i_n
(2B)	o_e
(3B)	ɔ_a
(4A)	u_#
(4B)	ɔ_o

b	
(13A)	#_ɔ
(14A)	ʃ_o

β	
(13B)	i_r
(14B)	ɔ_a

t̪	
(5A)	#_ɔ
(6A)	s_o
(7A)	ħ_o
(8A)	ʃ_o

θ	
(5B)	i_l
(6B)	o_e
(7B)	o_m
(8B)	o_o
(4B,7B,8A,8B, 12A,12B)	o_#

d̪	
(15A)	#_ɔ
(16A)	r_ɛ

ð	
(15B)	i_r
(16B)	o_ɛ

k	
(9A)	#_ ɔ
(10A)	s_o
(11A)	m_o
(12A)	z_o

x	
(9B)	i_r
(10B, 11B)	o_e
(12B)	o_o

g	
(17A)	#_ɔ
(18A)	ʃ_o

ɣ	
(17B)	i_m
(18B)	ɔ_a

D. It is evident that there are no minimal pairs in any of the tables above. The consonants in all the pairs are allophones of each other: we find alternations between the consonants in words derived from the same root (semantically and morphologically related words). In addition, they have complementary distributions on their left side – the plosive in each pair appears at the beginning of word or after a consonant, while the fricatives appear only after vowels. Thus, according to the data, the six plosives (p, b, t̪, d̪, k, and ɡ) are phonemes in Tiberian Hebrew, and the fricatives are their restricted allophones realized after vowels (ɸ, β, θ, ð, x, and ɣ, respectively).

E. (i) Formal rule for each pair of allophones:

- [p]-[ɸ]: p → ɸ / V_
- [b]–[β]: b → β / V_
- [t̪]–[θ]: t̪ → θ / V_
- [d̪]-[ð]: d̪ → ð / V_
- [k]–[x]: k → x / V_
- [ɡ]-[ɣ]: ɡ → ɣ / V_

(ii) Generalization: the observed transformations involve a change in manner of articulation, while the place of articulation and voice are preserved. The observed change is from a plosive to a fricative and, according to Section 15.2.1, it can be formulated as a change in the [continuant] feature from [-cont] (plosives involve discontinuous airflow through the oral cavity) to [+cont] (fricatives involve continuous airflow through the oral cavity). Thus, we can rewrite the rules above in a more transparent way:

- Formal rule: $C_{[\text{-son,-cont}]} \rightarrow [\text{+cont}]$ / V_
- Description: an obstruent (i.e., [-son]) non-continuant (i.e., [-cont]) consonant becomes continuant after a vowel.

The phonological process responsible for these alternations in called **spirantization**.

The generalized rule is more economic than the individual rules formulated in (i), but it can be made even more economic. Note that the input to the rule contains the feature [-cont] while the output contains the opposite value, namely [-cont]. This formulation is somewhat redundant. The rule changes the [cont] value of [-son] consonants that follow a vowel to [+cont]. Without any evidence to the contrary, we can assume that a [+cont, -son] consonant (i.e., a fricative) after a vowel will remain [+cont]. Thus, as long as the input is a [-son] consonant, the output will be [+cont], regardless of the [cont] value of the input. Therefore, it is not necessary to specify the [cont] value of the target and, consequently, we can revise the analysis as follows:

- Formal rule: $C_{[-son]} \rightarrow [+cont] / V\_$
- Description: an obstruent becomes continuant after a vowel.

F. Underlying representations:

UR	/jipnu/
Spirantization	jiɸnu
PR	[jiɸnu]

UR	/sot̪em/
Spirantization	soθem
PR	[soθem]

UR	/moker/
Spirantization	moxer
PR	[moxer]

Chapter 16 Morphophonemics

In the previous chapters, we discussed the concepts of phonemes and allophones, assuming that phonemes are mental representations of sounds and that allophones are various realizations of phonemes. The most compelling evidence for allophonic relations is obtained when the connection between the allophones is clear. This is usually the case when morphology is involved.

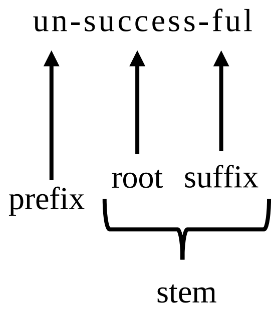
16.1. Basic morphology

In many languages, words can be divided into smaller units of meaning, or **morphemes**. In such languages, words can be either **monomorphemic** (i.e., composed of a single morpheme) or **polymorphemic** (i.e., composed of multiple morphemes). Examples for monomorphemic words include: *cat*, *walk*, *the*. These words are also called **free (unbound) morphemes** since they can be used as stand-alone words.

Morpheme: the minimal linguistic unit that carries meaning (i.e., cannot be divided into smaller meaningful units).

Polymorphemic words are often constructed by a process of **affixation**, in which **affixes** are concatenated to some base. Affixes are morphemes that have a grammatical role. They are also called **bound morphemes**, since they cannot be used as standalone words. The most common types of affixes are prefixes and suffixes. A **prefix** is attached in front of its base (e.g., *un-* is attached as a prefix to the base *real* to create *unreal*). A **suffix** is attached to the end of its base (e.g., *-ful* is attached as a suffix to the base *success* to create *successful*).

Any word (monomorphemic or polymorphemic) to which an affix is added serves as a **base** for affixation. A monomorphemic base (i.e., a base that cannot be decomposed further) is called a **root** (or **root word**). For example, *success* is a root word that forms the base for *successful*. A base which, in itself, is composed of several morphemes is called a **stem**. For example, the word *unsuccessful* is created by adding the prefix *un-* to the base word *successful*. In this case, *successful* is referred to as a stem and not as a root, because it can be further decomposed into *success* + *-ful*.



Prefixes and suffixes are added to their bases in a linear fashion, such that the integrity of both bases and affixes is maintained. In other types of affixation, the linear integrity of these elements is compromised to some degree. An **infix** is an affix inserted inside the base, thereby splitting it in two. Infixes are common in Austronesian languages. For example, Tagalog (the main language of the Philippines) uses an agentive *-um-* infix, as in *bilih* ‘to buy’ → *b-um-ilih* ‘X buys/bought’ (French, 1988). A **circumfix** is an affix consisting of a prefix and a suffix. In other words, a circumfix is a two-part affix that is split by the base. The German verbal past participle *ge- -t*, attached to a verbal root, is an example of a circumfix, as in *spiel* [ʃpiːl] → *gespielt* [gə-ʃpiːl-t] ‘play’ (de Lacy, 2007).

Both infixed and circumfixed words involve some entangling of the bases and affixes. A more extreme case of such entangling the **nonconcatenative morphology**, common in Semitic languages, such as Hebrew and Arabic. Such a system, also known as **root-and-pattern morphology**, involves non-linear interleaving of a **consonantal root** and a **morphological pattern**. The consonantal root is an abstract unit composed entirely of consonants (typically three), which defines some general concept. For example, the root *ktb* is related to the general concept of writing in Arabic.[49] The morphological pattern consists of one or more specified vowels and, occasionally, some specified consonants, in fixed positions. It represents various grammatical and lexical features, such as tense, aspect, number, and part-of-speech. The specified elements of the morphological pattern are interleaved with empty “C-slots”, reserved for the consonantal root. For example, in Classical Arabic, the *CaCaCa* pattern expresses past tense, singular, masculine, third-person verbs in the first **Binyan**[50] (e.g., [KaTaBa] ‘he wrote’),

[49] In functional terms, the notion of a consonantal root is similar to that of a root word – both refer to a basic unit of meaning. The difference is that the Semitic consonantal root (e.g., *ktb*) is not a standalone word, as opposed root words (e.g., *cat*).

[50] Verbs in Semitic languages are classified into one of several derivational classes called *binyanim* (Hebrew, lit.: ‘buildings’, singular form: *binyan*).

while the *maCCuːC* pattern expresses the passive participle form of verbs in the first Binyan (e.g., [maKTuːB] 'written') (McCarthy, 1981).[51]

Another type of morphological structure involves **reduplication**, that is, repetition of some part of the word. Reduplication is used in many languages and is especially common in Austronesian languages. For example, in Ilokano (Philippines), repetition of word-initial consonant-vowel-consonant sequence in nouns marks plurality (e.g., [púsa] 'cat' → [pus-púsa] 'cats'; [kaldíŋ] 'goat' → [kal-kaldíŋ] 'goats') (Hayes & Abad, 1989). The **reduplicant** (i.e., repeated part) is connected linearly to its base, similarly to prefixes and suffixes. However, its segmental content is not fully specified, since it changes from word to word (e.g., [pus] in [pus-púsa] and [kal] in [kal-kaldíŋ]). In this respect, reduplication is similar to root-and-pattern morphology, which also has underspecified elements (C-slots).

In addition to classifying polymorphemic words by the type of morphological structure, one can also distinguish such words on functional grounds. Morphology can be either **inflectional** or **derivational**. Inflectional morphemes specify information like number, gender, and tense. For example, the word *walks* is composed of the root *walk*, which gives the word its basic meaning, and the inflectional suffix *–s*, which specifies present tense (or plural, if *walk* is analyzed as a noun). Derivational morphemes create new words in the language. For example, the word *lockable* is composed of the root *lock* and the derivational suffix *–able*, which generates an adjective that specifies a quality related to locking (an entity described as *lockable* is something that can be locked). The distinction between inflectional and derivational morphology is not crucial for the purpose of this book.

16.2. Morphophonemic analysis

When polymorphemic words are created, the pronunciation of the individual morphemes can be affected by the interaction with the other morphemes in the word. The study of the phonological interactions between morphemes (or the interaction between morphological processes and phonological processes) is called **morphophonemics**, or **morphophonology**.[52] The following subsections demonstrate the application of phonological processes at various morphological boundaries.

[51] Uppercase letters in the phonetic transcriptions represent root consonants.

[52] This chapter focuses on interactions involving linear affixes (e.g., prefixes and suffixes). Analyzing the phonological consequences of other morphological structures is beyond the scope of this book.

16.2.1. **Alternations in affixes**

To demonstrate the process of morphophonemic analysis, consider the following data featuring English singular and plural nouns. The data in (16-1) is organized by **morphological paradigm**. A morphological paradigm is a list of words with the same underlying morphological structure. Thus, the nouns in the singular are presented in one column, and the corresponding plural nouns are presented in another column.

(16-1) English plural nouns

	Singular	Plural	
1.	kæt	kæts	‘cat’
2.	ɹuːf	ɹuːfs	‘roof’
3.	dʌk	dʌks	‘duck’
4.	ʃɑːp	ʃɑːps	‘shop’
5.	dɔɡ	dɔɡz	‘dog’
6.	ɡɹuːv	ɡɹuːvz	‘groove’
7.	sneɪl	sneɪlz	‘snail’
8.	hɛd	hɛdz	‘head’
9.	ʃuː	ʃuːz	‘shoe’
10.	bʌs	bʌsəz	‘bus’
11.	wɪt͡ʃ	wɪt͡ʃəz	‘witch’
12.	noʊz	noʊzəz	‘nose’
13.	d͡ʒʌd͡ʒ	d͡ʒʌd͡ʒəz	‘judge’

In line with previous chapters, we assume that every morpheme has an underlying phonological representation (UR). This is true for both lexical roots and affixes. The aim of **morphophonemic analysis** is to find these underlying representations based on the given data, and to formulate phonological rules that account for the variation between the hypothesized and the given forms. When analyzing data in morphological paradigms, we always begin with finding the UR of the affixes. In normal datasets, the number of affix forms is usually much smaller than the number of different bases, so it is much easier to determine the UR of an affix before determining the UR of all the bases than the other way around.[53]

To find the UR of an affix, we begin by scanning all the words in the relevant column, looking for a segment or a sequence of segments that are repeated throughout the column. Since affixes can display some variations, we also compare the forms of bases with and without an affix (or with different affixes), in order to determine the plausible boundaries of the affix.

[53] In the *planned* Chapter 22 (“Rule application and interaction”), we will analyze datasets in which phonological alternations are observed both in bases and affixes. In such cases, it is crucial to analyze the alternations in the affixes before analyzing the alternations in the bases.

In order to establish the UR of the plural suffix in (16-1), we note that all the words in the plural end in one of three ways: [-s] (words 1-4), [-z] (words 5-9), and [-əz] (words 10-13). We call such alternative pronunciations of a morpheme, **allomorphs**, in analogy with allophones. To make sure we identified the allomorphs correctly, we compare the singular and plural forms of nouns and verify that there are no segments left unassigned to a base or an affix. Indeed, we see that all plural nouns are composed of the singular form plus one of the plural allomorphs.

Following the logic of phonemic analysis (see Section 14.2), we list the environments in which each of the allomorphs appear.

[-s]		[-z]		[-əz]	
(1)	t_#	(5)	g_#	(10)	s_#
(2)	f_#	(6)	v_#	(11)	t͡ʃ_#
(3)	k_#	(7)	l_#	(12)	z_#
(4)	p_#	(8)	d_#	(13)	d͡ʒ_#
		(9)	uː_#		

Since we are dealing with a suffix, it always appears at the end of the word. Consequently, the right-side environment is the same in all cases and, thus, irrelevant for determining the UR of the suffix. Examining the left-side environment, we see that [-s] and [-əz] appear only after consonants, while [-z] appears after both consonants and vowels. In other words, [-z] appears in the most general environment of the three allomorphs. This suggests that the UR of the plural suffix is /-z/, with [-z] as the general allomorph and [-s] and [-əz] as the restricted allomorphs. This is illustrated in Figure 16.1:

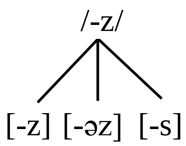


Figure 16.1 The allomorphs of the English plural suffix

It is noteworthy that the conclusion from the phonological analysis is at odds with the written form of the plural suffix was contains ⟨s⟩ and not ⟨z⟩! Indeed, phonological judgments can be shaped by both spontaneous speech perception and formal language education.

After establishing the UR of the plural suffix we can determine the UR of the pluralized words in (16-1). In general, the UR of an affixed word is the concatenation of the UR of the base and the UR of the affix (in the correct order!), with hyphens separating the morphemes. The UR is a transparent representation of the morphological structure of the word, before any phonological changes. Thus, the UR of [kæts] 'cats' is /kæt-z/ (strangely as it may seem…), the UR of [dɔgz] 'dogs' is /dɔg-z/, and the UR of [bʌsəz] 'buses' is /bʌs-z/.

Next, we would like to identify the environments that trigger the realization of /-z/ as one of its restricted allomorphs. [-s] appears after plosives and fricatives, but so does [-z]. Crucially, [-s] appears only after voiceless consonants, while [-z] appears only after voiced consonants. Thus, we can conclude that /-z/ is realized as [-s] after voiceless consonants. However, we see that [-əz] can appear after both voiced (e.g., [z]) and voiceless consonants (e.g., [s]). On closer examination, we notice that [-əz] appears only after alveolar and postalveolar fricatives and affricates, also known as stridents (see Section 15.2.3). Thus, we can conclude the distributions of the allomorphs of the English plural suffix:

(16-2) Allomorphs of the English plural suffix:
[-s]: after non-strident voiceless consonants
[-əz]: after stridents
[-z]: elsewhere

What phonological rules account for the alternations of the plural suffix? The z-s alternation can be formulated as:

(16-3) $z \rightarrow s$ / $C_{[\text{-voice,-strd}]}$ _#

or, by just specifying the changed feature ([z] and [s] differ in [voice]):

(16-4) $z \rightarrow$ [-voice] / $C_{[\text{-voice,-strd}]}$ _#

This notation allows us to see the motivation behind the z-s alternation: [z] becomes voiceless after another voiceless consonant. This is called **voicing assimilation**.

In the z-əz alternation there is no change of segment, but rather an insertion of the vowel [ə]. In formal notation, insertion of a segment (also called **epenthesis**) is written as turning a "nothing" into "something":

(16-5) $Ø \rightarrow ə / C_{[+strd]}\_z$

We use the empty set symbol, Ø, to denote an empty segmental “slot”. In the case of epenthesis, the slot was empty in the UR and was filled-in in the PR (in cases of deletion, as we shall see in the next section, the slot is emptied in the PR). The following diagram illustrates the idea of segmental slots with UR and the PR of *buses*. The top row of the table displays the segments of the UR, and the bottom row displays the corresponding segments of the PR. The rows are identical except for the fourth slot. This slot is occupied by [ə] in the PR. In the UR the slot is empty (the slot does not actually exist in the UR, but including it in the representation can help us understand the process of epenthesis).

	1	2	3	4	5
UR:	b	ʌ	s	Ø	z
PR:	b	ʌ	s	ə	z

What could be the motivation behind the insertion of [ə] between the two consonants? We note that [z] is a strident, so all the examined cases involve [ə] epenthesis between two stridents. Apparently, a sequence of stridents is “hard” to pronounce, and inserting a vowel between the consonants simply breaks apart this hard sequence. In fact, sequences of stridents are so difficult that speakers (not only of English) tend to avoid them altogether (try to think of English words containing letter sequences such as ⟨sz⟩, ⟨zs⟩, and ⟨zsh⟩).

After establishing the phonological rules, we can use them to show how the PR is derived from the UR (morpheme boundaries are marked only in the UR):

cats	
UR:	/kæt-z/
Voicing assimilation	kæts
PR:	[kæts]

buses	
UR:	/bʌs-z/
Vowel epenthesis	bʌsəz
PR:	[bʌsəz]

16.2.2. **Alternations in the base at morpheme boundaries**

The English plural suffix data demonstrates one type of phonological interaction between base words and affixes, namely the fact that alternations in an affix can be induced by segments of the base. However, often, it is not the affix but, rather, the base that shows phonological alternations across morphological paradigms. For example, the following table lists some Karok (or, Karuk) verbs in the imperative (A) and first-person singular (B).

(16-6) **Karok** (Language isolate; California, U.S.A) (*Ararahih'urípih: A Dictionary and Text Corpus of the Karuk Language*, 2023; Bright, 1957)

	(A) Imperative	**(B) 1SG.**	
1.	pasip	nipasip	'shoot'
2.	si:tva	niʃi:tva	'steal'
3.	suprih	niʃuprih	'measure'
4.	vupaksip	nivupaksip	'to start to cut'
5.	xus	nixus	'think, feel, realize'
6.	tatarariʃ	nitatarariʃ	'hold down'
7.	piʃvi:tkar	nipiʃvi:tkar	'tear up'

We can see that the imperative is expressed by a "null" affix, since there is no sequence of segments (or similar segments) that is repeated throughout column A in a fixed position. In addition, the first-person singular is expressed by the prefix /ni-/, since all the words in column B begin with [ni].

In addition, words 2 and 3 demonstrate an alternation between [s] and [ʃ] in the base. An analysis of the distribution of [s] and [ʃ] reveals that both consonants appear before [u] as well as at the end of the word. Therefore, the right-hand side of the environment does not account for the alternation. Examination of the left-side environment reveals that [ʃ] appears only after [i], while [s] appears after various vowels, consonants, as well as at the beginning of the word. Moreover, [s] does not appear after [i].

[s]		[ʃ]	
(1)	a_i	(2B)	i_i:
(2)	#_i:	(3B)	i_u
(3)	#_u	(6)	i_#
(4)	k_i	(7)	i_v
(5)	u_#		

Therefore, the analysis suggests that [s] and [ʃ] are allophones of /s/, and that the phoneme is realized as the restricted allophone [ʃ] after [i]. This can be represented by the following formal rule:

(16-7) s → ʃ / i_

The process demonstrated in the data is called **palatalization**, by which the place of articulation of a non-palatal consonant moves closer to the palate under the influence of a front or high vowel. Here, an alveolar consonant becomes postalveolar after [i] (high, front vowel)). According to the analysis, the UR of (2B) is /ni-siːtva/ 'I steal'. The PR [niʃiːtva] is derived from the UR as follows:

UR:	/ni-siːtva/
Palatalization	niʃiːtva
PR:	[niʃiːtva]

The Karok dataset demonstrates that affixes can induce alternations in the base consonant.

16.2.3. Alternations in the base at word boundaries

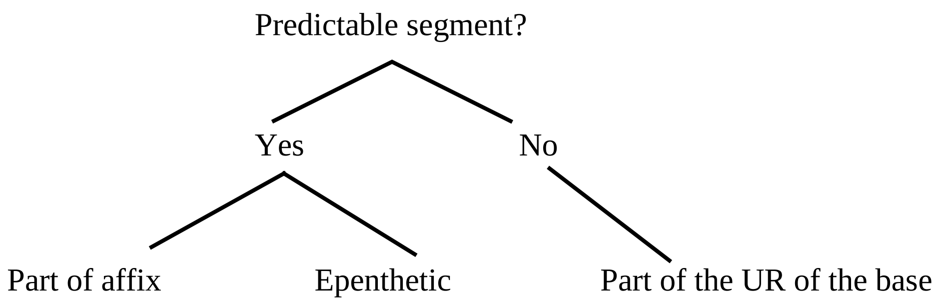
Alternations in the base can make it difficult to determine the underlying form of the base. While it seems intuitive that the bear root is the UR and that alternations occur due to interactions with an affix, it is not always the case. The reason is that, in addition to processes occurring at morpheme boundaries, there are some processes that occur at word boundaries (i.e., affecting the initial or final segment of the word). As a result, the UR of the base is sometimes observed in an affixed form, while it is the unaffixed form that exhibits some phonological process. Thus, when observing alternations in the base, one needs to determine whether they are caused by interactions with a morpheme boundary or with a word boundary. Consider the following data from Samoan:

(16-8) **Samoan** (Austronesian; Samoan Islands) (Odden, 2013)

	Simple	**Perfective**	
1.	tuː	tuːlia	'stand'
2.	tau	taulia	'cost'
3.	ʔalo	ʔalofia	'avoid'
4.	ili	ilifia	'blow'
5.	oso	osofia	'jump'
6.	soa	soaŋia	'have a friend'
7.	asu	asuŋia	'smoke'

Examining the data, we see that there is no sequence that is repeated throughout the verbs in the simple form. Thus, we can conclude that the simple form is "affixless" (or, expressed by a "null" affix). In addition, all the perfective forms end in [ia]. Thus, [ia] is part of the perfective suffix. The [ia] sequence, in turn, is preceded by some consonant that does not appear in the simple form. However, each perfective form exhibits a different consonant ([l], [f], or [ŋ]). Thus, it is not obvious whether these alternating consonants belong to the root, to the perfective suffix, or to neither. In order to account for these elusive consonants, we first examine the features of these segments and their environments. The aim of the analysis is to determine whether these alternations are predictable. If they are predictable, we may conclude that they are either part of the affix, or that they are epenthetic (like the epenthetic [ə] in (16-1)).

Segments can be predictable in two ways. Segments are **morphologically predictable** if they are part of an affix. That is, all or most of the words in a morphological paradigm contain the segment. **Phonologically predictable** segments are usually epenthetic segments – their appearance is predictable by the phonetic environment (like the [ə] inserted between stridents in the English plural suffix). Segments whose appearance cannot be predicted must be part of the underlying representation (UR) of the word in the lexicon. The status of alternating segments is summarized in the following diagram.



If we conclude that the alternations in the column are predictable, we can perform a phonemic analysis (i.e., by listing the environments of each alternating segment) in order to account for the alternations (i.e., finding the phoneme underlying the alternating segments and the rules that account for its allophones). Our data suggests that the alternating consonants in the perfective form do not share any features (place, manner, or voice). In addition, there is nothing in their environments that can account for the alternations (e.g., words #2 and #7 have alternating consonants, [l] and [ŋ], in the same environment u_i).

Therefore, we can conclude that the alternating consonants are neither part of the suffix nor inserted in the perfective form, since we cannot predict their occurrence. Thus, the UR of the perfective suffix is /-ia/, and the alternating consonants are part of the UR of the base (i.e., part of its lexical representation). In other words, the UR of the base is not represented faithfully by the unaffixed roots, since they miss the alternating consonants. Therefore, we can determine the UR of the bases as the perfective form without the perfective suffix. Writing down the hypothesized UR alongside the original data helps visualizing and understanding the observed alternations.

(16-9) Analyzed Samoan verbs

	Simple	Perfective		Base UR
1.	tuː	tuːlia	'stand'	tuːl
2.	tau	taulia	'cost'	taul
3.	ʔalo	ʔalofia	'avoid'	ʔalof
4.	ili	ilifia	'blow'	ilif
5.	oso	osofia	'jump'	osof
6.	soa	soaŋia	'have a friend'	soaŋ
7.	asu	asuŋia	'smoke'	asuŋ

Why are these consonants missing from the simple form? Examining the environments where we would expect to find them, we see that they all appear at the end of the UR. Thus, they would be in a word-final position, had they not been deleted. Moreover, there is no word in the data the ends with a consonant. Apparently, Samoan, like many other languages, avoids consonants at word endings, and when a UR ends with a consonant, it is deleted in the PR. The formal representation of this **consonant deletion** is very similar to that of epenthesis:

(16-10) C → Ø / _#

The rule above says that any consonant becomes "nothing" (i.e., deleted) at the end of a word. The following diagram illustrates the structural difference between the UR and the PR of word #2 [tau] 'cost'. The UR has four segmental slots, the last of which is emptied in the PR, leaving it with only three slots.

	1	2	3	4
UR:	t	a	u	l
PR:	t	a	u	Ø

16.2.4. **Morpheme-specific phonological processes**

The underlying motivation of many phonological processes is phonetic. Thus, even when observed at morpheme boundaries, we may assume that the phonological processes are not triggered by the morpheme boundaries themselves. Yet, some datasets demonstrate phonological processes that seem to apply only to specific affixes. To demonstrate morpheme-specific phonological processes let us return to the data of English adjectives analyzed in (15-5) and repeated here for convenience.

(16-11) English adjectives

	Adjective		**Negated adjective**	
1.	ækjɚət	'accurate'	ɪ**n**ækjɚət	'inaccurate'
2.	əfɛktɪv	'effective'	ɪ**n**əfɛktɪv	'ineffective'
3.	dɪpɛndənt	'dependent'	ɪ**n**dɪpɛndənt	'independent'
4.	səkjʊɻ	'secure'	ɪ**n**səkjʊɻ	'insecure'
5.	vælɪd	'valid'	ɪ**n**vælɪd	'invalid'
6.	fɻiːkwənt	'frequent'	ɪ**n**fɻiːkwənt	'infrequent'
7.	bælənst	'balanced'	ɪ**m**bælənst	'imbalanced'
8.	pɑːsəbəl	'possible'	ɪ**m**pɑːsəbəl	'impossible'
9.	pɝːfɛkt	'perfect'	ɪ**m**pɝːfɛkt	'imperfect'

In Section 15.5, we concluded that the UR of the consonant in the negation prefix is /n/ and that it is realized as [m] when attached to bases beginning with a bilabial plosive. This observation was captured by the formal rule:

(16-12) n → m / _ [bilabial plosive] , or in terms of distinctive features:

(16-13) n → [Lab] / _ $C_{[Lab,\ -cont,\ -son]}$

According to the formulation above, every /n/ in English should become labial when preceding bilabial plosive. However, there are plenty of counterexamples to this prediction, including instances of the *un-* prefix (e.g., *unpleasant* [ʌnplezənt] and *unbelievable* [ʌnbəliːvəbəl]) and even other instances of /n/ following [ɪ] (e.g., *pinball* [pɪnbɔːl] and *input* [ɪnpʊt]). Thus, it is obvious that the analysis proposed in (16-12) and (16-13) is not generally correct. A more precise formal analysis of the data must specify that it only applies to the *in-* prefix. Such formal analysis is given in (16-14).

(16-14) n → [Lab] / #ɪ_+$C_{[Lab,-son,-cont]}$

According to(16-14), /n/ becomes bilabial before bilabial consonants when it is part of the /in-/ prefix. The prefix structure is indicated by the environments of the rule. The left-side environment contains the word boundary symbol followed by [ɪ], to indicate that the process only applies at the beginning of the word. The right-side environment contains the '+' sign, which marks a **morpheme boundary**. This helps establishing that not every instance of the phoneme /n/ is affected by a following bilabial consonant – only those that are part of the /in-/ prefix.[54]

Section 14.3, introduced the general form of phonological rules: A → B / X_Y. This notation can be adjusted to represent processes that apply only to specific affixes. Including the '+' sign indicates a morpheme boundary in the environment of a rule phonological. The position of the '+' relative to the underscore ('_') and the position of the word boundary ('#') differentiate prefixes and suffixes. (16-15) illustrates the formal notation of prefix-specific processes. According to this notation, the UR sequence /XA+Y/ is realized in the PR as [XB+Y], where /XA/ and [XB] represent the prefix in the UR and the PR, respectively. Y is the base segment, assumed to be responsible for the alternations in the prefix. This analysis captures the variations in the /in-/ negation prefix, analyzed above (e.g., /ɪn-pɝːfɛkt/ → [ɪmpɝːfɛkt] 'imperfect'). Here, A represents the UR of the prefix consonant (i.e., /n/), B represents the PR of the prefix consonant (i.e., [m]), X is the initial part of the prefix (i.e., /ɪ/), and Y is the first base consonant (i.e., /p/).

(16-15) A → B / #X _+ Y

(16-16) illustrates the formal notation of suffix-specific processes. According to this notation, the UR sequence /X+AY/ is realized in the PR as [X+BY], where /AY/ and [BY] represent the suffix in the UR and the PR, respectively. X is the base segment, assumed to be responsible for the alternations in the suffix. This analysis captures the variations in the /-z/ plural suffix, analyzed in Section 16.2.1 (e.g., /kæt-z/ → [kæts] 'cats'. Here, A represents the UR of the suffix consonant (i.e., /z/), B represents the PR of the prefix consonant (i.e., [s]), X is the final base consonant (i.e., /t/), and Y represents additional hypothetical segments in the suffix following the alternating consonant. In this example, there are no such segments, thus Y is empty.

[54] It might be argued this rule should also apply to word-initial *in* that originated from the preposition *in* (e.g., *input*, *inbox*), which could be considered a separate morpheme (at least historically). Thus, it might be necessary to specify that the rule applies only to the negation *in*-. However, there is no standard notation for such specification within the theoretical framework used in this book.

(16-16) $A \rightarrow B / X +\_ Y\#$

16.3. Practice

1. **Kolami** (Dravidian; India) (Kenstowicz & Kisseberth, 1973)

	(A) Imperative	**(B) Present**	**(C) Future**	**(D) Past**	**Gloss**
1.	siː	siːatun	siːdatun	siːtan	‘give’
2.	ar	aratun	ardatun	artan	‘weep’
3.	suk	sukatun	sugdatun	suktan	‘wither’
4.	put	putatun	puddatun	puttan	‘cut’
5.	dag	dagatun	dagdatun	daktan	‘cough’
6.	mad	madatun	maddatun	mattan	‘forget’

A. Find the UR of the imperative, present, future, and past affixes.

B. Analyze the phonological alternations in the base consonants.

C. Determine the underlying representations of the bases.

Solution:

A. Imperative: Ø. There is no segment or sequence of segments repeated throughout the column.
Present: /-atun/. All the words in the column end with [atun].
Future: /-datun/. All the words in the column end with [datun].
Past: /-tan/. All the words in the column end with [tan].

B. Words 3-6 exhibit alternations in the voice of base plosives (e.g., the [k]-[g] alternation in (3B) [suk-atun] and (3C) [sug-datun]. Analyzing the distributions of voiced and voiceless plosives indicates that sequences of two plosives always agree in voice. Moreover, since the suffixes do not show any alternations, we can conclude that alternations in the base plosives are due to interactions with the following (i.e., suffix) plosives. We can also rule out potential interaction with word boundaries, since both voiced and voiceless plosives can occur word-finally (e.g., (3A) [suk] vs. (5A) [dag]). Therefore, we can propose the following account for the observed alternations:

Name of the process: Voicing assimilation
Description in words: An obstruent is assimilated in voice to a following obstruent
Formal rule: $C_{[\text{-son}]} \rightarrow [\alpha\text{voice}] / \_C_{[\text{-son}, \alpha\text{voice}]}$

C. The UR of the bases happen to be identical to the imperative forms (though it would be inaccurate to say that the UR IS the imperative):

	Imperative	Base	Gloss
1.	siː	/siː/	‘give’
2.	ar	/ar/	‘weep’
3.	suk	/suk/	‘wither’
4.	put	/put/	‘cut’
5.	dag	/dag/	‘cough’
6.	mad	/mad/	‘forget’

2. **Javanese** (Malay; Indonesia) (Kenstowicz & Kisseberth, 1979)

	(A) Noun	(B) This noun	Gloss
1.	t͡ʃaŋkɪr	t͡ʃaŋkire	‘cup’
2.	t͡ʃukʊr	t͡ʃukure	‘haircut’
3.	ket͡ʃap	ket͡ʃape	‘soy sauce’
4.	kulɪt	kulite	‘skin’
5.	səbap	səbabe	‘reason’
6.	murɪt	muride	‘student’
7.	bedʊk	beduge	‘mosque drum’

A. Find the UR of the ‘noun’ and ‘this noun’ affixes.
B. Analyze the phonological alternations in the bases.
C. Determine the underlying representations of the bases (ignore vowel alternations).

<u>Solution</u>:

A. Noun: Ø. There is no segment or sequence of segments repeated throughout the column.
This noun: /-e/. All the words in the column end with [e].

B. Words 5-7 exhibit alternations in the voice of base plosives (e.g., the [b]-[p] alternation in (5A) [səbap] and (5B) [səbab-e]). Analyzing the distributions of voiced and voiceless plosives indicates that voiced plosives do not occur word-finally. Moreover, both voiced and voiceless plosives occur intervocalically (e.g., (3B) [ket͡ʃape] and (5B) [səbabe]). Therefore, we can propose the following account for the observed alternations:

Name of the process: Final devoicing
Description in words: An obstruent (or plosive) becomes devoiced at the end of the word.
Formal rule: $C_{[\text{-son,-cont}]} \rightarrow [\text{-voice}]$ / _#

C. The UR of the bases:

	(A) Noun	Base	Gloss
1.	t͡ʃaŋkɪr	/t͡ʃaŋkɪr/	‘cup’
2.	t͡ʃukʊr	/t͡ʃukʊr/	‘haircut’
3.	ket͡ʃap	/ket͡ʃap/	‘soy sauce’
4.	kulɪt	/kulɪt/	‘skin’
5.	səbap	/səbab/	‘reason’
6.	murɪt	/murɪd/	‘student’
7.	bedʊk	/bedʊg/	‘mosque drum’

Note that, unlike in the Kolami data, here, the UR of the bases are not always identical to the UR of the unaffixed words (compare 5A and 5B). For that reason, one should remember that the UR of bases is generally not the same as the UR of words with null affixes.

3. **Koasati** (Muskogean; Louisiana, USA) (Odden, 2013)

	Noun	My Noun	
1.	apaht͡ʃá	amapaht͡ʃá	‘shadow’
2.	asikt͡ʃi	amasikt͡ʃi	‘muscle’
3.	ilkanó	amilkanó	‘right side’
4.	ifá	amifá	‘dog’
5.	iskí	amiskí	‘mother’
6.	pat͡ʃokkó:ka	ampat͡ʃokkó:ka	‘chair’
7.	bajá:na	ambajá:na	‘stomach’
8.	towá	antowá	‘onion’
9.	tá:ta	antá:ta	‘father’
10.	kastó	aŋkastó	‘flea’

A. List the allomorphs of the ‘My noun’ prefix and determine the appropriate environment of each allomorph.
B. Determine the underlying representation (UR) of the prefix.
C. Analyze the phonological process responsible for the changes in the prefix.
D. Determine the UR of the possessive forms 4, 8, 10 and show how the PR is derived from the UR.

Solution:

A. The allomorphs of the ‘My noun’ prefix are:
   [am]: before vowels and bilabial consonants.
   [an]: before alveolar consonants ([t]).

[aŋ]: before velar consonants ([k].

B. The UR of the 'My noun' prefix is /am-/: it appears in the most general environment – before vowels and consonants.

C. **Name of the process**: nasal place assimilation

Description in words: a nasal consonant assimilates in place to a following consonant.

Formal rule: $C_{[+nas]} \rightarrow [place_i]$ / _ $C_{[placei]}$

Note: this is the most general rule that can be derived based on the data. Other, less general formulations are also acceptable, including: a specific rule for /m/, a specific rule for a prefix, and a specific rule in which assimilation is triggered only by plosives (or obstruents). Additional data is needed to determine which formulation is the most accurate.

D. Determine the UR of the following possessive forms and show how the PR is derived from the UR:

i. (4) [amifá] 'my dog'

UR	/am-ifá/
Place assimilation	-
PR	[amifá]

Note: the hyphen in the rule application row indicates that the rule did not affect the UR.

ii. (8) [antowá] 'my onion'

UR	/am-towá/
Place assimilation	antowá
PR	[antowá]

iii. (10) [aŋkastó] 'my flea'

UR	/am-kastó/
Place assimilation	aŋkastó
PR	[aŋkastó]

Chapter 17 Common phonological processes

The purpose of this chapter is to review various phonological processes that are common cross-linguistically (although there are many others). I regard this review as a “linguistic trip” and, accordingly, use data from different languages to demonstrate the various processes (the selection of languages is rather arbitrary; the discussed processes are found in many other languages, too).

When possible, I discuss potential motivations underlying the processes. Often, the motivation is grounded in phonetics, i.e., the process stems from issues related to speech planning and perception. Thus, the present chapter forms a direct link between phonetics and phonology – phonological processes often start out as a spontaneous “solution” to a phonetic problem (e.g., of articulation). Initially, it may be used inconsistently by a small group of speakers. But once this solution is adopted and used consistently by the native speaker community, it becomes part of the grammar of the language and gains the status of a phonological rule. From that point on, it will be used productively by the speakers, and the initial phonetic motivation underlying the phonological rule will not be needed to account for the process. The productivity of the phonological rule would be most evident in the formation of new words and the borrowing of words from other languages. As we shall see, the written forms of words sometimes provide good evidence for the productivity of phonological processes.

From a structural point of view, we can divide the various phonological processes to those affecting specific features of segments and those affecting whole segmental units. The first group of processes includes mainly various forms of assimilation, while the second group includes deletion, epenthesis, and metathesis.

17.1. Feature changing processes

In feature changing processes, segments surface with one or more features altered compared to their underlying value. The change could be motivated by an interaction with a neighboring segment, morpheme boundaries or word boundaries.

17.1.1. **Assimilation**

Assimilation is a family of phonological processes, in which one or more features of a segment are modified to match those of a nearby segment. Assimilation processes can be characterized by the direction of assimilation, the features involved, and the degree of assimilation.

17.1.1.1 Voicing assimilation

The following data from Modern Hebrew was analyzed in Sections 15.5 and 15.6. The data provides two alternative pronunciations for each word: "careful" speech, and the "rapid-speech" pronunciation, which is typical of spontaneous speech, which does not pay special attention to articulation.

(17-1) Modern Hebrew (Semitic)[55]

	Careful pronunciation ("UR")	**Rapid-speech pronunciation ("PR")**	
1.	kvut͡sa	gvut͡sa	'group'
2.	misgɛʁɛt	mizgɛʁɛt	'frame'
3.	hɛsbɛʁ	hɛzbɛʁ	'explanation'
4.	χɛʃbon	χɛʒbon	'account'
5.	tguva	dguva	'reaction'
6.	mitbaχ	midbaχ	'kitchen'
7.	zχuχit	sχuχit	'glass'
8.	zkɛna	skɛna	'old.SG.F'
9.	bχina	pχina	'exam'
10.	savta	safta	'grandma'
11.	mitlɛ	mitlɛ	'hanger'
12.	smiχa	smiχa	'blanket'
13.	ʃnija	ʃnija	'second'
14.	mamtak	mamtak	'candy'
15.	psantɛʁ	psantɛʁ	'piano'
16.	halχana	halχana	'composition'

Words (1-10) exhibit alternations in the voicing of plosives and fricatives between the two pronunciations of a word. Assuming the careful pronunciation represents the UR of these words and the rapid-speech pronunciation represent the PR, we can say that the [voice] feature of realized consonants tends to match the [voice] feature of the following consonant. These are

[55] Data compiled by the author.

examples of **voicing assimilation**, a process by which a consonant assimilates in voice to a neighboring consonant. This process typically involves obstruent consonants and, words (11-16) in (17-1) confirm that sonorants neither undergo nor trigger such assimilation. Therefore, we can write the following formal rule for the data above:

(17-2) $C_{[\text{-son}]} \rightarrow [\alpha\text{voice}] \ / \ \_C_{[\text{-son},\ \alpha\text{voice}]}$

Rule (17-2) states that an obstruent takes the [voice] of the following obstruent, regardless of its valence. The **indexed** feature [αvoice] stands for the union {[+voice], [-voice]}. Thus, (17-2) is equivalent to the union of the rules in (17-3):

(17-3) i. $C_{[\text{-son}]} \rightarrow [+\text{voice}] \ / \ \_C_{[\text{-son},\ +\text{voice}]}$

ii. $C_{[\text{-son}]} \rightarrow [-\text{voice}] \ / \ \_C_{[\text{-son},\ -\text{voice}]}$

Voicing assimilation is a very common process in spoken Hebrew, and some of the example in (17-1) are classic textbook examples used by language teachers in Israel. Evidence for the psychological reality of the process can also be seen sometimes in writing, especially in children during their first schooling years, who tend to use spelling that reflects the pronunciation of words. The following sign placed over a bakery in Israel shows deliberate misspelling of the word סבתא /savta/ ‘grandma’, in which the letter ב was replaced by the letter פ (in cursive Hebrew: פ). This spelling represents the common spoken form of the word [safta], where the underlying consonant /v/ assimilates in voice to the following /t/.

Figure 17.1 Voicing assimilation in Hebrew. The sign in the picture, placed over a bakery, states the name of the business “העוגות של ספתא” [ha-ugot ʃɛl safta] ‘grandma’s cakes’, where ספתא [safta] is a misspelling of סבתא [savta] ‘grandma’. The misspelled form represents the common spoken form of the word, which exhibits voicing assimilation of the underlying /v/ to the following /t/. (Photo taken by the author)

In the Hebrew data in (17-1), a consonant assimilates in voice to the <u>following</u> consonant. This is called **regressive** (or, **anticipatory**) **assimilation**, since the value of the [voice] feature “spreads” backwards (or being “anticipated”), as illustrated in Figure 17.2.

/savta/ → [safta] [-voice]

Figure 17.2 Regressive (anticipatory) assimilation. The value of the [voice] feature spreads backwards from [t] to /v/ to yield the surface form [f].

Assimilation can apply in the opposite direction, as well. Consider the following data featuring English verbs in the base form as well as the past and third person singular present forms.

(17-4) English verbs

	Base form	**Past**	**3SG.PRS**	
1.	ɹeɪn	ɹeɪnd	ɹeɪnz	‘rain’
2.	ɹɑb	ɹɑbd	ɹɑbz	‘rob’
3.	smaɪl	smaɪld	smaɪlz	‘smile’
4.	kɹaɪ	kɹaɪd	kɹaɪz	‘cry’
5.	hɑp	hɑpt	hɑps	‘hop’
6.	wɑk	wɑkt	wɑks	‘walk’
7.	kɔf	kɔft	kɔfs	‘cough’

The past form in (17-4) shows alternations between [d] and [t]. Examining the environments in which the alternations are realized, we see that [d] appears after voiced consonants and vowels (words 1-4), while [t] appears after voiceless consonants (words 5-7). Thus, we may conclude that the UR of the past suffix is /-d/ (note that the UR refers to the phonological representation of the suffix, and not to its written form ⟨-ed⟩). Similarly, the present suffix alternates between [z] after voiced consonants and vowels (words 1-4) and [s] after voiceless consonants (words 5-7). Therefore, its UR is /-z/.

Both past and present suffixes exhibit alternations between voiced and voiceless counterparts that can be attributed to voicing assimilation, in this case, **progressive** (or, **perseveratory**) **voicing assimilation**: a consonant assimilates in voice to the <u>preceding</u> consonant. The term “progressive” means that the value of the [voice] feature “spreads” forward (“perseveratory” means that voicing perseveres, i.e., continues longer than expected), as illustrated in Figure 17.3.

/w ɑ k - z/ → [w ɑ k s]
[-voice]

Figure 17.3 Progressive (perseveratory) assimilation. The value of the [voice] feature spreads forward from [k] to /z/ to yield the surface form [s].

This process can be formulated as follows:

(17-5) $C_{[\text{-son}]} \rightarrow [\alpha\text{voice}] \,/\, C_{[\text{-son},\ \alpha\text{voice}]}\_$

The formulation is very similar to that of regressive assimilation. The difference is in the position of the environment relative to the position in which the process applies (left or right to it). The direction of assimilation tends to be fixed per language, but can be different across languages (e.g., voicing assimilation is regressive in Hebrew but progressive in English).

17.1.1.2 Place assimilation

Another common type of assimilatory process affects place of articulation. In its most common form, nasals are assimilated in place to a neighboring obstruent. Consider the following data from Balantak.

(17-6) **Balantak** (Austronesian; Indonesia) (van den Berg & Busenitz, 2012)

	Indefinite	**One X**	
1.	urat	saŋurat	‘knuckle’
2.	ilio	saŋilio	‘day’
3.	kau	saŋkau	‘tree’
4.	gunsal	saŋgunsal	‘breath’
5.	taːʔ	santaːʔ	‘word’
6.	daŋan	sandaŋan	‘span’
7.	sinisoʔ	sansinisoʔ	‘index finger’
8.	parot	samparot	‘bunch, bundle’
9.	bandeʔ	sambandeʔ	‘half’
10.	bituʔon	sambituʔon	‘month’

The ‘One X’ prefix has three allomorphs: [sam], [san], and [saŋ]. Analysis of their distributions reveals that [saŋ] appears before vowels and velar obstruents (words 1-4), [san] appears before alveolar obstruents (words 5-7), and [sam] appears before bilabial obstruents (words 8-10). Therefore, we can conclude that the UR of the prefix is /-saŋ/, since the allomorph [saŋ] appears in the most general environment. The other allomorphs of the prefix are derived by matching the place of articulation of the nasal consonant of the prefix to that of the following consonant. This is achieved by a process of **nasal place assimilation**, in which a nasal consonant assimilates in place to a neighboring (here, following) obstruent.

Following the logic discussed in Section 15.5, we can represent the alternations in the prefix via the following formal rule:

(17-7) $C_{[+nas]} \rightarrow [place_i]$ / _$C_{[-son,place_i]}$

(17-7) reads: a nasal consonant assimilates in place to a following obstruent. The indexed feature $[place_i]$ stands for the union of all possible places of articulation, namely: {[Lab], [Cor], [Dor]}. Therefore, (17-7) is equivalent to the union of the following rules:

(17-8) $C_{[+nas]} \rightarrow [Lab] / \_C_{[-son, Lab]}$

$C_{[+nas]} \rightarrow [Cor] / \_C_{[-son, Cor]}$

$C_{[+nas]} \rightarrow [Dor] / \_C_{[-son, Dor]}$

Note that since all the observed alternations involve the prefix consonant /ŋ/, we could have written a specific rule for this consonant, instead of a more general rule that applies to all nasal consonants. However, the data does not show evidence for any other nasal + obstruent sequences with non-matching places of articulation (words 4 and 9 are the only words in the data that have a sequence of nasal + obstruent in which the nasal is not part of the prefix. In both words the alveolar nasal /n/ is followed by an alveolar obstruent). Therefore, we can assume that the place assimilation rule is general.

Nasal place assimilation is common in many other languages. For example, as discussed in Section15.5, the consonant /n/ of the Latin negation prefix /in-/ (used also in English and French) assimilates in place of articulation to the following consonant (compare *accurate* → ***in****accurate* and *possible* → ***im****possible*). While phonological processes are often limited to the spoken domain, in these examples of nasal place assimilation, the change has also affected the standard written form.

17.1.1.3 Palatalization

The following table lists the perfective and subjunctive forms of several verbs in Duwai. The perfective is marked by a suffix /-o/ and the subjunctive by the suffix /-i/ (the vowel alternation in word 1 and the tone diacritics marked on vowels are irrelevant for our purpose). In words (1-4) the last consonant of the base alternates between an alveolar in the perfective form and postalveolar in the subjunctive. Moreover, in words (1-2), there is an additional alternation between a plosive in the perfective and an affricate in the subjunctive. Finally, word (5) demonstrates an alternation between plain and palatalized alveolar nasal consonants.

(17-9) **Duwai** (Chadic; Nigeria) (Daskum, 2009; Schuh, 2002)

	Perfective	**Subjunctive**	
1.	ɨ̀fto	àft͡ʃì	‘lie down’
2.	kàːdo	kàːd͡ʒi	‘bite’
3.	màːso	màːʃi	‘buy’
4.	tùːzo	tùːʒi	‘tip to pour’
5.	wàːn	wàːnʲi	‘milk’

In both paradigms, the left-side environment of the alternating consonants is the same. Thus, the alternations are triggered by the right-side environment, namely, the suffix vowel. However, the exact mechanism is not evident from the data. In order to understand the process observed in (17-9), let us examine additional words.

(17-10) Additional words in Duwai (Daskum, 2009; Schuh, 2002)

6.	ɨtkà	'body, self'
7.	mɨ̀dgai	'shepherd'
8.	wàsko	'mix uniformly'
9.	zàːzgo	'rub, scratch'
10.	sâgi	'razor, small knife'
11.	ɽiːmi	'kapok tree, silk cotton tree'

The additional data allows us to examine the alternating consonants in additional environments. In particular, we see that the alveolar consonants can appear before consonants and vowels other than [i]. In addition, words (10-11) show that velar, retroflex and bilabial consonants can appear before [i]. Therefore, we may conclude that the alternating consonants in (17-9) are alveolar in the UR and become postalveolar or palatalized before [i]. Thus, the UR of (2) [kàːd͡ʒi] 'bite.subjunctive' is /kàːd-i/.

The data in (17-9) demonstrates the process of **palatalization**, a shift in the place of articulation of a segment towards the hard palate. The process is usually triggered by front and/or high vowels. Note that the term *palatalization* also refers to a type of secondary articulation, which does not necessarily result from a phonological process (see Section 6.2).

Palatalization can be realized in one of several ways: the segment becomes palatal, palatalized, or postalveolar (recall that postalveolar consonants are also sometimes called palatoalveolar because of their proximity to the hard palate; see Chapter 4). It can be restricted to a specific group of segments (e.g., velar consonants) or apply more generally. In addition, it may involve a change of manner, usually from a plosive to a fricative or an affricate. In Duwai, palatalization affects only alveolar consonants, which surface as postalveolar before [i]. In addition, alveolar plosives are realized as affricates. In Chapter 16, we saw that in Karok, the alveolar /s/ is realized as a postalveolar /ʃ/ after [i].

Finally, some evidence for palatalization that targets velar obstruents, specifically, is found in words with Latin influence (i.e., the palatalized forms either developed within Latin or in languages influenced by Latin). One notable case is the realization of the letter ⟨c⟩ as an alveolar or postalveolar fricative or affricate before front vowels. For example, the Latin form *magicus* [mägɪkʊs̠] ‘magical.SG.M.NOM’ (borrowed from Ancient Greek), developed into [mɑːd͡ʒikus] in Ecclesiastical Latin, and later into [maʒik] and [mad͡ʒɪk] (‘magic’) in French and English, respectively. In these forms, the original velar /ɡ/ was stabilized as postalveolar. A similar alternation of /k/ is observed when a suffix beginning with a front vowel is added (e.g., ‘magician’: English: [məd͡ʒɪʃən], French: [maʒɪsən]). Similar velar-(post)alveolar alternations are found in word pairs such as [mɛdɪsən] ‘medicine' – [mɛdɪkəl] ‘medical’, and [fənɛtɪks] ‘phonetics’ – [foʊnətɪʃən] ‘phonetician’.

Palatalization is a type of place assimilation and, sometimes, also a type of manner assimilation. The position of the tongue in the consonant becomes closer to the position of the tongue in the neighboring vowel. Two underlying mechanisms can account for the various realizations of palatalization: fronting and raising. Figure 17.4 shows the position of the tongue in the high front vowel [i], which typically triggers palatalization, and the consonants /k/ and /t/, which can undergo palatalization.

The transition between a velar consonant (e.g., /k/) and a front vowel involves **fronting** of the tongue. Palatalization occurs when the fronting becomes synchronized with the articulation of the consonant. In this case, the assimilation is in terms of the horizontal position of the tongue. By contrast, the transition between an alveolar consonant (e.g., /t/) and a high vowel involves **raising** of the tongue. Palatalization then occurs if the raising becomes synchronized with the articulation of the consonant. In this case, the assimilation is in terms of the vertical position of the tongue.

As noted above, palatalization often involves a change in manner of articulation from plosive to fricative or affricate. The mechanism seems to be reduced degree of obstruction in the oral tract induced by the neighboring vowel (see also “spirantization”, below).

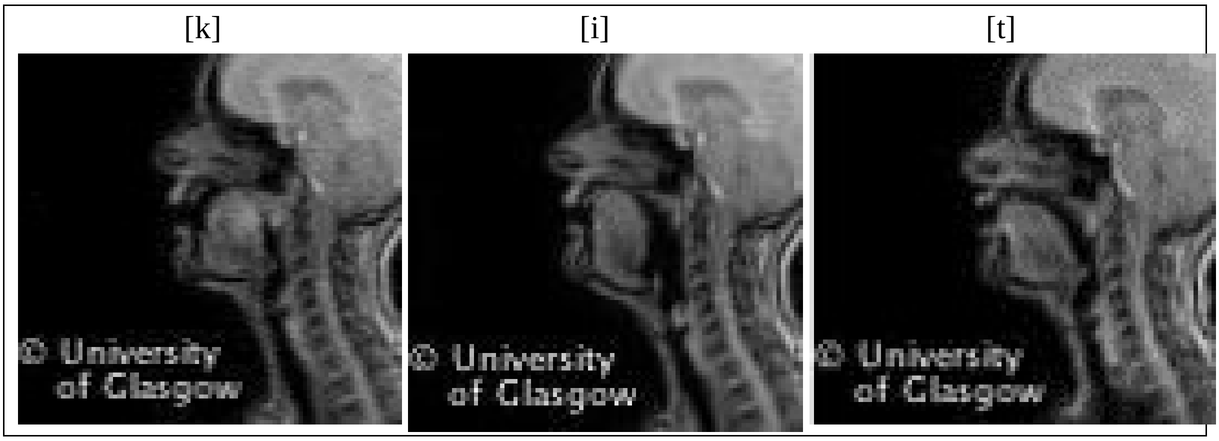


Figure 17.4 Tongue position in [k], [i], and [t]

17.1.1.4 Spirantization

In Section 15.7, we analyzed the following data from Tiberian Hebrew:

(17-11) Plosive-fricative alternations in Tiberian Hebrew

		A		**B**	
1.	‘turn’	**p**ɔnɔ	(3SG.M.PST)	ji**ɸ**nu	(3PL.M.FUT)
2.	‘count’	lis**p**or	(INF)	so**ɸ**er	(3SG.M.PRS)
3.	‘dig’	laħ**p**or	(INF)	ħɔ**ɸ**ar	(3SG.M.PST)
4.	‘fly’	ʕu**ɸ**	(2SG.M.IMP)	ʕɔ**ɸ**oθ	(3PL.F.PRS)
5.	‘hang’	**t̪**ɔlɔ	(3SG.M.PST)	ji**θ**lu	(3PL.M.FUT)
6.	‘seal’	lis**t̪**om	(INF)	so**θ**em	(3SG.M.PRS)
7.	‘sign’	jaħ**t̪**om	(3SG.M.FUT)	ħo**θ**moθ	(3PL.F.PRS)
8.	‘drink’	liʃ**t̪**oθ	(INF)	ʃo**θ**oθ	(3PL.F.PRS)
9.	‘mine’	**k**ɔrɔ	(3SG.M.PST)	ji**x**ru	(3PL.M.FUT)
10.	‘hire’	lis**k**or	(INF)	so**x**er	(3SG.M.PRS)
11.	‘sell’	jim**k**or	(3SG.M.FUT)	mo**x**er	(3SG.M.PRS)
12.	‘win’	liz**k**oθ	(INF)	zo**x**oθ	(3PL.F.PRS)
13.	‘create’	**b**ɔrɔ	(3SG.M.PST)	ji**β**rɔ	(3SG.FUT)
14.	‘break’	niʃ**b**or	(1PL.FUT)	ʃɔ**β**ar	(3SG.M.PST)
15.	‘demand’	**d̪**ɔraʃ	(3SG.M.PST)	ji**ð**roʃ	(3SG.M.FUT)
16.	‘tyrannize’	t̪ir**d̪**ɛ	(3SG.F.FUT /.2SG.M.FUT)	ro**ð**ɛ	(M.PRS)
17.	‘wean’	**g**ɔmal	(3SG.M.PST)	t̪i**ɣ**mol	(3SG.F.FUT/2SG.M.FUT)
18.	‘err’	jiʃ**g**ɛ	(3SG.M.FUT)	ʃɔ**ɣ**a	(3SG.M.PST)

An analysis of the distribution of plosives and fricatives in the data reveals that plosives appear at the beginning of words and after consonants, while fricatives appear only after vowels. Since plosives and fricatives alternate in different conjugations of the same verb, we can conclude that plosives are phonemic in Tiberian Hebrew, and that fricatives are allophones of plosives realized after vowels. The phonological process turning plosives into fricatives under the influence of vowels is called **spirantization** (*spirant* is an alternative name for fricatives). In Hebrew, spirantization occurs after vowels.

Spirantization is a type of manner assimilation: plosives involve full obstruction of the oral tract, vowels do not involve any obstruction, while in fricatives there is intermediate obstruction. In other words, spirantization makes the degree of obstruction more similar across consonants and neighboring vowels. The use of [+cont] in the formal analysis makes the spirantization mechanism more evident. Since vowels are [+cont] (see Section 15.2.1), the process of spirantization matches the value of the [cont] feature of the underlying plosives to that of the neighboring vowels. Spirantization in Tiberian Hebrew can be formally analyzed as:

(17-12) $C_{[\text{-son}]} \rightarrow [\text{+cont}]\ /\ V\_$

17.1.1.5 Full (complete) assimilation

The assimilatory processes considered so far altered a single feature (or two) in the target consonant. The data in (17-13) demonstrates **full (complete) assimilation**, namely a process by which a consonant acquires all the features a neighboring consonant. This data features definite and indefinite nouns in Palestinian Arabic. According to words (1-6), the UR of the definite article is /il-/. In words (7-12), the underlying /l/ assimilates completely to a following coronal consonant.[56] The fact that the assimilation is complete is evident especially in words 7-8, in which /l/ acquires both the manner and voice of the following consonant. Note that, in Arabic, this process is specific to /l/ in the definite article (e.g., /il-taːʒ/ → [ittaːʒ] 'the crown'). An underlying /l/ that is not part of the definite article does not assimilate to a following coronal consonant (e.g., [iltizaːm] 'commitment'').

(17-13) **Palestinian Arabic** (Semitic) (Laks, 2011)

	Indefinite	**Definite**			**Indefinite**	**Definite**	
1.	bhaːr	ilbhaːr	'spice'	7.	taːʒ	ittaːʒ	'crown'
2.	muʕallem	ilmuʕallem	'teacher'	8.	suːq	issuːq	'market'
3.	kitaːb	ilkitaːb	'book'	9.	zeːt	izzeːt	'oil'
4.	hawa	ilhawa	'air'	10.	ʒisr	iʒʒisr	'bridge'
5.	ftˤur	ilftˤur	'breakfast'	11.	raːs	irraːs	'head'
6.	ʕiːd	ilʕiːd	'holiday'	12.	loːn	illoːn	'color'

17.1.1.6 Vowel harmony

In the data examined so far, the features of some consonants were modified under the influence of neighboring consonants and vowels. In this section, we shall examine alternations in vowels. The following data illustrates case marking in Turkish nouns. Columns A and B demonstrate the nominative and genitive cases of singular nouns, respectively, while columns C and D demonstrate case marking on plural nouns.

[56] The letters of the Arabic alphabet representing coronal consonants are known as "sun letters").

(17-14) Nominative and genitive cases in Turkish (Odden, 2013)

	A. Nom SG	**B. Gen** SG	**C. Nom** PL	**D. Gen** PL	
1.	ip	ip-in	ip-ler	ip-ler-in	‘rope’
2.	ev	ev-in	ev-ler	ev-ler-in	‘house’
3.	biber	biber-in	biber-ler	biber-ler-in	‘pepper’
4.	t͡ʃɨkɨʃ	t͡ʃɨkɨʃ-ɨn	t͡ʃɨkɨʃ-lar	t͡ʃɨkɨʃ-lar-ɨn	‘exit’
5.	kɨz	kɨz-ɨn	kɨz-lar	kɨz-lar-ɨn	‘girl’
6.	sap	sap-ɨn	sap-lar	sap-lar-ɨn	‘stalk’
7.	adam	adam-ɨn	adam-lar	adam-lar-ɨn	‘man’

According to columns B and D, the genitive suffix has two allomorphs: [-in] (words 1-3) and [-ɨn] (words 4-7). According to columns C and D, the plural suffix also has two allomorphs: [-ler] (words 1-3) and [-lar] (words 4-7). In other words, each suffix has two allomorphs that differ in their vowel.

So far, when analyzing alternations in an affix, we examined the immediate environments in which the alternations occur. However, the data in (17-14) poses a challenge to this approach. In the plural suffix, the alternating vowels, [e] and [a], always occur between the consonants [l] and [r]. As for the genitive suffix, the right side environment of the alternating vowels is always [n] and, although the left side changes, we can find both vowels in the same environment (e.g., both [i] and [ɨ] occur in the environment p_n in words (1B) and (6B)).

Therefore, we can conclude that the immediate phonetic environment is not responsible for the vowel alternations in the data. In order to solve the puzzle, we note that the immediate environments in these cases contain only consonants. Thus, it might be helpful to examine the base vowel preceding the suffixes. Below are the environments of the alternating genitive vowels, where C represents the consonant preceding the alternating vowel, which we already determined to be irrelevant for our purpose.

(17-15) Environments for vowel alternations in the genitive suffix in Turkish

[-i]		[-ɨ]	
(1B)	iC_n	(4-5B)	ɨC_n
(2-3B)	eC_n	(6-7B)	aC_n

The chart above shows that each genitive vowel is preceded by different base vowels: [-i] is preceded by [i] and [e], while [-ɨ] is preceded by [ɨ] and [a]. In order to understand the mechanism underlying the observed alternations, let us draw a vowel chart for the data, according to the practice outlined in Section 15.3.

(17-16) Vowel chart for the data in (17-14)

	[-back]	central	[+back]
[+high,-low]	i	ɨ	
[-high,-low]	e		
[-high,+low]			a

When drawing the vowel chart, we followed the convention of treating [a] as a back rather than front vowel (i.e., we treat [a] as if it were written as [ɑ]). This convention stems from a combination of several facts. First, due to typographical conventions, ⟨a⟩ and ⟨ɑ⟩ are used in different typefaces as the default representation of the same letter (the first letter of the alphabet in many writing systems). Second, languages typically have only one low unrounded vowel. As a result, transcribers are usually not concerned with choosing the most phonetically accurate symbol between the two. Third, the back vowel [ɑ] is more common cross-linguistically than the front vowel [a]. Overall, the use of either [a] or [ɑ] in phonetic transcriptions is mainly a matter of convenience rather than a true phonetic representation of the vowel in question. Authors of phonological texts often note explicitly that they use [a] to represent a low, back unrounded vowel. This is what Odden (2013) does regarding the data analyzed here. Thus, [a] is classified as [+back].

A second issue arising from (17-16) is the classification of [ɨ]. Phonetically, it is a central vowel. However, according to the binary system of vowel features described in Section0 15.3.2, central vowels are not regarded as an independent class. Rather, they are classified as [+back] or [-back] in each language according to their behavior in the language. The analysis of the genitive suffix in Turkish suggests that [ɨ] patterns with [a], since the genitive suffix [-ɨn] always follows one of these vowels and never a front vowel. This suggests that [ɨ] behaves phonologically like a back vowel in Turkish.[57] Therefore, we can re-write the vowel chart for Turkish as:

(17-17) Revised vowel chart for the data in (17-14)

	[-back]	[+back]
[+high,-low]	i	ɨ
[-high,-low]	e	
[-high,+low]		a

[57] Note that in IPA, a back, high, unrounded vowel is represented by [ɯ] and not by [ɨ]. Assuming the given transcription is faithful to the actual pronunciation, the data supports the claim that the phonological features of vowels are designed to reflect phonological behavior rather than articulation (see Section 15.3).

We can summarize the analysis of the alternating vowels in the genitive suffix as follows: [-in] appears when the last vowel of the base is [-back], while [-ɨn] appears when the last vowel of the base is [+back]. What is more, the vowel chart in (17-17) suggests a potential motivation for the observed alternation: the backness of the genitive suffix vowel always matches the backness of the preceding base vowel. In other words, the alternation of the vowel in the genitive suffix is determined by assimilation to the preceding vowel. But, unlike other cases of assimilation we examined earlier, the data in (17-14) demonstrates assimilation between non-adjacent segments (i.e., the assimilated vowels are separated by consonants). Such **"long-distance" assimilation** is also called **harmony**. In particular, assimilation of one vowel to another vowel is called **vowel harmony**.

As in other cases of assimilation, the alternation between segments (here, [i] and [ɨ]) suggests that they are allophones of the same phoneme. Thus, the next step is determining the general allophone which, in this case, also means determining the UR of the genitive suffix. The analysis suggests that a [+back] vowel in the suffix ([ɨ]) always follows a [+back] vowel, and a [-back] ([i]) in the suffix always follows a [-back] vowel. In other words, we have complementary distributions that are equally general. Thus, it is impossible to determine, based on the given data, which genitive vowel is the general allophone. We can only conclude that the UR of the genitive suffix contains a high, unrounded vowel, but we do not know whether it is [+back] or [-back]. We say that the UR is **underspecified** with respect to the backness of the vowel, that is, the lexical representation of the genitive suffix does not specify the backness of the vowel. The UR of the genitive suffix can be written as: /-$V_{[+high,-round]}$n/.

Following the logic applied to the genitive suffix, we can analyze the plural suffix. We observe that [-ler] appears when the last vowel of the base is front (i.e., [-back]), while [-lar] appears when the last vowel of the base is [+back]. Since [e] is [-back] and [a] is [+back], we conclude that the backness of the plural suffix is determined according to the backness of the last base vowel. In other words, the plural suffix demonstrates vowel harmony, similarly to the genitive suffix. Again, the data does not indicate which form of the suffix is more general.

When analyzing the behavior of the plural suffix, it is important to note that in addition to the backness change, there also seems to be a height change. There is no obvious motivation for this change. Assuming that the transcribed data represents the pronunciation accurately, we can speculate that the height change is epiphenomenal to the backness change and is required by the phonetic inventory of Turkish. In other words, the main change is in backness and the change of height is required since Turkish does not have an equivalent vowel with the same height as the UR vowel (i.e., a front low unrounded vowel or a back close-mid unrounded vowel).

Though it may sound complicated, in practice, the height change has little effect on the analysis. In fact, by slightly revising the vowel chart, grouping [a] and [e] under [-high], the alternation becomes clearer. Consequently, we may write the UR of the plural suffix as: /-l$V_{[\text{-high,-round}]}$r/.

(17-18) Revised vowel chart for the data in (17-14)

	[-back]	[+back]
[+high]	i	ɨ
[-high]	e	a

The analysis of the genitive and plural suffixes reveals that both behave similarly: the suffix vowel assimilates in backness to the preceding vowel, across one or two consonants. Formally, this can be written as:

(17-19) $V_{[\text{-round}]} \rightarrow [\alpha\text{back}] \;/\; V_{[\alpha\text{back, -round}]}C_0\,\_$

(17-19) reads: an unrounded vowel assimilates in backness to a preceding unrounded vowel across any number of consonants. C_0 is a formal way to represent a sequence of consonants of any length (zero or more). Although our data does not show evidence that vowel harmony applies when the vowels are three or more consonants apart, we have no evidence to the contrary. Generally, it is assumed that if vowels can assimilate (or, harmonize) across two consonants, then the number of intervening consonants can be any number, at least in theory.

In the analysis above, we indicated that both assimilated vowels are [-round]. Since the data in (17-14) did not contain any [+round] vowels, we could have chosen to eliminate the [-round] specification in (17-19) and generalize the rule to any vowel. While this approach is valid (see also Section 15.6), one may wish to be more cautious and look for more direct evidence for the behavior of segments. The following data provides such direct evidence for the behavior of rounded vowels in Turkish.

(17-20) More genitive and plural in Turkish (Odden, 2013)

	A. Nom SG	**B. Gen SG**	**C. Nom PL**	**D. Gen PL**	
8.	jyz	jyz-yn	jyz-ler	jyz-ler-in	‘face’
9.	køj	køj-yn	køj-ler	køj-ler-in	‘village’
10.	pul	pul-un	pul-lar	pul-lar-ɨn	‘stamp’
11.	ok	ok-un	ok-lar	ok-lar-ɨn	‘arrow’
12.	son	son-un	son-lar	son-lar-ɨn	‘end’

Column B in (17-20) shows that the genitive suffix has two additional allomorphs that were not observed in (17-14): [-yn] and [-un], both containing rounded vowels. It is also noteworthy that in all the examples above, the vowel of the root is also rounded.

To better understand the behavior of the genitive suffix, it would be useful to update the vowel chart in (17-18) according to the new data:

(17-21) Revised vowel chart for the data in (17-20)

	[-back]		[+back]	
[+high]	i	y	ɨ	u
[-high]	e	ø	a	o

Overall, the genitive suffix has four allomorphs:

- [-in]: appears when the last vowel of the base is [-back] and [-round] ([i] is also [-back] and [-round]).
- [-ɨn]: appears when the last vowel of the base is [+back] and [-round] ([ɨ] is also [+back] and [-round]).
- [-yn] appears when the last vowel of the base is [-back] and [+round] ([y] is also [-back] and [+round]).
- [-un] appears when the last vowel of the base is [+back] and [+round] ([u] is also [+back] and [+round]).

With the revised vowel chart, we can determine that the genitive suffix always contains a high vowel and is underspecified with respect to backness and roundedness, which are determined by the backness and roundedness of the last vowel of the base. Thus, the UR of the suffix is: /$V_{[+high]}$n/.

Interestingly, the genitive suffix does not show the roundedness alternation in column D when it is preceded by the plural suffix. Moreover, the plural suffix itself does not exhibit roundedness alternation at all (e.g., the plural form of (8) is [jyz-ler] and not [jyz-lør]). Thus, our earlier conclusion about the plural suffix remains: it contains a non-high, non-round vowel that is underspecified with respect to backness. The UR of the plural suffix is then: /l$V_{[-high,-round]}$r/.

Going over all the data in (17-20) and (17-14), we can see that the vowels in both genitive and plural suffix agree in backness with their base. Since the genitive vowel is always [+high] and the plural vowel is always [-high], we may speculate that the backness of a suffix vowel is affected by the backness of the base vowel, regardless of vowel type. Moreover, the data shows that all the vowels in a given word (both in roots and suffixes) agree in backness, i.e., all the vowels in a word are either [+back] or [-back]. Such agreement in features across a number of vowels is a key signature of harmony. The backness vowel harmony rule in (17-19) can now be generalized to any vowel:

(17-22) $V \rightarrow [\alpha back]$ / $V_{[\alpha back]}C_0$ _

The formula above reads: a vowel assimilates in backness to a preceding vowel across any number of consonants.

While both genitive and plural suffixes are affected by backness harmony, only the genitive suffix undergoes roundedness harmony. A possible explanation for the different behaviors of the genitive and plural suffixes is due to the different vowels contained in them: the genitive suffix contains [+high] vowels, while the plural suffix contains [-high] vowels. Thus, we hypothesize that only [+high] suffix vowels are assimilated in roundedness to the base vowel,[58] and formally:

(17-23) $V_{[+high]} \rightarrow [\beta round]$ / $V_{[\beta round]}C_0$ _

The formula above reads: a high vowel assimilates in roundedness to a preceding vowel across any number of consonants. It is important to note that, while vowel harmony can skip intervening consonants, it does not skip intervening vowels. Thus, the genitive vowel in column

[58] An alternative analysis would be to link the behavior of the vowels to the nature of the specific suffixes (i.e., plural versus genitive). However, we normally prefer to analyze phonological processes in phonological terms and apply affix-specific analysis only when it is impossible to account for the behavior of different affixes in pure phonological terms.

D of (17-20) does not assimilate in roundedness to the root vowel because harmony is blocked by the intervening vowel of the plural suffix.

To summarize the analysis of the Turkish data, we examined two suffixes containing a vowel whose features are underspecified in the UR and determined by processes of vowel harmony. The vowel of the genitive suffix is affected by two processes: it assimilates both in backness and roundedness to the preceding vowel. By contrast, the vowel of the plural suffix assimilates only in backness to the preceding vowel. This difference was accounted for by a constraint that limits the application of roundedness harmony to high vowels.

Generally, vowel harmony can be quite complex to analyze with several elements requiring attention (as a matter of fact, vowel harmony is not a single process but a family of processes, analogous to the union of the various assimilatory processes in consonants). First, one needs to realize that vowel harmony is taking place by noticing consistent vowel alternations that cannot be accounted for by the surrounding consonants. Then, it is recommended to draw a vowel chart for the data according to the features listed in Section 15.3. Finally, one can analyze the various aspects of the process and write a phonological rule. The following checklist can help analyzing the data:

(17-24) VH checklist:

1. **Vowel chart**: group vowels according to [high], [low] and [back]. Pairs of [+round] and [-round] vowels should be organized clearly within cells ([+round] on the right). The [ATR] feature may be relevant, especially when the language has both close (high) and near-close vowels or both close-mid and open-mid vowels (see Section 15.3.3). If the data contains central vowels, their behavior should be examined to determine whether they should be grouped with front or back vowels. Consequently, the chart should be updated after target and trigger vowels are analyzed and there is sufficient information to classify central vowels as patterning with front or back vowels.

2. **The harmonized features**: list pairs of alternating vowels and check which features change within pairs.

3. **Target vowels**: check whether only vowels with specific features undergo harmony.

4. **Trigger vowels**: check whether only vowels with specific features trigger harmony

5. **Shared features**: check whether harmony applies only when both participating vowels share certain features.

6. **Number of intervening consonants**: check whether harmony can apply across more than one consonant.

7. **Type of intervening consonant**: check whether harmony applies only if the vowels are separated by specific consonants (in some languages vowel harmony applies across glottal consonants but not across other consonants).

8. **Morpheme boundaries**: if all the vowels in the word (both root and affixes) share the harmonized features, morpheme boundaries are irrelevant. However, if affix vowels assimilate to root vowels, but the root contains vowels that do not share the relevant feature, then morpheme boundary is an integral part of the process.

After collecting all the information from the checklist, write the phonological rule. Use Greek letters to indicate agreement in features. If the rule involves agreement in more than one feature, use a different letter for each feature (α, β, γ). This is relevant both for harmonized features and required shared features. Intervening consonants are indicated according to the following convention: if harmony applies across one consonant but not two, use **(C)**, which stands for zero or one consonants. If harmony applies across two consonants, use $\mathbf{C_0}$, which stands for any number of consonants (zero or more). Finally, if harmony applies only when participating vowels belong to different morphemes, use '+' to indicate morpheme boundary in the rule.

For example, the following checklist can be used to analyze the roundedness harmony in Turkish:

(17-25) Roundedness harmony in Turkish:

1. Vowel chart: see (17-21).
2. The harmonized feature: [βround] (the roundedness of the genitive vowel is determined by the roundedness of the last root vowel)
3. Target vowels: $V_{[+high]}$ (only high vowels are affected by roundedness harmony).
4. Trigger vowels: V – there are no specific constraints on trigger vowels.

5. Shared features: none – there is no evidence that harmony applies only if the participating vowels share a certain feature.
6. Number of intervening consonants: C_0 – harmony applies across two intervening consonants. There is no evidence for an upper limit on the number of intervening consonants.
7. Type of intervening consonant: any – there is no evidence that harmony applies only across specific intervening consonants.
8. Morpheme boundaries: not required – there is no evidence that harmony applies only if the participating vowels are separated by a morpheme boundary. No words contain a mixture of rounded and unrounded vowels within the same morpheme.

Assimilation – summary

There are different types of assimilation. The main motivation seems to be ease of articulation: assimilation reduces the number of required articulatory movements (speech requires fast adjustments). We have seen that assimilation can operate in different directions – both regressive and progressive assimilations are observed. Both directions are motivated, but each has a different underlying mechanism. Regressive (anticipatory) assimilation indicates planning ahead – the articulators are positioned early in anticipation of the next segment. Progressive (perseveratory) assimilation suggests a "failure" to adapt the articulators, so they remain in the configuration appropriate for the previous segment.

17.1.2. Dissimilation

The following dataset illustrates the UR and PR forms of some phrases in Ainu. Examples (3-4) show that when the UR contains a sequence of two /ɾ/s, the first of them is realized as [n]. This is an example of **dissimilation**, a process by which one segment changes so as to become less similar to another segment. Possible motivations for dissimilation could be avoiding repetition of consonants and maintaining distinctiveness of word boundaries (word boundaries are often blurred in rapid speech).

(17-26) **Ainu** (Language isolate; Northern Japanese island of Hokkaido) (Suzuki, 1998)[59]

	UR	PR	
1.	kuko**ɾ** **k**uɾ	kuko**ɾ** **k**uɾ	‘my husband’
2.	ko**ɾ** **m**at	ko**ɾ** **m**at	‘his wife’
3.	ko**ɾ** **ɾ**ametok	ko**n** **ɾ**ametok	‘his bravery’
4.	kuko**ɾ** **ɾ**usuj	kuko**n** **ɾ**usuj	‘I want to have (something)’

17.1.3. Final devoicing

In Section 16.3 we analyzed the following data from Javanese:

	(A) Noun	(B) This noun	Gloss
1.	t͡ʃaŋkɪr	t͡ʃaŋkire	‘cup’
2.	t͡ʃukʊr	t͡ʃukure	‘haircut’
3.	ket͡ʃap	ket͡ʃape	‘soy sauce’
4.	kulɪt	kulite	‘skin’
5.	səbap	səbabe	‘reason’
6.	murɪt	muride	‘student’
7.	bedʊk	beduge	‘mosque drum’

The words in column B are generated by a suffix /-e/ added to the nouns in column A. In addition to suffixation, words (5-7) demonstrate a voicing alternation (e.g., the [p]-[b] alternation in (5) [səbap] ‘reason’ – [səbabe] ‘this reason’). In these three words voiced plosives appear between vowels and voiceless plosives appear word-finally. Since voiceless plosives also appear between vowels in the data (e.g., [p] in (4) [ket͡ʃape]), and voiced plosives do not appear word-finally, we conclude that underlying voiced plosives become devoiced at the end of the word. Therefore, the UR of noun (5) is /səbab/. The alternations in the data are due to **final devoicing**, a process by which an obstruent becomes voiceless word-finally. Formally, it can be represented as:

(17-27) $C_{[\text{-son}]} \rightarrow$ [-voice] / _#

[59] As of the beginning of the 21st century, Ainu is considered a critically endangered language with only a small number of remaining native speakers (see http://www.endangeredlanguages.com/lang/1212).

17.2. Processes affecting whole segments

The processes discussed in the previous section altered specific features of target segments. There are also phonological processes that affect segments as wholes.

17.2.1. Deletion

Consider some masculine and feminine singular past tense verbs in Hebrew.

(17-28) Hebrew verbs

	A. Masculine	B. Feminine	
1.	baat	baata	‘kick’
2.	ʃaal	ʃaala	‘ask’
3.	dibɛʁ	dibʁa	‘speak’
4.	sipɛʁ	sipʁa	‘tell’
5.	ʃamaʁ	ʃamʁa	‘guard, keep’
6.	masaʁ	masʁa	‘handed’
7.	amaʁ	amʁa	‘say’

Verbs in the two paradigms exhibit a morphological difference: verbs in the feminine are formed by adding the /-a/ suffix to masculine verbs. In addition, words (3-6) exhibit a phonological alternation – the second vowel in the masculine form ([ɛ] in (3-4), [a] in (5-6)) is absent in the feminine form. These are examples of **vowel deletion**. Thus, the UR of (3B) is /dibɛʁ-a/ ‘speak.3SG.F.PST’.

The deletion site seems to be enclosed between a vowel-consonant sequence on the left and a consonant-vowel sequence on the right, or VC_CV. The process is formally represented as:

(17-29) V → Ø / VC_CV

It reads: “a vowel is deleted (becomes “nothing”) after a VC and before a CV sequence”.

The VC_CV structure is known as a “two-sided open syllable” and is a common site of vowel deletion cross linguistically. For example, consider the alternative pronunciations of the word ‘family’: [fæməli] / [fæmli]. In the second pronunciation the vowel [ə] is deleted at a VC_CV environment.

Deletion is a common process cross-linguistically and can affect both vowels and consonants. For example, in Section 16.2.3, we analyzed consonant deletion at the end of words in Samoan. Possible underlying motivations for deletion could be avoiding articulatorily or perceptually challenging sequences. Common cases for deletion include: (i) consonant deletion at the end of the word, (ii) consonant deletion in a sequence of consonants (avoiding specific sequences or consonant sequences in general), (iii) vowel deletion in a sequence of vowels, and (iv) vowel deletion in a "two-sided open syllable".

17.2.2. **Epenthesis (insertion)**

Consider the nominative and dative forms of some nouns in Turkish.

(17-30) Turkish (Odden, 2013)

	A. Nominative	**B. Dative**	
1.	oda	odaja	'room'
2.	balo	baloja	'ball'
3.	imlaː	imlaːja	'spelling'
4.	arɨ	arɨja	'bee'
5.	at	ata	'horse'
6.	halk	halka	'folk'
7.	kojun	kojuna	'sheep'
8.	sarp	sarpa	'steep'

While the nominative case seems to be a null affix, the dative case suffix has two allomorphs: [-ja] in words (1-4), and [-a] in words (5-8). [-ja] appears after bases ending in vowels, while [-a] appears after bases ending in consonants. Determining the UR of the dative suffix is not straightforward. However, it can be deduced by some phonological reasoning.

If the UR of the dative suffix is /-ja/, this suggests that [j] is deleted in words (5-8) (i.e., the UR of (5B) is: /at-ja/). Since each base in words (5-8) ends with a different consonant, we may hypothesize that [j] is deleted to avoid a sequence of consonants. However, according to words (3), (6), and (8) Turkish does allow sequences of consonants.[60] Thus, assuming the UR of the dative suffix is /-ja/ lacks a clear motivation.

If the UR of the dative suffix is /-a/, this means that [j] is inserted in words (1-4) (i.e., the UR of (1B) is: /oda-a/). Note that the bases in words (1-4) end with a vowel, so the dative forms of these words would contain a sequence of vowels in the UR. Moreover, no word in the data contains a sequence of vowels. Therefore, we may hypothesize that the consonant [j] is inserted between two underlying vowels to avoid surface vowel sequences. Thus, the Turkish data

[60] The data does not show this, but Turkish also allows [j] after consonants, as in, e.g., [antalja] 'Antalya'.

demonstrates **consonant epenthesis** (or **consonant insertion**). The process is formally represented as:

(17-31) $\varnothing \rightarrow j$ / V_V

(17-31) reads: "[j] is inserted (or, a "nothing" becomes [j]) between vowels."

Epenthesis is a common process cross-linguistically and can affect both vowels and consonants. For example, in Section 16.2.1, we analyzed vowel epenthesis between strident consonants in English. Similarly to deletion, epenthesis could also be a strategy to avoid articulatorily or perceptually challenging sequences. Common cases for epenthesis include: (i) vowel epenthesis at the end of words ending with a consonant (ii) vowel epenthesis in a sequence of consonants (breaking specific sequences or consonant sequences in general), (iii) consonant epenthesis in a sequence of vowels, and (iv) consonant epenthesis at the beginning of the words that start with a vowel.

<u>Epenthesis, deletion, affixation</u>

Processes of epenthesis and deletion can be easily confused with each other and with affixation. Thus, it is important to understand the nature and recognize the typical signature of each of them. Affixation is a <u>morphological process</u> that affects the meaning of words (plurality, tense, etc.). An affix appears (in principle) in all the words in a morphological paradigm. Alternations in affixes are phonologically predictable and usually result in one of several allomorphs.

By contrast, deletion and epenthesis are <u>phonological processes</u>. They do not apply to an entire morphological paradigm, but only to a subset of words that contain a specific phonetic environment. The features of epenthetic segments are fixed (or highly predictable). It usually involves a <u>specific</u> segment that appears in a specific environment. For example, the Turkish epenthetic consonant [j] is inserted between vowels, while in English, the epenthetic vowel [ə] is inserted between similar or identical consonants (e.g., two stridents).

The hallmark of deletion is when <u>several</u> different segments belonging to a broad <u>class</u> are missing from a specific environment. Unlike epenthetic segments that are highly predictable, the exact features of deleted segment are much unpredictable. For example, 16.2.3, the perfective forms of Samoan verbs contain the consonants [l], [f], and [ŋ] that are absent from the simple verb forms. These consonants do not share any features and, therefore, cannot be epenthetic or part of the perfective suffix, but rather deleted in surface forms of simple verbs.

17.2.3. Metathesis

The following dataset lists the masculine and neuter forms of some adjective in Faroese. Examination of the data indicates that the masculine is expressed by the suffix /-ʊr/, and the neuter is expressed by the suffix /-t/.

(17-32) **Faroese** (Germanic; Faroe Islands) (Lockwood, 1955)

	Masculine	**Neuter**	
1.	tung-ʊr	tuŋg-t	‘heavy’
2.	smal-ʊr	smal-t	‘narrow’
3.	fɛsk-ʊr	fɛks-t	‘fresh’
4.	rask-ʊr	raks-t	‘energetic’
5.	dansk-ʊr	daŋks-t	‘Danish’

In addition to the change of suffix, words (3-5) demonstrate a change of order between adjacent consonants – the masculine form of these words contain the sequence [sk], while the neuter form contains the sequence [ks]. This change of order is known as **metathesis**, whose assumed motivation is to prevent “ill-formed” sequences. In the case of Faroese, it is commonly hypothesized that metathesis swaps [s] and [k] before [t] to avoid the sequence [skt].[61]

Another case of metathesis occurs in Hebrew. (17-33) lists the UR and PR of third person singular past form of verbs in Binyan hitpa’el. Verbs in this paradigm contain a prefix that ends with /t/. When attached to bases that begin with a strident consonant, the suffix /t/ switches places with the base strident (e.g., /hi**ts**apeʁ/ → [hi**st**apeʁ] ‘get a haircut’).

(17-33) Metathesis in **Hebrew** (Semitic)

	UR	**PR**	
1.	hitlabeʃ	hitlabeʃ	‘get dress’
2.	hitxamem	hitxamem	‘heat.intr’
3.	hitgabeʁ	hitgabeʁ	‘overcome’
4.	hitsapeʁ	histapeʁ	‘get a haircut’
5.	hit͡ʃapeʁ	hi͡ʃtapeʁ	‘improve.intr’
6.	hit͡tsadek	hit͡stadek	‘justify oneself’

[61] Metathesis does not have a formal representation within linear phonology theories.

17.3. Practice

1. **Lithuanian** (Dambriūnas et al., 1993; Mathiassen, 1996)

The data in (17-34) demonstrates two prefixes in Lithuanian, each having various meanings. Prefix 1 may indicate: arrival, separation, return to original position, completion. Prefix 2 may indicate: around, from all sides, at the side of, completion. The diacritics above vowel indicate pitch accent, which is irrelevant for this exercise (more about pitch accent in the *planned* Chapter 20 ("Tone and pitch-accent")).

(17-34) Lithuanian prefixes

Prefix 1			**Prefix 2**		
1.	at-eĩti	'arrive'	1.	ap-eĩti	'circumvent'
2.	at-im̃ti	'take away'	2.	ap-ieʃkóti	'search everywhere'
3.	at-nèʃti	'bring'	3.	ap-láistːti	'sprinkle all around'
4.	at-rìʃti	'untie'	4.	ap-ródiːti	'show around'
5.	at-kir̃sti	'cut off'	5.	ap-kur̃sti	'become deaf'
6.	at-skìrti	'separate'	6.	ap-sùpti	'surround'
7.	ad-bùkti	'become blunt'	7.	ab-dauʒiːti	'damage'
8.	ad-gáuti	'get back'	8.	ab-gíːdiːti	'heal to some extent'
9.	ad-vaʒiúoti	'arrive by vehicle'	9.	ab-ʒélti	'become overgrown'

A. State the allomorphs of each prefix, the environments in which each allomorph appears, and determine the UR of the prefixes. Explain your answers.

B. Analyze the phonological process responsible for the alternations in the prefixes.

Solution

A. Prefix 1 has two allomorphs: [at] before vowels (words 1-2), sonorant consonants (words 3-4), and voiceless obstruents (words 5-6); [ad] before voiced obstruents (words 7-9). The UR of prefix 1 is /at-/ since this allomorph appears in the most general environment. Prefix 2 has two allomorphs: [ap] before vowels (words 1-2), sonorant consonants (words 3-4), and voiceless obstruents (words 5-6); [ab] before voiced obstruents (words 7-9). The UR of prefix 2 is /ap-/ since this allomorph appears in the most general environment.

B. Both prefixes demonstrate the following process:

Name of the process: Voicing assimilation

Description in words: An obstruent assimilates in voice to a following obstruent.

Formal rule: $C_{[-son]} \rightarrow [\alpha voice] / \_C_{[-son,\ \alpha voice]}$

2. **Lamba** (Niger-Congo; Zambia) (Odden, 2013)

	Plain	**Passive**	**Neuter**	
1.	t͡ʃitɑ	t͡ʃitwɑ	t͡ʃitikɑ	‘do’
2.	tulɑ	tulwɑ	tulikɑ	‘dig’
3.	pɑtɑ	pɑtwɑ	pɑtikɑ	‘scold’
4.	fisɑ	fiswɑ	fiʃikɑ	‘hide’
5.	lɑsɑ	lɑswɑ	lɑʃikɑ	‘wound’
6.	mɑsɑ	mɑswɑ	mɑʃikɑ	‘plaster’
7.	t͡ʃetɑ	t͡ʃetwɑ	t͡ʃetekɑ	‘spy’
8.	soŋkɑ	soŋkwɑ	soŋkekɑ	‘pay tax’
9.	t͡ʃesɑ	t͡ʃeswɑ	t͡ʃesekɑ	‘cut’
10.	kosɑ	koswɑ	kosekɑ	‘be strong’

A. Fill in the vowel chart for Lamba.

	[-back]	[+back]
[+high]		
[-high,-low]		
[-high,+low]		

The following analysis has been proposed for the data above. For each of the statements below, determine whether it is correct or not. If a statement is incorrect, explain why it is incorrect (e.g., provide examples that refute the statement) and provide the correct analysis.

B. The following underlying forms have been proposed for the verbal suffixes in Lamba.

i. Plain: Ø. Explanation: this is the plain form of the verb and, therefore, it has no suffix.
ii. Passive: /-wɑ/. Explanation: all the words in the column end with [wɑ].
iii. Neuter: /-$V_{[-back,-round]}$kɑ/. Explanation: all the words in the column end with a front, unrounded vowel and [kɑ]. The first vowel of the suffix is underspecified because there is no way to tell which of its allophones appears in the more general environment.

C. The alternations in the base in words 4-6 have been claimed to result from the following process:

- Name of the process: palatalization

- Description in words: a process in which the place of articulation of a segment becomes [+palatal].
- Formal rule: C → [+palatal] / _$V_{[+high]}$

D. The alternations in the neuter suffix have been claimed to result from the following process:

- Name of the process: vowel lowering
- Description in words: a high suffix vowel becomes [-high].
- Formal rule: V → [-high] / +_

E. The UR of [kosekɑ] (10, neuter) is claimed to be /kos-eka/ because this is the form we see in the data.

<u>Solution</u>

A. Vowel chart:

	[-back]	[+back]
[+high]	i	u
[-high,-low]	e	o
[-high,+low]		ɑ

B. Suffixes:

i. Plain: Incorrect. The explanation is irrelevant. The UR is determined based on phonological considerations, not semantic considerations. The UR of the plain form is /-ɑ/ since all the words in that column end with [ɑ].

ii. Passive: Correct.

iii. Neuter: Incorrect. The UR is /-ikɑ/. The first vowel of the suffix is predictable. It is [e] when the last vowel of the base is [-high,-low] and [i] otherwise (i.e., when the last vowel of the base is [+high] or [+low]).

C. The alternations in the base in words 4-6: Partially correct. The process is indeed palatalization, but the description and formal rule are lacking. First, the description is too general. It refers to any segment, while the data shows the process applies specifically to

/s/ (other consonants like /t/ and /l/ do not palatalize in the data). Second, the description does not specify the environment in which the process applies. Third, the description and formal rule are inconsistent. According to the formal rule, the process applies to any consonant (not "segment" as specified in the description), and it applies before a high vowel (this is not mentioned in the description). Fourth, the formal rule is also inaccurate. The data shows that only /s/ is affected by the rule. Thus, for example, /t/ in (2, plain) [tulɑ] does not become [+palatal] before [u] and /l/ in (2, neuter) [tulikɑ] does not become [+palatal] before [i].

The correct analysis:

- Name of the process: palatalization
- Description in words: /s/ becomes [+palatal] before a high vowel.
- Formal rule: s → [+palatal] / _$V_{[+high]}$

Note: alternative analyses are possible (e.g., the rule is triggered specifically by [i]).

D. The alternations in the neuter suffix: Incorrect. Although, the observed alternation is a change of a high vowel into a non-high vowel, the process responsible for this alternation is vowel harmony. According to the proposed analysis, every instance of high vowel in a suffix should surface as a non-high vowel. However, this alternation is not observed in words #1-6 (e.g., /i/ does not become [e] in (2, neuter) [tulikɑ]).

The correct analysis:

- Name of the process: vowel harmony
- Description in words: a vowel becomes [-high] after a [-high,-low] vowel across any number of consonants.
- Formal rule: V → [-high] / $V_{[-high,-low]}C_0$_

Note: alternative analyses are possible (e.g., the rule might apply specifically to /i/, or only to high vowels in a suffix).

E. The UR of [kosekɑ] (10, neuter): Incorrect. Phonological analysis reveals that the UR of the base of (10) is /kos/ (the base appears in all the columns of row 10 with no alternations), and that the UR of the neuter suffix is /-ika/ (see above). Therefore, the UR of (10, neuter) is /kos-ika/.

3. **Hungarian** (Uralic; Hungary) (Odden, 2013)

	Noun	**In noun**	**From noun**	
1.	kalap	kalabban	kalapto:l	‘hat’
2.	ku:t	ku:dban	ku:tto:l	‘well’
3.	ʒa:k	ʒa:gban	ʒa:kto:l	‘sack’
4.	rab	rabban	rapto:l	‘prisoner’
5.	ka:d	ka:dban	ka:tto:l	‘tub’
6.	fal	falban	falto:l	‘wall’
7.	a:gʲ	a:gʲban	a:kʲto:l	‘bed’
8.	ʃro:f	ʃro:vban	ʃro:fto:l	‘screw’

A. Find the UR of the ‘In noun’ and ‘From noun’ suffixes. Explain.

The following analysis has been proposed for the data above. For each of the statements below, determine whether it is correct or not. If a statement is incorrect, explain why it is incorrect (e.g., provide examples that refute the statement) and provide the correct analysis.

B. The alternations in the bases have been claimed to result from the following process:

- Name of the process: final devoicing
- Description in words: a plosive at the end of the word becomes voiceless.
- Formal rule: $C_{[\text{-son,-cont}]} \rightarrow$ [-voice] / _#

C. [b] and [p] are said to be allophones of /b/ in Hungarian. [p] is the restricted allophone that appears only at the end of the word, [b] is the general allophone and appears elsewhere.

<u>Solution</u>

A. UR of suffixes:
   i. ‘In noun’: /-ban/. All the words in the column end with [ban].
   ii. ‘From noun’: /-to:l/. All the words in the column end with [to:l].

B. The alternations in the bases: Incorrect. The analysis predicts that all word-final plosives should be [-voice]. However, nouns 4, 5, and 7 end with a voiced plosive. Also, the proposed analysis misses the fact that fricatives (which are [+cont]) also exhibit the voicing alternation (word 8). In addition, rows supposedly showing “final devoicing” (e.g., row 1) also show devoicing before [t] in the ‘from noun’ column. A complete analysis of the data must include the following rule of voicing assimilation, which accounts for all the alternations in the data, thus making final devoicing redundant.

The correct analysis:

- Name of the process: voicing assimilation
- Description in words: an obstruent assimilates in voice to a following obstruent.
- Formal rule: $C_{[-son]} \rightarrow [\alpha voice] / \_C_{[-son, \alpha voice]}$

C. [b] and [p]: Partially correct. /p/ is also a phoneme in Hungarian with [b] as one of its allophones, as can be seen in row 1. In general, due to voicing assimilation, [b] and [p] are allophones of each other. The fact that they are distinct phonemes can be inferred from the fact that they can appear in the same phonetic environment – at the end of the word after [a] (nouns 1 and 4).

4. **Maori** (Austronesian; New Zealand) (Kenstowicz & Kisseberth, 1979)

	Verb	**Gerundive**	
1.	awhi	awhitaŋa	‘embrace’
2.	hopu	hopukaŋa	‘catch’
3.	aru	arumaŋa	‘follow’
4.	tohu	tohuŋaŋa	‘point out’
5.	mau	mauraŋa	‘carry’
6.	wero	werohaŋa	‘stab’

A. It has been proposed that the UR of the gerundive suffix is /-Caŋa/, since all the words in the gerundive column end with a consonant + [aŋa]. Is this statement correct? Explain.

B. Analyze the phonological alternations in the data.

C. Determine the UR of the verb and gerundive forms of words in (5) ‘carry’.

   i. Verb – [mau]:

   ii. Gerundive – [mauraŋa]:

Solution

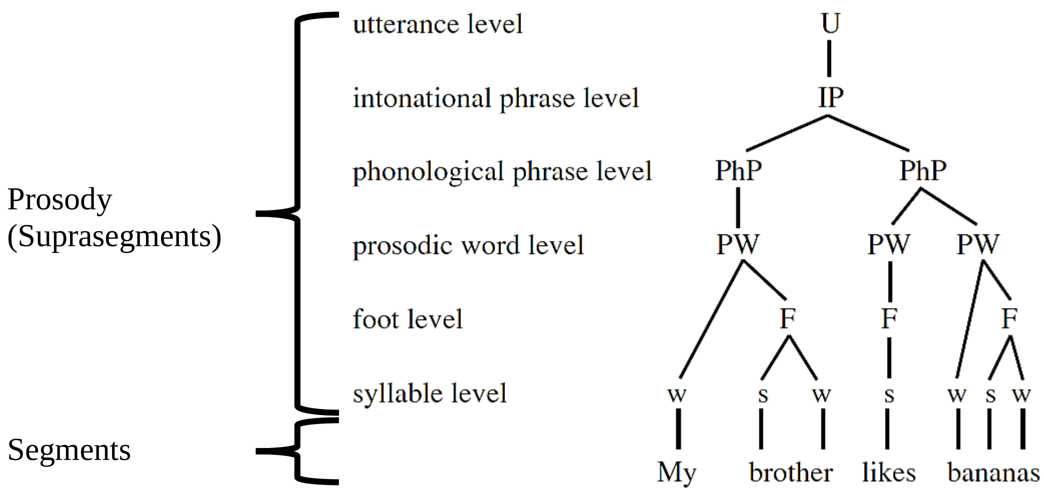
A. The UR of the gerundive suffix: Incorrect. The UR is /-aŋa/: all the words in that column end with [aŋa]. The consonant preceding the sequence [aŋa] is different in different words and cannot be predicted. Therefore, it is likely a part of the base and not part of the suffix.

B. The phonological alternations in the data:
   - Name of the process: consonant deletion
   - Description in words: a consonant is deleted at the end of the word.
   - Formal rule: C → Ø / _#

C. The UR of the verb and gerundive forms of words in (5) ‘carry’:
   i. Verb – [mau]: /maur/
   ii. Gerundive – [mauraŋa]: /maur-aŋa/

Part V: Prosodic phonology

The fifth part of the book explores *prosody*, fondly called the "music" of language". In practice, prosodic phonology studies **suprasegmental** phenomena: phonological entities whose domain is larger than a single segment. The figure below, adapted from Gerken & McGregor (1998), illustrates the relation between the segmental and prosodic (or suprasegmental) domains.



As can be seen in the figure, prosody has several levels, each of which is associated with different functions and phenomena. Prosody can be associated with word-level, lexical phenomena, such as word stress in stress languages, like English, and tone patterns in tonal languages, like Chinese. In addition, prosody includes phenomena such rhythm and intonation, which reflect the syntax and discourse levels of oral communication. Each chapter in this part focuses on a different phonological level and function.

Chapter 18 Syllables

The following table contains the masculine and feminine forms of some adjectives in Hebrew:

(18-1) Hebrew adjectives

	Masculine	**Feminine**	
1.	gadol	gdola	‘big’
2.	kaʁ̞ov	kʁ̞ova	‘close’
3.	taluj	tluja	‘hung’
4.	ʃatul	ʃtula	‘planted’

It is easy to see that feminine adjectives are formed by adding an -a suffix to masculine adjectives. In addition, the first vowel /a/ is deleted in the feminine form. Thus, the UR of (1) ‘big.F’ is /gadola/. This vowel deletion process is common cross-linguistically, effectively shortening a three consonant-vowel sequence (C**V**CVCV → CCVCV). It can be described by the following formal rule:

(18-2) V→ Ø / C_CVCV

Let us examine two more adjectives:

(18-3) Additional Hebrew adjectives

	Masculine	**Feminine**	
5.	lavuʃ	lɛvuʃa	‘clothed’
6.	ʁ̞aχok	ʁ̞ɛχoka	‘far’

According to (18-1), the UR of ‘clothed.F’ and ‘far.F’ are /lavuʃ-a/ and /ʁ̞aχok-a/, respectively. In addition, we expect the first vowel to be deleted according to (18-2). Thus, the surface forms should have been [lvuʃa] ‘clothed.F’, and [ʁ̞χoka] ‘far.F’. At first glance, it might seem that first vowel /a/ was not deleted, but rather changed into [ɛ]. However, the motivation for a vowel change is unclear.

Alternatively, we may hypothesize that the vowel /a/ was indeed deleted, as in the other adjectives, and that [ɛ] is an epenthetic vowel inserted following the deletion of /a/. This hypothesis is in accordance with the observation that [ɛ] is the default epenthetic vowel in Hebrew (see more in the *planned* Chapter 22 (“Rule application and interaction”)). A possible motivation for vowel epenthesis in this case would be to break consonant sequences that are ill-formed in Hebrew. However, the sequences [lv] and [ʁ̥χ] are found in other Hebrew words, such as [hit͡sta**lv**ut] 'intersecting' and [ha**ʁ̥χ**ava] 'widening', respectively. Thus, it is not the sequence of consonants *per se* that triggers vowel epenthesis, and the motivation for the process remains obscure at this point. To explain this puzzling observation we need to look at these sequences on a higher level – the level of the syllable.

18.1. The syllable

A **syllable** is a sub-word structural unit that groups segments. A word may be composed of one or more syllables, each of which, groups one or more segments. For example, the word [bəluːn] 'balloon' is composed of two syllables: [bə] and [luːn]. In IPA transcriptions, syllables belonging to a single word are separated by a period (e.g., [bə.luːn]), as demonstrated in the SUPRASEGMENTALS section of the IPA sheet. The relation between segments, syllables, and the word level in [bə.luːn] is illustrated in Figure 18.1.

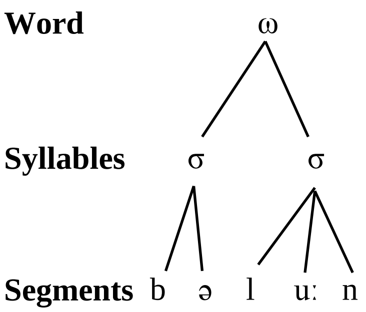


Figure 18.1 Segments, syllables, and word

A word consisting of a single syllable is a **monosyllabic** word (or a **monosyllable**), while a word with more than one syllable is a **polysyllabic** word. The table below provides the basic terms describing the number of syllables in a word:

Table 18.1 Number of syllables in a word

No. of syllables	**Term (adjective)**	**Example**	
1	Monosyllabic	dɔg	'dog'
2	Disyllabic / bisyllabic	bə.luːn	'balloon'
3	Trisyllabic	ek.sə.lənt	'excellent'
4	Quadrisyllabic	mæg.nɪ.fə.sənt	'magnificent'
5	Pentasyllabic	mæg.nɪ.fə.sənt.li	'magnificently'
6	Hexasyllabic	ɪn.vaɪ.rə.men.təl.ɪst	'environmentalist'

18.2. The psychological reality of the syllable

Syllabification (diving, or parsing, a word into syllables) is done rather intuitively and along similar principles across languages. However, the syllable is a theoretical concept. Speakers with no linguistic education don't necessarily know what a syllable is. In addition, unlike segments, which have rather clear acoustic correlates, such as the characteristic formant structure of vowels (Figure 18.2), there are no clear acoustic correlates of syllables.

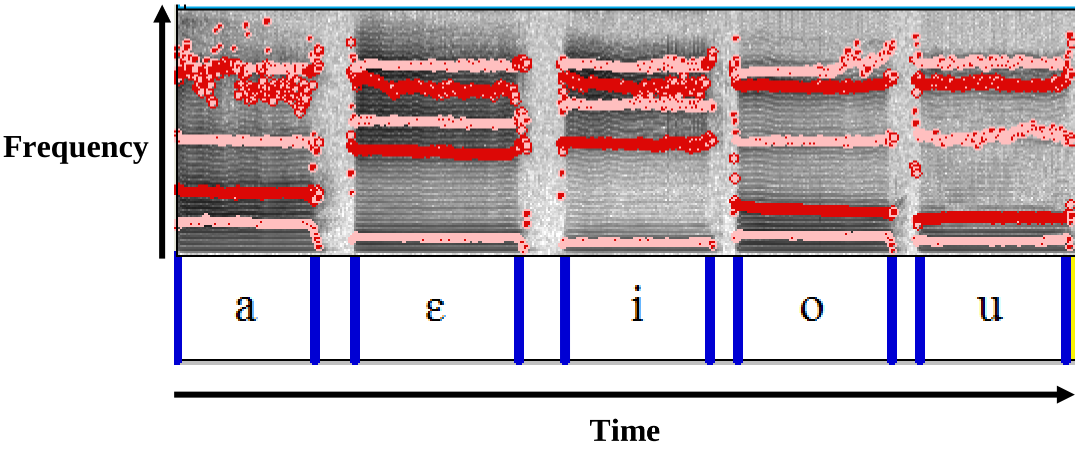


Figure 18.2 Typical formant structure of Hebrew vowels

Yet, despite the lack of direct acoustic evidence, there are other types of compelling evidence for the psychological reality of syllables ("psychological reality" is a phrase commonly used by linguists when arguing that a certain linguistic concept is not merely a theoretical idea, but a true part of the linguistic system of speakers). First, the most obvious evidence is the fact that speakers do seem to have intuitions about syllables. For instance, when asked to pronounce a word slowly, speakers tend to break it at syllable boundaries (e.g., the word *phonology* would be broken down into four fragments: [fə]-[nɑː]-[lə]-[d͡ʒi]).

Second, as suggested by the Hebrew data in (18-3), certain phonological processes are affected by the organization of segments into syllables (see more below). Third, the writing systems of some languages use **syllabary** – a set of symbols representing syllables, rather than individual segments. An example for such a writing system is Katakana, one of the alternative writing systems of Japanese, which includes symbols such as カ, which stands for the sequence /ka/, and キ, which stands for the sequence /ki/.

Finally, some language games may utilize syllable in a sophisticated way. For example, Filipino Tagalog-speaker teenagers have a language game called *Baligtad* ('inside-out, upsidedown, backward, inverted‘), which involves reordering of phonological elements. One variant of the game involves reordering of syllables, as demonstrated in (18-4) (French, 1988).

(18-4) Baligtad: a Tagalog language game

	Tagalog	Baligtad	
1.	pa.ŋit	ŋit.pah	‘ugly’
2.	ma.gan.dah	da.ma.gan	‘beautiful’
3.	ka.ma.tis	tis.ka.mah	‘tomato’

18.3. Syllable structure

Every syllable has one “important” segment – the **nucleus**. The nucleus is usually a vowel (but see below). Consonants preceding the syllable nucleus belong to the **onset** of the syllable, while consonants following the syllable nucleus belong to the **coda** of the syllable (*coda* is ‘tail’ in Italian). Every syllable has a nucleus, while the onset and coda are optional. For example, *undo* [ʌn.duː] is a disyllabic word, with no onset in the first syllable, and no coda in the second, as demonstrated in (18-5). A syllable that has a coda (i.e., ending with a consonant) is called a **closed syllable** (e.g., [ʌn] in [ʌn.duː] ‘undo’), while a syllable with no coda (i.e., ending with a vowel) is called an **open syllable** (e.g., [duː] in [ʌn.duː] ‘undo’).

(18-5) The syllable structure of *undo*

	Onset	Nucleus	Coda
Syllable 1	-	ʌ	n
Syllable 2	d	uː	-

The combination of the nucleus and coda is often called **rime** (or, **rhyme**), since this is the part of the syllable used for rhyming. For example, *b**ig*** and *d**ig*** rhyme, since they have identical rimes ([ɪɡ]), even though their onsets are different ([b] vs. [d]). By contrast, *big* and *bit*, which have identical nuclei (and onsets) but different codas, rhyme to a lesser degree, and so do *big* and *bug*, which have identical codas but different nuclei. The must general syllable structure is illustrated in Figure 18.3.

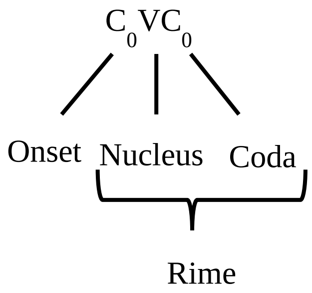


Figure 18.3 General syllable structure

The sequence C_0 in Figure 18.3 indicates zero or more consonants in either the onset or coda position of the syllable. When more than one consonant occupies the same syllabic position, we refer to such a sequence of consonants as a **consonant cluster** . A syllable onset containing a consonant cluster is called a **complex onset** (e.g., [kl] is a complex onset in [klæs] ‘class’). A syllable coda containing a consonant cluster is called a **complex coda** (e.g., [ns] is a complex coda in [dæns] ‘dance’).

18.4. Syllabification

As claimed above, syllabification is done rather intuitively and along similar principles across languages. All these principles are governed by one general principle, according to which segments are grouped into syllables according to their sonority. **Sonority** is roughly defined as ‘a degree of prominence assigned to a speech sound according to its intrinsic “loudness” or “airflow”’ (Trask, 1996). Accordingly, the various segments can be ordered following a **sonority hierarchy**, from vowels (the most sonorous segments) to plosives (the least sonorous segments). This is illustrated in Figure 18.4.

$$\underbrace{\text{Vowels} > \text{Glides} > \text{Liquids} > \text{Nasals}}_{\text{Sonorants [+son]}} > \underbrace{\text{Fricatives} > \text{Affricates, Plosives}}_{\text{Obstruents [-son]}}$$

Figure 18.4 The sonority hierarchy

Despite being rather elusive, sonority is a useful notion when it comes to syllabification. Data from various languages indicate that different classes of segments along the sonority hierarchy are subjected to different constraints on their positional distribution. In practice, the granularity of the hierarchy may change somewhat from language to language. For example, in some languages, all obstruents may be treated as having similar degree of sonority (see below). In other language, voiced obstruents might be considered as more sonorous than their voiceless counterparts (e.g., voiced plosive > voiceless plosive), and low vowels might be more sonorous than high vowels.

Syllabification, in general, follows (to a good approximation) a single principle, called the **sonority sequencing generalization (SSG)**, or **sonority sequencing principle (SSP)**. According to the more restrictive ("strong") version of this principle, the sonority level must rise from the syllable margins towards the nucleus, while according to the more permissive ("weak") version, the sonority level must not fall from the syllable margins towards the nucleus[62]. The two versions of SSG differ in their treatment of **sonority plateau** – a sequence of segments with the same sonority level (e.g., [kt], [sf]). Sonority plateaus are allowed by the weak SSG, but not by the strong SSG. SSG is defined independently for the syllable onset (going from the beginning of the syllable towards the nucleus) and coda (going from the end of the syllable towards the nucleus), as illustrated in Figure 18.5.

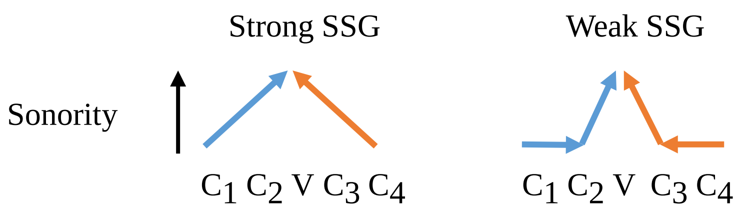


Figure 18.5 Sonority sequencing generalization (SSG)

Some languages follow the strong version of SSG (i.e., sonority decrease or plateau towards the syllable nucleus are disallowed). Other follow the weak version of SSG (i.e., sonority plateau within the syllable is allowed, but sonority decrease towards the syllable nucleus is disallowed). Yet, some languages allow violations of even the weak version of SSG (i.e., sonority may decrease towards the syllable nucleus).

For example, English follows the strong SSG in the onset position: a complex onset with rising sonority is allowed (e.g., [klæs] 'class'; Figure 18.6A), while a complex onset with sonority plateau (e.g., *[ktæs]) or fall (e.g., *[lkæs]) are disallowed.[63] On the other hand,

[62] Despite the fact that the SSG is defined at the syllable level, it is usually discussed only in the context of consonant clusters.

[63] In English, all obstruents have the same "effective" degree of sonority. Thus, a sequence of any two obstruents is disallowed in the syllable onset (e.g., *[kfas], *[pzam]). The only exception to this rule involves sequences of

English follows the weak SSG in the coda position, since it allows sonority plateau in the coda (e.g., [wɑːkt] 'walked'; Figure 18.6B). Hebrew follows the weak SSG in both onsets and codas, since sonority plateaus are allowed in both positions (onset: *כתומה* [ktu.ma] 'orange.F'; Figure 18.6C, coda: *בדקת* [ba.dakt] 'check.2SG.F.PST'). Russian is one of relatively few languages that allow falling sonority towards the syllable nucleus (e.g., ртуть [rtutʲ] 'mercury'; Figure 18.6D).

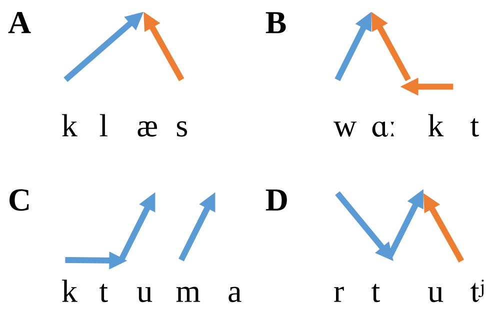


Figure 18.6 SSG and sonority contours. (A) Rising sonority in the onset satisfies the strong version of SSG (English: [klæs] 'class'). (B) Sonority plateau in the coda satisfies the weak version of SSG (English: [wɑːkt] 'walked'). (C) Sonority plateau in the onset satisfies the weak version of SSG (Hebrew: [ktu.ma] 'orange.F'). (D) Sonority decrease in the onset violates the SSG (Russian: [rtutʲ] 'mercury').

The most intuitive principle of syllabification derived from SSG is: 'one vowel per syllable', where a vowel can be either a monophthong (a single quality vowel like [a]) or a diphthong (a vowel with a changing quality, like [aɪ]). In other words, non-adjacent vowels cannot be in the same syllable. For example, [bɪ.liːv] is a valid syllabification of *believe*, while [b.ɪliːv] is an invalid syllabification, since it contains two non-adjacent vowels ([ɪ] and [iː]).

Generally, in words composed of several consonant-vowel sequences, each of the consonant-vowel sequences constitutes a syllable (e.g., [pə.d͡ʒɑː.mə] 'pajama'). This is a thumb rule, which I call "'yes onset' is better than 'no onset'" (as we shall see in the next chapter, English seems to deviate from this rule in certain cases).

Variation in syllabification occurs mainly in cases of mid-word consonant sequences or vowel sequences. Syllabification in such cases is guided primarily by SSG. For example, [ʃɛl.tɚ] is a plausible syllabification of 'shelter', while [ʃɛ.ltɚ] is implausible, since the sequence [lt] in the onset of the second syllable violates SSG. On the other hand, if there are

voiceless stridents ([s] and [ʃ]) plus a plosive or a fricative in the syllable onset (e.g., [spɔːɹt] 'sport', [sfɪɹ] 'sphere', and [ʃtɪk] 'shtick'). This anomalous behavior of voiceless stridents, which is observed in other languages as well, poses a challenge to phonological theory.

two alternative syllabifications that satisfy the SSG, both are theoretically acceptable. Some knowledge of the language is needed to determine the correct syllabification (if there is one). For example, [sɛt.lɚ] and [sɛ.tlɚ] are both possible syllabifications of ‘settler’ (though, the first seems to be preferred).

Variation also exists in the case of vowel sequences. However, the problem of vowel sequences is partly an issue of transcription conventions. Sequences of non-identical vowels (e.g., [ai]) can either belong to the same syllable or to two separate syllables. In the former case, the sequence of vowels is called a **diphthong**. A diphthong is essentially a single vowel that begins with one quality and ends with another. Thus, both parts of the diphthong belong to the nucleus of the same syllable (a diphthong can be considered a **complex nucleus**). Combining the two vowels under a tie-bar (e.g., [a͡ɪ]) indicates that they are part of a single segment.

However, tie-bars are usually not used to mark diphthongs. Consequently, non-identical vowel sequences in linguistic data are generally ambiguous. For example, native speakers of English typically perceive *bite* a monosyllable containing a diphthong [baɪt]. By contrast, native speakers of Hebrew typically perceive the similarly sounding word *בית* [bait] ‘house’ as disyllabic, with adjacent vowels in separate syllables (i.e., [ba.it]). Such a sequence of adjacent vowels that belong to separate syllables is called a **hiatus**. Since some languages prefer diphthongs over hiatus, while others prefer hiatus over diphthongs, it is generally not possible to determine how non-identical vowel sequences should be treated just by looking at the data.

Similarly, a sequence of identical vowels can be either a single long vowel in the nucleus of a single syllable, or a hiatus involving two identical vowels in separate syllables. Often, long vowels are marked as such to avoid ambiguity (e.g., [bɪ.liːv] ‘believe’). However, in some datasets, long vowels are also written as a sequence of two vowels (e.g., [bɪ.liiv] ‘believe’). Since some languages have long vowels and no hiatus while other languages have hiatus and no long vowels, some knowledge of the target language is required in order to decide how vowel sequences should be analyzed.

While the nucleus of the syllable is usually a vowel, some languages allow syllables with no vowels. In such cases, the most sonorous consonant of the syllable occupies the nucleus. A consonant in the nucleus of a syllable is called a **syllabic consonant**. The existence of syllabic consonants lead to a revised definition of the syllable nucleus as the ‘sonority peak of the syllable’. In some dialects of English, nasals and liquids can function as syllabic consonants. For example, the pronunciation of *button* in some English dialects is [bʌtn], which violates SSG if treated as monosyllabic. To avoid this violation, the word is treated as disyllabic with a syllabic nasal consonant in the second syllable (i.e., [bʌt.n̩] or [bʌ.tn̩]). The little vertical line under the consonant marks as a syllabic consonant. Interestingly, some languages even allow words with no vowels at all. For example, Czech and Slovak have a famous tongue twister: *strč prst skrz krk* [str̩t͡ʃ pr̩st skr̩s kr̩k] 'stick a finger through the throat'.[64] All the words in the phrase are monosyllables with a syllabic [r] at the nucleus.

18.5. Syllable types

Each language has a list of allowed syllable types. To determine the allowed syllable types according to a given dataset, we divide words into syllable and, then, replace each vowel with a "V" and each consonant with a "C". Rewriting words in this form allows us to see the various types of syllable existing in the language. Note that all diacritics expect for the long symbol (e.g., [aː]) and the syllabicity mark (e.g., [n̩]) should be suppressed, and consonants linked by a tie-bar should be represented as a single consonant. Thus, each onset and coda consonant is represented by a single "C", and we distinguish four types of nuclei: (i) a short vowel (e.g., *bit* [bɪt] → C**V**C), (ii) a long vowel (e.g., *beat* [biːt] → C**V**ːC), (iii) a diphthong (e.g., *bite* [baɪt] → C**VV**C), and (iv) a syllabic consonant (e.g., *button* [bʌ.tn̩] → CV.C**C̩**).

In the following exercise, a partial inventory of the possible syllables of English is determined based on a sample of words. First, the words on the list are syllabified, and the syllable structure of each word is determined as explained above (18-6). Then, the different syllable types are summarized (18-7). Each type is accompanied by an example from the data. The relevant syllable in each word is highlighted for clarity.

(18-6) Syllable structure analysis of English words

	Word (syllabified)		**Syllable structure**
1.	bə.luːn	'balloon'	CV.CVːC
2.	ʃɛl.tɚ	'shelter'	CVC.CV
3.	klæs	'class'	CCVC
4.	ɔl.təɹ	'alter'	VC.CVC
5.	baɪt	'bite'	CVVC
6.	ə.hɛd	'ahead'	V.CVC
7.	dæns	'dance'	CVCC
8.	wɑːkt	'walked'	CVːCC
9.	bʌ.tn̩	'button'	CV.CC̩

[64] https://en.wikipedia.org/wiki/Str%C4%8D_prst_skrz_krk

(18-7) A partial syllable inventory of English

Type	Word	Example	Type	Word	Example
CV	1	**bə**.luːn	CVVC	5	baɪt
CVːC	1	bə.**luːn**	V	6	**ə**.hɛd
CVC	2	**ʃɛl**.tɚ	CVCC	7	dæns
CCVC	3	klæs	CVːCC	8	wɑːkt
VC	4	**ɔl**.təɹ	CC̩	9	bʌ.**tn̩**

18.6. Syllables and phonological processes

In the last section of this chapter, we return to the puzzle that motivated the initial discussion of the chapter, namely phonological processes related to syllable structure.

18.6.1. Processes motivated by well-formedness of syllable structure

As explained in 18.4, syllabification is guided by SSG, such that sequences of consonants that potentially violate SSG, will be assigned to separate syllables. However, in some cases, there is no way to parse a word into syllables without violating the SSG. In such cases, certain phonological processes may apply to correct the structure of the word. For example, recall the Hebrew adjectives shown at the beginning of the chapter and repeated here:

(18-8) Hebrew adjectives

1.	gadol	gdola	'big'
2.	kaʁ̥ov	kʁ̥ova	'close'
3.	taluj	tluja	'hung'
4.	ʃatul	ʃtula	'planted'
5.	lavuʃ	lɛvuʃa	'clothed'
6.	ʁ̥aχok	ʁ̥ɛχoka	'far'

According to the pattern observed in words 1-4, the feminine form of words 5-6 should have been *[lvuʃa] 'clothed.F', and *[ʁ̥χoka] 'far.F', respectively. However, the word-initial sequences of consonants in these words have a falling sonority contour, which would lead to a violation of SSG, if parsed as a complex onset, as illustrated in Figure 18.7A-B. The potential violation of SSG is avoided by vowel epenthesis that breaks the problematic consonant cluster. The epenthesis is followed by **resyllabification**, i.e., reassigning segments to syllables, in a way that satisfies SSG (e.g., /lvu.ʃa / → *[lɛvu.ʃa] → [lɛ.vu.ʃa] 'clothed.F'). This is illustrated in Figure 18.7C-D.

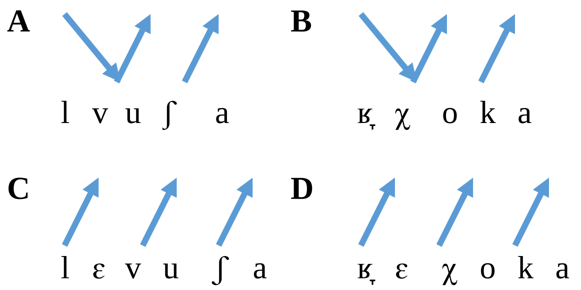


Figure 18.7 Vowel epenthesis motivated by syllable structure in Hebrew. Consonant clusters that violate SSG in (A) /lvuʃa/ 'clothed.F' and (B) /ʁ̥χoka/ 'far.F' are modified by vowel epenthesis followed by resyllabification in (C) [lɛ.vu.ʃa] 'clothed.F' and (D) [ʁ̥ɛ.χo.ka] 'far.F'.

18.6.2. **Processes applying to specific syllabic positions**

In addition to phonological processes motivated by syllable structure, there are processes that do not affect syllable structure, but are sensitive to syllabic position. Thus, some processes affect only consonants in the syllable onset, while others affect only consonants in the syllable coda. The following dataset illustrates **syllable-final devoicing** in Low German, a dialect of German spoken in Northern Germany. The data shows that the voiced obstruent /d/ is devoiced only when in the syllable coda (a word-final consonant is a special case of coda consonant).

(18-9) Syllable-final devoicing in Low German (Gussenhoven & Jacobs, 2017)

UR	**PR**	
kɪnd	kɪnt	'child'
kɪnd-ɪʃ	kɪn.dɪʃ	'childish'
kɪnd-lɪç	kɪnt.lɪç	'childlike'

In formal notation, the rule can be written as:

(18-10) $C_{[-son]} \rightarrow [-voice] \,/\, \_\,]_{\sigma}$, where $]_{\sigma}$ marks the end of the syllable.

The "mirror image" of this process is **syllable-initial voicing**, involving the voicing of syllable-initial voiceless obstruents. This is process can be observed in the following data from Mohawk, an Iroquoian language spoken by the Mohawk people in parts of the United States and Canada. Note that underlying (non-glottal) voiceless obstruents surface as voiced only when in the syllable onset.

(18-11) Syllable-initial voicing in Mohawk (Halle & Clements, 1983)

	UR	**PR**	
1.	je-k-hrek-s	jek.reks	‘I push it’
2.	hra-jʌ̃tho-s	ra.jʌ̃t.hos	‘he plants’
3.	hra-kʌ̃-s	raː.gʌ̃s	‘he sees her’
4.	ʌ̃-k-hʌ̃te-ʔ	ʌ̃k.hʌ̃ː.deʔ	‘I shall go ahead’
5.	k-esak-s	geː.zaks	‘I look for it’
6.	hra-ket-as	ra.geː.das	‘he scrapes’

In formal notation, the rule can be written as:

(18-12) $C_{[\text{-son}]} \rightarrow [\text{+voice}]\ /\ {}_{\sigma}[\ \_$, where ${}_{\sigma}[$ marks the beginning of the syllable.

Chapter 19 Stress

19.1. Segmental contrasts

In Chapter 14, I introduced the concept of a *phoneme*, as an abstract segment – the mental representation of speech sound. One of the main motivations behind the introduction of the concept was the fact that different phonemes of the same language (or, dialect) *contrast* with each other. That is, replacing one phoneme with another in a given word can create a different word. For example, the list of words in (19-1) demonstrate phonemic contrasts by changing the first consonant.

(19-1) Phonemic contrasts in English

1.	[miː]	'me'
2.	[biː]	'be'
3.	[piː]	'pea'
4.	[tiː]	'tea'

In Chapter 15, I introduced the notion of the *distinctive feature*. Accordingly, the data in (19-1) can be viewed not as contrasts of phonemes, but rather as contrasts of distinctive features. Each word on the last differs in a single feature from the word above and below it. The difference between word 1 and 2 is in manner of articulation (i.e., nasal, or [+nas], versus plosive, or [-nas]). The difference between word 2 and 3 is in voice (i.e., voiced, or [+voice] versus voiceless, or [-voice]). Finally, the difference between word 3 and 4 is in place of articulation (i.e., bilabial, or [Lab] versus alveolar, or [Cor]).

19.2. Stress

The data in (19-1) demonstrates **lexical contrasts** (i.e., meaning differences) achieved through segmental contrasts. However, lexical contrasts can also be generated on prosodic levels. Consider the word *object*. This word can function either as a noun, or as a verb. In spoken English, the exact use of the word can be determined by the way it is pronounced. Figure 19.1 illustrates some acoustics properties of *object* as a noun (A) and as a verb (B), in British pronunciation according to the online Cambridge dictionary.[65] The diagram in each panel includes the pitch (blue curve) and intensity (yellow curve) contours. The position and value of the maximal point is marked on each curve. In addition, the boundary between syllables is

[65] https://dictionary.cambridge.org/dictionary/english/object

marked below each acoustic diagram, and each syllable is transcribed. Finally, the duration of each syllable in seconds is indicated at the bottom of the diagram.

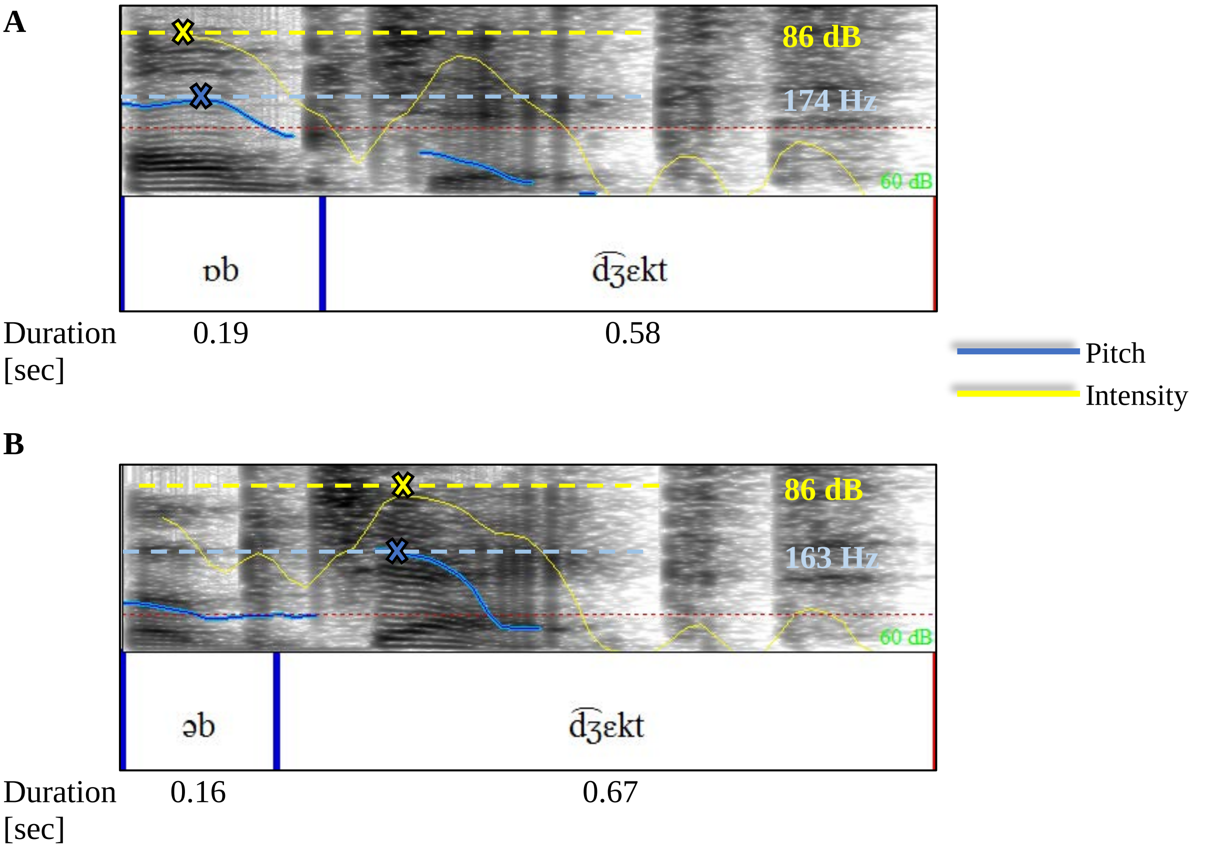


Figure 19.1 Stress contrasts in English. (A) *object* (noun), (B) *object* (verb)

Several differences become apparent when comparing the two figures. The first syllable in the noun is pronounced on a higher pitch (mean: 150 Hz) than the second syllable (mean: 81 Hz). In particular, the maximum point of the pitch curve is reached during the first syllable. The same is true, to a lesser extent, for the intensity contour: the loudest point occurs during the articulation of the first syllable. The properties of the pitch and intensity contour are reversed in the verb: here, the overall pitch level is higher on the second syllable (mean: 123 Hz) than on the first syllable (mean: 98 Hz), and both maximum points are reached during the second syllable.

In addition to the pitch and intensity differences, we can see that, the second syllable is longer than the first syllable in both diagrams. This is not surprising since the second syllable contains four segments, while the first syllable contains only two. However, the relative difference between the syllables is different for the two words. The second syllable is four times longer than the first syllable in the verb, compared to a 3:1 ratio in the noun.

Overall, we observe that, in the noun, the first syllable is relatively longer, louder and higher-pitched (compared to the second syllable), and the same is true for the second syllable of the verb (compared to the first syllable). These observations can be summarized by saying that the first syllable of the noun and the second syllable of the verb are stressed.

Generally, **stress** is a certain type of prominence of syllables. That is, a **stressed syllable** is perceptually more prominent than other surrounding syllables, which are said to be **unstressed**. In IPA transcription, stress is marked by a small superior vertical line at the beginning of the stressed syllable (the *primary stress* diacritic found in the SUPRASEGMENTALS section of the IPA sheet). For example, the pronunciation of *object* as a noun (stress on the first syllable) is transcribed: [ˈɒb.d͡ʒɛkt], while the pronunciation of *object* as a verb (stress on the second syllable) is transcribed: [əb.ˈd͡ʒɛkt].

19.3. Acoustic correlates of linguistic stress

Following the analysis of Figure 19.1, we can state the general acoustic properties of stress, as follows: stressed syllables tend to be: (i) louder, (ii) longer, and (iii) higher-pitched (or on a rising pitch contour) compared to unstressed syllables. Of the three characteristics, the last two are the most reliable acoustic properties of stress. Languages may use one or both of them to realize stress. Languages that have phonemic vowel length (i.e., long and short vowels can create minimal pairs) tend to rely more on pitch differences to realize stress.

It should be noted that determining stress based on the acoustic properties listed above is not straightforward. Because of physiological factors, pitch is affected not only by stress. First, initiating pitch requires effort from the vocal cords. As a result, pitch tends to be high at the beginning of voiced sections (beginning of word, after voiceless segment) and then drops to a lower level. Second, maintaining vocal cord vibrations also requires effort. As a result, pitch tends to decrease gradually over a stretch of continuous speech. For these reasons, pitch tends to be high on the first syllable and low on the last syllable, regardless of stress (see more in the *planned* Chapter 21 (“Intonation”)).

One can determine that a non-initial syllable is stressed if pitch decreases towards the end of the first syllable and then increases or jumps towards the middle of the second syllable. For example, in the recording of the noun *object* demonstrated above, pitch level is high until the middle of the first, stressed syllable. Then, pitch decreases throughout the second, unstressed syllable. In the verb *object*, pitch level is low on the first, unstressed syllable. Then, the pitch level jumps at the beginning of the second, stressed syllable, before it starts decreasing.

Duration also depends on factors other than stress. First, it depends on the number of segments (a syllable with 3 segments will likely be longer than a syllable with 2 segments). Second, syllables tend to be longer at the end of speech streams (e.g., end of sentence). In a minimal pair of words that differ by the position of stress, we compare the relative durations of syllables across words. In the example of the *object* noun-verb pair, the durations of syllables in the noun are: syllable 1 = 0.19 sec, syllable 2 = 0.58 sec, and the ratio *duration of syllable 1* / *duration of syllable 2* is 0.33. In the verb *object*: syllable 1 = 0.16 sec, syllable 2 = 0.67 sec, and the ratio *duration of syllable 1* / *duration of syllable 2* = 0.24. The higher the ratio *duration of syllable 1* / *duration of syllable 2*, the longer syllable 1 is (i.e., more stressed) compared to syllable 2. The lower the ratio, the longer syllable 2 is (i.e., more stressed) compared to syllable 1. Therefore, in the current example, the first syllable is longer relative to the second syllable in the noun compared to the verb. Therefore, the first syllable is more stressed in the noun compared to the verb.

A comparison of the transcriptions of the noun-verb *object* ([ˈɒb.d͡ʒɛkt] versus [əb.ˈd͡ʒɛkt]) indicates another aspect related to stress, namely a difference in the quality of vowels. Although not indicated by the written form, the first letter ⟨o⟩ is pronounced as different vowels in the two words. In particular, it is pronounced as [ə] in the verb, when unstressed. Pronouncing unstressed vowels as [ə] is a prominent feature of English. This pronunciation is the result of **vowel reduction**, a phonological process whereby an underlying unstressed vowel receives a “neutral” quality.

Vowel reduction is found in many languages, though, the exact manifestation may change across languages. In English, unstressed vowels are typically reduced to [ə] (sometimes to [ɪ]), although not all unstressed vowels are affected, as can be seen in the transcription of the noun *object*, in which the second, unstressed vowel is not reduced to [ə]. Another language that uses vowel reduction is Russian, in which unstressed /o/ and/a/ are reduced to [ɐ] in the first pretonic syllable (i.e., before a stressed syllable) and to [ə] in other unstressed syllables (Asherov et al., 2016). For example, хорошо ‘good’ has three underlying /o/ vowels, but only the last of them is pronounced as such, since it is stressed. The other two are reduced as explained above: [xərɐˈʂo].

19.4. Stress placement

Cross-linguistically, stress tends to be close to word edges. In some languages, it is placed close to the end of the word, while in others, it is closer to the beginning. Table 19.1 lists the common terms describing stress patterns according to the position of the syllable bearing the stress.

Table 19.1 Names of stress patterns by position[66]

Name	**Position (syllable)**	**Pattern**	**Example**
Relative to word end			
Ultimate (final)	Last	σσσσσ́	Hebrew: [mɛ.ha.jɛ.la.ˈdim] ‘from the children’
Penultimate	Second-to-last	σσσσ́σ	Hebrew: [ha.klɛ.man.ˈti.not] ‘the clementines’
Antepenultimate	Third-to-last	σσσ́σσ	Macedonian: [ko.li.ˈt͡ʃest.vo.to] ‘the quantity’
Relative to word beginning			
Initial	First	σ́σσσσ	Pintupi: [ˈpu.liŋ.ka.la.tʲu] ‘we (sat) on the hill’
Pen-initial	Second	σσ́σσσ	Shiwilu: [i.ˈkɘr.mu.tuʔ.lɘk] ‘I have a headache’

Typologically, languages in which syllables can be characterized as stressed or unstressed can classified into three groups.[67] One group includes languages in which the position of stress is fixed. For example, in Hungarian, stress is assigned to the first syllable in all polysyllabic words. By contrast, in French, stress always falls on the last syllable. A second group of languages (e.g., Latin, Classical Arabic) have a ‘regular stress rule’. The position of stress in these languages is not fixed. However, it can be predicted by a rule that assigns stress to a certain syllable based on phonological properties (e.g., the presence of a long vowel or coda consonants). Finally, some languages (e.g., Hebrew, English) are said to have *lexical stress*. In these languages, the position of the stress is largely unpredictable. That is, speakers of these languages need to learn the stress pattern for individual words.

Among the languages with a fixed stress pattern, ultimate- and initial-stress are the most common patterns. Cross-linguistic surveys (e.g., Gordon, 2002) have found that each of these

[66] References for examples: Macedonian (Kenstowicz & Kisseberth, 1979), Pintupi (Kager, 2007), Shiwilu (Gussenhoven & Jacobs, 2017).

[67] In the next chapter (*planned*), I discuss classes of languages that do not use stress.

patterns characterizes the stress system of at least 30% of the surveyed languages. These patterns are followed by penultimate stress, which is the fixed stress pattern in 25-29% of the languages. Finally, peninitial (4-5.5% of the languages) and antepenultimate (2-3.5% of the languages) are the rarest stress patterns.

19.5. Stress as a distinctive feature

In some languages, stress position can be used as a distinctive feature (i.e., to differentiate the meaning of words), similarly to segmental features. For example, English has minimal pairs of disyllabic words in terms of stress position: in these pairs, ultimate stress usually indicates that the word is used as a verb, while penultimate stress usually marks it as a noun or adjective. Some examples for such minimal pairs are listed in (19-2).

(19-2) Stress in minimal pairs in English

	Ultimate		**Penultimate**	
1.	əb.ˈd͡ʒɛkt	‘object (verb)’	ˈɒb.d͡ʒɛkt	‘object (noun)’
2.	pɚ.ˈmɪt	‘permit (verb)’	ˈpɝː.mɪt	‘permit (noun)’
3.	kən.ˈtɛnt	‘content (adj)’	ˈkɑːn.tɛnt	‘content (noun)’
4.	ɹ̠ɪ.ˈkɔːɹ̠d	‘record (verb)’	ˈɹ̠ɛk.ɚd	‘record (noun)’

In the last example, we see a peculiar property of English. The assumed syllable structure of *record* (noun) is CVC.VC. This contradicts the “universal” observation, according to which, syllables with an onset are preferred over syllables without an onset (see Section 18.4). The explanation seems to be that in English (at least in some dialects), a short stressed vowel “attracts” coda consonants, while a long, stressed vowel does not (Treiman & Zukowski, 1990). Thus, [ˈhæmɚ] ‘hammer’ is parsed as [ˈhæm.ɚ], since the short stressed [æ] attracts the following [m]. By contrast, the long stressed [iː] in *cheesy* does not attract a coda (i.e., [ˈtʃiː.zi]).

It is also noteworthy that the placement of stress may vary across dialects of English. For example, in some dialects *research* (noun) is pronounced with penultimate stress (e.g., [ˈɹiː.sɝː.t͡ʃ]), while in others it is pronounced with ultimate stress (e.g., [ɹɪ.ˈsɜː.t͡ʃ]).

Another contrastive use of stress position is found in Modern Hebrew, in which some disyllabic given names, derived from content words, are distinguished from their origin by a stress shift. In these minimal pairs, the original content word (noun, verb, or adjective) has ultimate stress, while the derived given name is pronounced with penultimate stress.

(19-3) Stress in minimal pairs in Hebrew

	Ultimate		Penultimate	
1.	mɛ.ˈiʁ	‘enlightening’	ˈmɛ.iʁ	‘Meir’ (male name)
2.	pni.ˈna	‘pearl’	ˈpni.na	‘Pnina’ (female name)
3.	χa.ˈim	‘life’	ˈχa.im	‘Chaim/Haim’ (male name)

Note that the stress shift in these cases is a feature of colloquial Hebrew. Some people bearing any of these names may prefer to be addressed via the ultimate stress version. Using the ultimate stress version is also more common in formal social situations. This means that the ultimate stress word is essentially ambiguous between the content word and the derived given name, while the penultimate stress version refers unambiguously to the given name.

19.6. Degrees of stress and stress rules

Stress is essentially a quantitative phenomenon. Thus, a syllable can receive various amounts of stress in pronunciation. Some languages use this fact to distinguish three degrees of stress: a syllable might receive primary stress, secondary stress, or no stress. **Primary stress** is the highest degree of stress. Only one syllable in a word can have primary stress. **Secondary stress** is a lesser degree of stress (e.g., in terms of pitch level or duration). In languages with secondary stress, multiple syllables in a single word can have secondary stress. Secondary stress usually appears on alternating syllables within a word (e.g., the first, third, and fifth). Finally, syllables without primary or secondary stress are said to be **unstressed**.

In IPA, secondary stress is marked by an inferior vertical line at the beginning of the syllable. Certain long words in English are said to have a syllable bearing secondary stress, in addition to the primary stressed syllable. For example, in *California*, the penultimate syllable bears the primary stress, while the first syllable bears secondary stress [ˌkæl.ə.ˈfɔːr.njə].

In languages with a predictable stress pattern it is possible to write a formal rule that determines the position of stress.

19.6.1. Primary stress

Stress tends to be closer to one edge of the word (beginning or end). However, determining the relevant edge is not always easy task. For example, if the given data contains only disyllabic words with the first syllable stressed, it could be interpreted as either initial stress or penultimate stress, both are common cross-linguistically. Thus, in order to determine the relevant edge for stress placement we need long words (typically, 4-5 syllables). For example, consider the following data from Polish.

(19-4) Stress in Polish (Newlin-Łukowicz, 2012)

ˈgrɔʃ	‘penny’
ˈdvuj.ka	‘two’
rɛ.ˈpɔr.tɛr	‘reporter.NOM’ (“reporter”)
rɛ.pɔr.ˈtɛ.ra	‘reporter.GEN’ (“of reporter”)
rɛ.pɔr.tɛ.ˈrɔ.vi	‘reporter.DAT’ (“to reporter”)

The first word in (19-4) is monosyllabic and stressed. The other words have two syllables or more and, in all of them, stress is on the penultimate syllable. Thus, we may conclude that Polish has a fixed penultimate stress in polysyllabic words. Formally, we can write the following stress rule for Polish polysyllabic words:

(19-5) $\sigma \rightarrow$ [+stress] / $\_\sigma]_{\omega}$, where σ stands for a syllable, and $]_{\omega}$ marks the end of the word.

Thus, (19-5) says that stress is assigned to a syllable when followed by another syllable and word boundary. Similarly, the stress rule for monosyllabic words is:

(19-6) $\sigma \rightarrow$ [+stress] / $\_]_{\omega}$

The two rules can be combined using the parenthesis notation. When part of the environment in a phonological rule is put between parentheses, the rule has two modes of application. First, the parentheses are removed, sparing their content. If the given word contains the environment for the rule minus the parentheses, the rule applies. If the environment is not suitable for the rule, the content of the parenthesis is excluded from the rule to enable the application of the rule. Using this notation, we can combine (19-5) and (19-6):

(19-7) $\sigma \rightarrow$ [+stress] / $\_(\sigma)]_{\omega}$

This should be read as follows: the parentheses are removed as a first step. In other words, (19-5) is attempted first. It applies when the word contains a syllable which is followed by another syllable and word boundary (i.e., polysyllabic words). When these conditions are not met, that is, when the word does not contain a syllable which is followed by another syllable and word boundary (i.e., monosyllabic words), then the content of the parenthesis is excluded from the rule, and (19-6) applies.

In a similar fashion, we can write a formal rule that assigns stress to the second syllable (i.e., peninitial stress) in polysyllabic words and to the first (and only) syllable in monosyllabic words:

(19-8) $\sigma \rightarrow$ [+stress] / $_{\omega}$[_(σ) , where $_{\omega}$[marks the beginning of the word.

19.6.2. **Primary and secondary stress**

While only one syllable can bear the primary stress of a word, there can be multiple syllables bearing secondary stress. Secondary stress usually applies iteratively to alternating syllables. For example, the following data illustrates the stress pattern of Warao, a language isolate spoken in Venezuela.

(19-9) Primary stress and secondary stress in Warao (Howard, 1972)

1.	ji.ˌwa.ra.ˈna.e	‘he finished it’
2.	ˌna.ho.ˌro.a.ˌha.ku.ˈta.i	‘the one who ate’
3.	e.ˌna.ho.ˌro.a.ˌha.ku.ˈta.i	‘the one who causes him to eat’
4.	ˌja.pu.ˌru.ki.ˌta.ne.ˈha.se	‘verily to climb’

Like Polish, Warao places primary stress on the penultimate syllable. In addition, every other syllable from the penultimate to the beginning of the word (i.e., fourth, sixth, or eighth syllable when counting from the end) receives secondary stress. To distinguish the different degrees of stress, we use the following notation:

Degree	**Unary feature**	**Binary feature**
Primary stress	[1stress]	[+stress]
Secondary stress	[2stress]	[+stress]
Unstressed	[Østress]	[-stress]

Consequently, the primary stress rule for Warao is similar to the stress rule for Polish, with [+stress] substituted for [1stress]:

(19-10) $\sigma \rightarrow$ [1stress] / _σ]ω

The secondary stress rule then assigns secondary stress to every syllable followed by a sequence of another syllable plus any stressed syllable (either primary or secondary stress):

(19-11) $\sigma \rightarrow$ [2stress] / $\_\sigma\sigma_{[+\text{stress}]}$

Note that the stress pattern in Warao allows us to determine that sequences of vowels are divided into different syllables (i.e., hiatus), since that treating them as diphthongs would result in secondary stress on consecutive syllables (e.g., the sequence ˌroaˌha in (2) ˌnahoˌroaˌhakuˈtai).

19.7. Compound and phrasal stress

Although stress is a word-level feature, it can interact with higher linguistic levels. Thus, in certain languages, the morpho-syntactic status of words can affect their stress pattern. Isolated words in English have the default lexical stress pattern, with one syllable bearing primary stress (e.g., [ˈblæk] 'black', [ˈbɔɹd] 'board'). However, when words are combined in phrases or compounds the default stress pattern is modified (Chomsky & Halle, 1968).

In multi-word phrases, the head of the phrase receives the default stress, which is referred to as **phrasal (nuclear) stress**. The stress level of other words is weakened. For example, in [ˌblæk ˈbɔɹd] 'black board', *board* is the head of the phrase and, thus, receives its default primary stress. By contrast, *black* is a dependent of *board* and, thus, its stress degree is weakened relative to *black* in isolation (i.e., it receives secondary stress in the phrase *black board*).

When words are combined to form a new single word (i.e., a compound), primary stress is assigned to the first component (not the head!). This is called **compound stress**. By contrast, the stress of the head is weakened even below secondary stress and the syllable becomes unstressed. For example, in *blackboard,* which is a compound derived from *black board*, the dependent receives primary stress, while the head is unstressed (i.e., [ˈblæk.bɔɹd]).

19.8. Practice

1. **Araucanian** (Chile) (Echeverría & Contreras, 1965)

1.	wuˈle	‘tomorrow’
2.	t̪iˈpanto	‘year’
3.	eˈlumuˌju	‘give us’
4.	eˈluaˌenew	‘he will give me’
5.	kiˈmufaˌluwuˌlaj	‘he pretended not to know’

i. Divide the words into syllables.

ii. Determine the allowed types of syllables in Araucanian, classify them into open and closed syllables and give an example for each type.

iii. Describe the position of the primary and secondary stress.

<u>Solution</u>

i. Syllabification:

1.	wu.ˈle	‘tomorrow’
2.	t̪i.ˈpan.to	‘year’
3.	e.ˈlu.mu.ˌju	‘give us’
4.	e.ˈlu.a.ˌe.new	‘he will give me’
5.	ki.ˈmu.fa.ˌlu.wu.ˌlaj	‘he pretended not to know’

ii. Syllable types:

Type	**Word**	**Open/closed**	**Example**
CV	1	Open	**wu**.ˈle
CVC	2	Closed	t̪i.ˈ**pan**.to
V	3	Open	**e**.ˈlu.mu.ˌju

iii. Primary stress: pen-initial (second syllable).

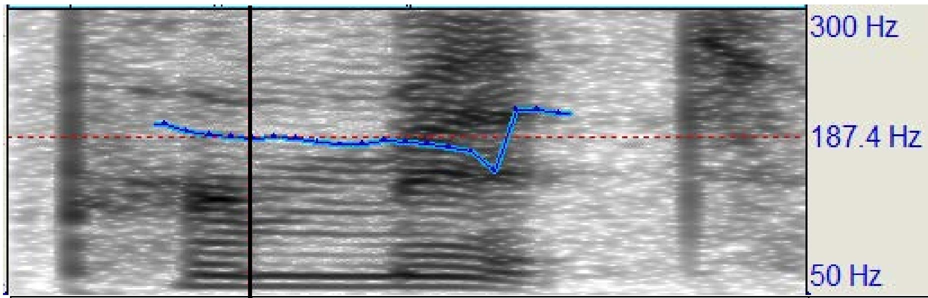
Secondary stress: alternating even syllables (fourth, sixth…)

Note: secondary stress can help us determine that sequences of vowels (e.g., [ua] in eˈluaˌenew) should be divided into separate syllables. That is, the language prefers a hiatus over a diphthong. If we analyze vowel sequences as diphthongs then primary and secondary stress would assign to adjacent syllables, which is less common.

2. **Contrastive stress in English**

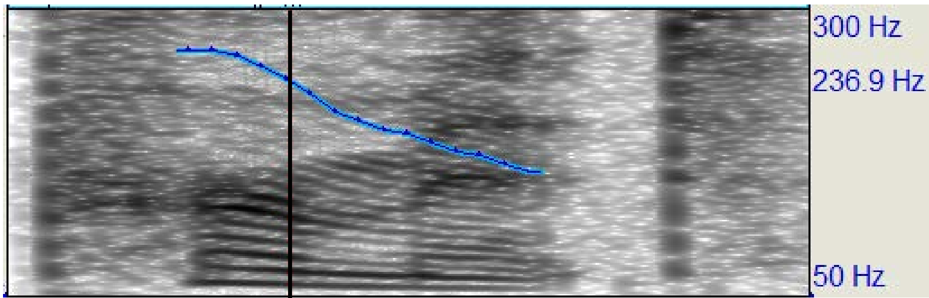
English has some minimal pair words with respect to stress position: (1) a disyllabic word with ultimate stress, which functions as a verb, and (2) a disyllabic word with penultimate stress, which functions as a noun or an adjective. The charts below illustrate the spectrogram and pitch contour (blue curve) of such a minimal pair for the written form *permit*: (1) [pɚˈmɪt] ‘permit (verb)’, and (2) [ˈpɝmɪt] ‘permit (noun)’. Analyze the figures and determine which is more likely to represent which word.

A.



	per	mit
Duration [sec]	0.17	0.38

B.



	per	mit
Duration [sec]	0.18	0.32

<u>Solution</u>

Figure A represents (1) [pɚˈmɪt] ‘permit (verb)’ and Figure B represents (2) [ˈpɝmɪt] ‘permit (noun)’.

<u>Explanation</u>: (i) *Duration*: the ratio of durations syllable 1/syllable 2 is higher in Figure B (0.18/0.32 = 0.56) compared to Figure A (0.17/0.38 = 0.45). This suggests that, in Figure B, syllable 1 is more stressed (like in the noun) and that in Figure A, syllable 2 is more stressed (like in the verb).

(ii) *Pitch*: in Figure B, pitch is high until the middle of the first syllable and then drops throughout the second syllable. This suggests that syllable 1 is more stressed (like in that noun). In Figure A, the pitch is relatively flat, with a small jump in the middle of the second syllable. Thus, it is likely that in Figure A, syllable 2 is more stressed (like in the verb).

Phonology review

This is a review exercise for Parts IV and V of the book. The data used in the exercise is a made-up language dubbed “Jupiterian”. Using an artificial data set makes it possible to cover a wider range of phenomena than what is normally found in a single language.

Jupiterian (made-up language)

A	**Singular**	**Plural**	
1.	æm	zeŋæm	‘man’
2.	ɡir	zeŋɡir	‘ring’
3.	ken	zeŋken	‘neck’
4.	pihsiʃ	zempihsiʃ	‘spaceship’
5.	bɑk	zembɑk	‘club’
6.	feːʃ	zemfeːʃ	‘chef’
7.	deh	zendeh	‘head’
8.	slɑk	zenslɑk	‘class’
9.	tɑh	zentɑh	‘hat’
10.	trɑ	zentrɑ	‘heart’
11.	ʃif	zenʃif	‘fish’

1. List the allomorphs of the plural prefix and state their environments.

2. What is the underlying representation (UR) of the plural prefix? Explain your answer.

3. Analyze the alternations in the plural prefix.

- Name of the process:
- Description in words:
- Formal rule:

4. Determine the UR of the following plural nouns:
   i. (3) [zeŋken] ‘necks’:
   ii. (5) [zembɑk] ‘clubs’:
   iii. (8) [zenslɑk] ‘classes’:

B	Singular	Plural	
12.	num	zeŋənum	‘moon’
13.	row	zeŋərow	‘war’
14.	lup	zeŋəlup	‘pool’
15.	jer	zeŋəjer	‘ray’
16.	mef	zeŋəmef	‘frame’

5. According to the analysis of table A, what should have been the UR of the following plural nouns?
   i. (12) [zeŋənum] ‘moons’:
   ii. (14) [zeŋəlup] ‘pools’:
   iii. (15) [zeŋəjer] ‘rays’:

6. Analyze the alternations in table B.

- Name of the process:
- Description in words:
- Formal rule:

C	Indefinite	Definite	
17.	æm	æmys	‘man’
18.	gir	girys	‘ring’
19.	deh	dehys	‘head’
20.	jer	jerys	‘ray’
21.	tɑh	tɑhus	‘hat’
22.	num	numus	‘moon’
23.	row	rowus	‘war’

7. Fill in the vowel chart for Jupiterian according to table C:

	[-back]		[+back]	
[+high]				
[-high,-low]				
[-high,+low]				

8. List the allomorphs of the definite suffix and state their environments.

9. What is the UR of the definite suffix? Explain your answer.

10. Analyze the alternations in the definite suffix.

- Name of the process:
- Description in words:
- Formal rule:

D	Indefinite	Definite	
24.	lut	lusus	‘tool’
25.	bɑk	bɑxus	‘club’
26.	pɑt	pɑsys	‘tap’
27.	nep	neɸys	‘pen’
28.	suːm	suːmus	‘mouse’
29.	tuːf	tuːfus	‘foot’
30.	rɑs	rɑsus	‘star’

11. Analyze the alternations in the consonants in table D.

- Name of the process:
- Description in words:
- Formal rule:

12. Determine the UR of the nouns in (24):
    i. [lut] ‘tool’:
    ii. [lusus] ‘the tool’:

13. Are [s] and [t] distinct phonemes or allophones of the same phoneme in Jupiterian? Explain.

E	Indefinite	Definite	
31.	næf	nævys	‘van’
32.	us	uzus	‘zoo’
33.	dox	doɣus	‘gold’
34.	noɸ	noβus	‘bone’
35.	rɑʃ	rɑʒus	‘jar’

14. Analyze the alternations in the consonants in table E (hint: compare to table D).

- Name of the process:
- Description in words:
- Formal rule:

15. Determine the UR of the nouns in (31):
    i. [næf] ‘van’:
    ii. [nævys] ‘the van’:

F	Indefinite	Definite	
36.	sæl	sælgys	‘glass’
37.	der	derbys	‘bread’
38.	kom	komsus	‘smoke’
39.	kol	kolkus	‘clock’
40.	ʃip	ʃipsys	‘speech’

16. Analyze the alternations in the consonants in table F.

- Name of the process:
- Description in words:
- Formal rule:

17. Determine the UR of the nouns in (36):
    i. [sæl] ‘glass’:
    ii. [sælgys] ‘the glass’:

G	Nominative	Genitive	
41.	æm	esæm	‘man’
42.	gir	ezgir	‘ring’
43.	ken	esken	‘neck’
44.	pihsiʃ	espihsiʃ	‘spaceship’
45.	bɑk	ezbɑk	‘club’
46.	feːʃ	esfeːʃ	‘chef’
47.	deh	ezdeh	‘head’
48.	slɑk	esslɑk	‘class’
49.	tɑːh	estɑːh	‘hat’
50.	ʃif	esʃif	‘fish’
51.	num	esnum	‘moon’
52.	row	esrow	‘war’
53.	lup	eslup	‘pool’
54.	jeːr	esjeːr	‘ray’
55.	mef	esmef	‘frame’
56.	zem	ezzem	‘maze’

18. List the allomorphs of the genitive (‘of’) prefix and state their environments.

19. What is the UR of the genitive prefix? Explain your answer.

20. Analyze the alternations in the genitive prefix.

- Name of the process:
- Description in words:
- Formal rule:

21. Determine the UR of the following genitive nouns:

    i. (41) [esæm] ‘of man’:

    ii. (42) [ezgir] ‘of ring’:

    iii. (43) [esken] ‘of neck’:

22. List all the possible syllable types in Jupiterian (according to tables A-G). Classify the types according to open and closed syllables. Give one example for each type: specify a word containing the relevant syllable type plus the word number and column in the data. If the word is polysyllabic, divide it into syllables and mark the relevant syllable (see example below).

Type	**Open/closed**	**Word**	**Example**
CVC	Closed	2, plural	**<u>zeŋ</u>**.gir

Solution

1. List the allomorphs of the plural prefix and state their environments.
   - [zeŋ]: before vowels and dorsal consonants.
   - [zem]: before labial consonants.
   - [zen]: before coronal consonants.

2. What is the underlying representation (UR) of the plural prefix? Explain your answer.

/zeŋ-/: it appears in the most general environment (before vowels and consonants).

3. Analyze the alternations in the plural prefix.

- Name of the process: (nasal) place assimilation
- Description in words: a nasal consonant assimilates in place to the following obstruent.
- Formal rule: $C_{[+nas]} \rightarrow [\text{place}_i] \ / \ \_C_{[\text{-son, placei}]}$

Notes:

1. It’s possible to write a specific rule for /ŋ/.
2. At this point, it is also possible to analyze the process as assimilation to any consonant.

4. Determine the UR of the following plural nouns:
   i. (3) [zeŋken] ‘necks’: /zeŋ-ken/
   ii. (5) [zembɑk] ‘clubs’: /zeŋ-bak/
   iii. (8) [zenslɑk] ‘classes’: /zeŋ-slak/

5. According to the analysis of table A, what should have been the UR of the following plural nouns?
   i. (12) [zeŋənum] ‘moons’: /zeŋ-num/
   ii. (14) [zeŋəlup] ‘pools’: /zeŋ-lup/
   iii. (15) [zeŋəjer] ‘rays’: /zeŋ-jer/

6. Analyze the alternations in table B.

- Name of the process: vowel epenthesis

- Description in words: a vowel (ə) is inserted between two sonorant consonants.
- Formal rule: $\text{Ø} \rightarrow \text{ə} \,/\, C_{[+son]}\, \_\, C_{[+son]}$

<u>Note</u>: an alternative analysis: insertion between a nasal (or, specifically, /ŋ/) and a sonorant.

7. Fill in the vowel chart for Jupiterian according to table C:

	[-back]		[+back]	
[+high]	i	y		u
[-high,-low]	e			o
[-high,+low]	æ		ɑ	

8. List the allomorphs of the definite suffix and state their environments.
   - [ys]: after a base containing a [-back] vowel.
   - [us]: after a base containing a [+back] vowel.

9. What is the UR of the definite suffix? Explain your answer.

$/\text{-}V_{[+high,+round]}s/$: all the definite words end with some high, rounded vowel + /s/. The backness of the vowel is determine completely by the backness of the base vowel and, therefore, it is underspecified.

10. Analyze the alternations in the definite suffix.

- Name of the process: vowel harmony
- Description in words: a high, rounded vowel assimilates in backness to a preceding vowel across any number of consonants.
- Formal rule: $V_{[+high,+round]} \rightarrow [\alpha back] \,/\, V_{[\alpha back]}C_0\_$

<u>Note</u>: at this point, it is also possible to assume that harmony applies across one consonant at most, since there are no sequences of consonants in the data above.

Alternative rule: $V_{[+high,+round]} \rightarrow [\alpha back] \,/\, V_{[\alpha back]}(C)\_$

Later data will show that harmony applies across sequences of two consonants as well.

11. Analyze the alternations in the consonants in table D.

- Name of the process: spirantization
- Description in words: a (voiceless) plosive becomes fricative between vowels.
- Formal rule: $C_{[-son,-voice]} \rightarrow [+cont]$ / V_V

12. Determine the UR of the nouns in (24):
    i. [lut] 'tool': /lut/
    ii. [lusus] 'the tool': /lut-$V_{[+high,+round]}$s/

13. Are [s] and [t] distinct phonemes or allophones of the same phoneme in Jupiterian? Explain.

They are both.

- There are near-minimal pairs – [t] and [s] appear in the same environment: #_uː in word 28 (suːm) and 29 (tuːf). This suggests that they are distinct phonemes
- They have complementary distributions: [s] appears between vowels, [t] does not. Moreover, there is evidence that /t/ becomes [s] between vowels due to the process of spirantization. That is, they are also allophones of /t/.

14. Analyze the alternations in the consonants in table E (hint: compare to table D).

- Name of the process: final devoicing
- Description in words: an obstruent becomes devoiced at the end of the word.
- Formal rule: $C_{[-son]} \rightarrow [-voice]$ / _#

<u>Note</u>: an alternative analysis: only fricatives become devoiced at the end of the word. However, there are no voiced plosives at the end of a word anywhere in the data.

15. Determine the UR of the nouns in (31):
    i. [næf] 'van': /næv/
    ii. [nævys] 'the van': /næv-$V_{[+high,+round]}$s/

16. Analyze the alternations in the consonants in table F.

- Name of the process: consonant deletion

- Description in words: a consonant is deleted at the end of the word after another consonant.
- Formal rule: C → Ø / C_#

<u>Note</u>: an alternative analysis: only obstruents are deleted. There are no sequences of sonorants at the end of the word anywhere in the data, so we cannot determine which analysis is the correct one.

17. Determine the UR of the nouns in (36):
    i. [sæl] 'glass': /sælg/
    ii. [sælgys] 'the glass': /sælg-$V_{[+high,+round]}$s/

18. List the allomorphs of the genitive ('of') prefix and state their environments.
    - [es]: before vowels, sonorants, and voiceless obstruents.
    - [ez]: before voiced obstruents.

19. What is the UR of the genitive prefix? Explain your answer.

/es-/: it appears in the most general environment (before vowels and consonants).

20. Analyze the alternations in the genitive prefix.

- Name of the process: voicing assimilation
- Description in words: an obstruent assimilates in voice to the following obstruent.
- Formal rule: $C_{[-son]} \rightarrow [\alpha voice]$ / $\_C_{[-son,\alpha voice]}$

<u>Note</u>: an alternative analysis: a rule that turns specifically /s/ to [+voice] before a [+voice] obstruent.

21. Determine the UR of the following genitive nouns:
    i. (41) [esæm] 'of man': /es-æm/
    ii. (42) [ezgir] 'of ring': /es-gir/
    iii. (43) [esken] 'of neck': /es-ken/

22. List all the possible syllable types in Jupiterian (according to tables A-G). Classify the types according to open and closed syllables. Give one example for each type: specify a word containing the relevant syllable type plus the word number and column in the data. If the word is polysyllabic, divide it into syllables and mark the relevant syllable (see example below).

Type	**Open/closed**	**Word**	**Example**
CVC	Closed	2, plural	**zeŋ**.gir
VC	Closed	1, singular	æm
CV	Open	1, plural	**ze**.ŋæm
CCVC	Closed	8, singular	slɑk
CCV	Open	10, singular	trɑ
CV:C	Closed	28, indefinite	suːm
CV:	Open	28, definite	**su**ː.mus
V	Open	32, definite	**u**.zus

Appendices

Appendix A Unicode values of IPA symbols

Occasionally, one may wish to find a certain IPA symbol in a certain character set. This can be quite challenging given the large number of available characters and the existence of some similarly looking characters. For this, and other purposes, it is useful to know the Unicode values of the symbols – codes which uniquely identify each character on the computer. The tables in this appendix list the Unicode values of IPA symbols. Each table represents a different section of the IPA sheet. In addition, some useful symbols that are not official IPA symbols are also included in the tables. These symbols are marked by a star.

Table A1 Unicode values of IPA pulmonic consonants

Features			**Symbol**	**Unicode**
Place	**Manner**	**Voice**		
Alveolar	Affricate	Voiced	ʣ*	02A3
Alveolar	Affricate	Voiceless	ʦ*	02A6
Alveolar	Approximant	Voiced	ɹ	0279
Alveolar	Fricative	Voiced	z	007A
Alveolar	Fricative	Voiceless	s	0073
Alveolar	Lateral approximant	Voiced	l	006C
Alveolar	Lateral fricative	Voiced	ɮ	026E
Alveolar	Lateral fricative	Voiceless	ɬ	026C
Alveolar	Nasal	Voiced	n	006E
Alveolar	Plosive	Voiced	d	0064
Alveolar	Plosive	Voiceless	t	0074
Alveolar	Tap/flap	Voiced	ɾ	027E
Alveolar	Trill	Voiced	r	0072
Alveolo-palatal	Affricate	Voiced	ʥ*	02A5
Alveolo-palatal	Affricate	Voiceless	ʨ*	02A8
Bilabial	Fricative	Voiced	β	03B2
Bilabial	Fricative	Voiceless	ɸ	0278
Bilabial	Nasal	Voiced	m	006D
Bilabial	Plosive	Voiced	b	0062
Bilabial	Plosive	Voiceless	p	0070
Bilabial	Trill	Voiced	ʙ	0299
Dental	Fricative	Voiced	ð	00F0
Dental	Fricative	Voiceless	θ	03B8
Glottal	Fricative	Voiced	ɦ	0266
Glottal	Fricative	Voiceless	h	0068

Features			Symbol	Unicode
Place	**Manner**	**Voice**		
Glottal	Plosive	Voiceless	ʔ	0294
Labiodental	Approximant	Voiced	ʋ	028B
Labiodental	Fricative	Voiced	v	0076
Labiodental	Fricative	Voiceless	f	0066
Labiodental	Nasal	Voiced	ɱ	0271
Labiodental	Tap/flap	Voiced	ⱱ	2C71
Palatal	Approximant	Voiced	j	006A
Palatal	Fricative	Voiced	ʝ	029D
Palatal	Fricative	Voiceless	ç	00E7
Palatal	Lateral approximant	Voiced	ʎ	028E
Palatal	Nasal	Voiced	ɲ	0272
Palatal	Plosive	Voiced	ɟ	025F
Palatal	Plosive	Voiceless	c	0063
Pharyngeal	Fricative	Voiced	ʕ	0295
Pharyngeal	Fricative	Voiceless	ħ	0127
Postalveolar	Affricate	Voiced	ʤ*	02A4
Postalveolar	Affricate	Voiceless	ʧ*	02A7
Postalveolar	Fricative	Voiced	ʒ	0292
Postalveolar	Fricative	Voiceless	ʃ	0283
Retroflex	Approximant	Voiced	ɻ	027B
Retroflex	Fricative	Voiced	ʐ	0290
Retroflex	Fricative	Voiceless	ʂ	0282
Retroflex	Lateral approximant,	Voiced	ɭ	026D
Retroflex	Nasal	Voiced	ɳ	0273
Retroflex	Plosive	Voiced	ɖ	0256
Retroflex	Plosive	Voiceless	ʈ	0288
Retroflex	Tap/flap	Voiced	ɽ	027D
Uvular	Fricative	Voiced	ʁ	0281
Uvular	Fricative	Voiceless	χ	03C7
Uvular	Nasal	Voiced	ɴ	0274
Uvular	Plosive	Voiced	ɢ	0262
Uvular	Plosive	Voiceless	q	0071
Uvular	Trill	Voiced	ʀ	0280
Velar	Approximant	Voiced	ɰ	0270
Velar	Fricative	Voiced	ɣ	0263
Velar	Fricative	Voiceless	x	0078
Velar	Lateral approximant	Voiced	ʟ	029F
Velar	Nasal	Voiced	ŋ	014B
Velar	Plosive	Voiced	g*	0067

Features			Symbol	Unicode
Place	**Manner**	**Voice**		
Velar	Plosive	Voiced	ɡ	0261
Velar	Plosive	Voiceless	k	006B

* Non-IPA symbol

Table A2 Unicode values of IPA non-pulmonic consonants

Section	Description	Symbol	Unicode
Ejectives	Ejective	ʼ	02BC
Clicks	Alveolar lateral click	ǁ	01C1
	Bilabial click	ʘ	0298
	Dental click	ǀ	01C0
	Palatoalveolar click	ǂ	01C2
	(Post)alveolar click	ǃ	01C3
Voiced implosives	Bilabial implosive	ɓ	0253
	Dental/alveolar implosive	ɗ	0257
	Palatal implosive	ʄ	0284
	Uvular implosive	ʛ	029B
	Velar implosive	ɠ	0260

Table A3 Unicode values of IPA other symbols

Description	Symbol	Unicode
Tie bar	͡	0361
Tie-bar below	͜	035C
Voiceless epiglottal fricative	ʜ	029C
Voiceless labial-velar fricative	ʍ	028D
Voiced alveolo-palatal fricative	ʑ	0291
Voiced labial-palatal approximant	ɥ	0265
Simultaneous ʃ and x	ɧ	0267
Voiced alveolar lateral flap	ɺ	027A
Epiglottal plosive	ʡ	02A1
Voiced epiglottal fricative	ʢ	02A2
Voiceless alveolo-palatal fricative	ɕ	0255
Voiced labial-velar approximant	w	0077

Table A4 Unicode values of IPA vowels

Openess	**Backness**	**Roundedness**	**Symbol**	**Unicode**
Near-close	Front	Rounded	ʏ	028F
Close-mid	Front	Rounded	ø	00F8
Mid	Central	Unrounded	ɚ*†	025A
Open-mid	Front	Unrounded	ɛ	025B
Open-mid	Central	Unrounded	ɜ	025C
Open-mid	Central	Unrounded	ɝ*†	025D
Open-mid	Central	Rounded	ɞ	025E
Near-open	Front	Unrounded	æ	00E6
Near-close	Back	Rounded	ʊ	028A
Near-close	Front	Unrounded	ɪ	026A
Open-mid	Back	Unrounded	ʌ	028C
Close	Back	Unrounded	ɯ	026F
Close	Central	Rounded	ʉ	0289
Close-mid	Central	Unrounded	ɘ	0258
Mid	Central	Unrounded	ə	0259
Close-mid	Central	Rounded	ɵ	0275
Open	Front	Rounded	ɶ	0276
Close	Central	Unrounded	ɨ	0268
Close-mid	Back	Unrounded	ɤ	0264
Open-mid	Back	Rounded	ɔ	0254
Open	Front	Unrounded	a	0061
Close-mid	Front	Unrounded	e	0065
Close	Front	Unrounded	i	0069
Close	Back	Rounded	u	0075
Open-mid	Front	Unrounded	œ	0153
Near-open	Central	Unrounded	ɐ	0250
Open	Back	Unrounded	ɑ	0251
Open	Back	Rounded	ɒ	0252
Close-mid	Back	Unrounded	o	006F
Close	Front	Rounded	y	0079

* Non-IPA symbol

† Rhotacized

Table A5 Unicode values of IPA diacritics

Description	Symbol	Unicode	Example
Advanced	̟	031F	u̟
Advanced Tongue Root	̘	0318	e̘
Apical	̺	033A	t̺
Aspirated	ʰ	02B0	tʰ
Aspirated (breathy) voiced	ʱ*	02B1	bʱ
Breathy voiced	̤	0324	a̤
Centralized	̈	0308	ë
Creaky voiced	̰	0330	a̰
Dental	̪	032A	t̪
Glottalized	ˀ*	02C0	tˀ
Labialized	ʷ	02B7	tʷ
Laminal	̻	033B	t̻
Lateral release	ˡ	02E1	tˡ
Less rounded	̜	031C	o̜
Linguolabial	̼	033C	t̼
Lowered	̞	031E	e̞
Mid-centralized	̽	033D	e̽
More rounded	̹	0339	o̹
Nasal release	ⁿ	207F	tⁿ
Nasalized	̃	0303	ẽ
No audible release	̚	031A	p̚
Non-syllabic	̯	032F	e̯
Palatalized	ʲ	02B2	tʲ
Pharyngealized	ˤ	02E4	tˤ
Prenasalized (bilabial)	ᵐ*	1D50	ᵐb
Prenasalized (velar)	ᵑ*	1D51	ᵑk
Prestopped (alveolar)	ᵈ*	1D48	ᵈn
Prestopped (bilabial)	ᵇ*	1D47	ᵇm
Prestopped (velar)	ᶢ*	1DA2	ᶢŋ
Raised	̝	031D	e̝
Retracted	̠	0320	e̠
Retracted Tongue Root	̙	0319	e̙
Rhoticity	˞	02DE	a˞
Syllabic	̩	0329	n̩
Syllabic	̍	030D	ŋ̍
Velarized	ˠ	02E0	tˠ
Velarized or pharyngealized	̴	0334	ɫ
Voiced	̬	032C	s̬

Description	Symbol	Unicode	Example
Voiceless	◌̥	0325	n̥
	◌̊	030A	ŋ̊

Table A6 Unicode values of IPA suprasegmentals

Description	Symbol	Unicode	Example
Extra-short	˘	0306	ă
Half-long	ˑ	02D1	aˑ
Linking (absence of a break)	‿	203F	z‿v
Long	ː	02D0	aː
Major (intonation) group	‖	2016	e‖
Minor (foot) group	|	007C	e|
Primary stress	ˈ	02C8	ˌfoʊnəˈtɪʃən
Secondary stress	ˌ	02CC	
Syllable break	.	002E	ɹi.ækt

Table A7 Unicode values of IPA tones and word accents†

Description	Symbol	Unicode	Example
Downstep	ꜜ	A71C	ꜜma
Extra high	̋	030B	e̋
Extra high	˥	02E5	ma˥
Extra low	̏	030F	ȅ
Extra low	˩	02E9	ma˩
Falling	̂	0302	ê
	˥˩	02E5+ 02E9	ma˥˩
Global fall	↘	2198	↘
Global rise	↗	2197	↗
High	́	0301	é
	˦	02E6	e˦
High rising	᷄	1DC4	e᷄
	˧˥	02E7+ 02E5	ma˧˥
Low	̀	0300	è
	˨	02E8	˨
Low rising	᷅	1DC5	e᷅
	˩˧	02E9+ 02E7	ma˩˧
Mid	̄	0304	ē
	˧	02E7	˧

Description	Symbol	Unicode	Example
Rising	ˇ	030C	ě
	˩˥	02E9+ 02E5	ma˩˥
Rising-falling	᷈	1DC8	e᷈
Upstep	ꜛ	A71B	ꜛma

† Contour tones can be represented by sequences of tone level letters. For example, a rising-falling tone can be represented by the sequence: mid tone (02E7) + extra-high tone (02E5) + mid tone (02E7) (e.g., ma˧˥˧).

Appendix B Recordings by native speakers of various languages

Links to pages from the UCLA Phonetics Lab Data website (Ladefoged, 2005) containing recordings of words uttered by native speakers of various languages that are referred to in this book. Some of these files came from a CD accompanying the fifth edition of Peter Ladefoged's "A Course in Phonetics" (Ladefoged, 2006). Other files appear on various pages of the "Index of languages" page of the UCLA Phonetics Lab Data.

Language	URL
!Xóõ	phonetics.ucla.edu/course/chapter6/xong/!xong.html
Aghul	phonetics.ucla.edu/appendix/languages/agul/agul.html
Akan	phonetics.ucla.edu/appendix/languages/akan/akan.html
Arabic	phonetics.ucla.edu/appendix/languages/arabic/arabic.html
Badaga	phonetics.ucla.edu/appendix/languages/badaga/badaga.html
Bura	phonetics.ucla.edu/appendix/languages/bura/bura.html
Chinese	phonetics.ucla.edu/course/chapter10/chinese/chinese.html
Danish	phonetics.ucla.edu/course/chapter9/danish/danish.html
Finnish	phonetics.ucla.edu/appendix/languages/finnish/finnish.html
French	phonetics.ucla.edu/course/chapter11/french/french.html
Hindi	phonetics.ucla.edu/course/chapter6/hindi/hindi.html
Idoma	phonetics.ucla.edu/appendix/languages/idoma/idoma.html
Kele	phonetics.ucla.edu/vowels/chapter13/kele.html
Lakhota	phonetics.ucla.edu/course/chapter6/lakhota/lakhota.html
Malayalam	phonetics.ucla.edu/course/chapter7/malayalam/malayalam.html
Norwegian	phonetics.ucla.edu/appendix/languages/norwegian/norwegian.html
Russian	phonetics.ucla.edu/appendix/languages/russian/russian.html
Sindhi	phonetics.ucla.edu/course/chapter6/sindhi/sinhi.html
Wangurri	phonetics.ucla.edu/appendix/languages/wangurri/wangarri.html
Zulu	phonetics.ucla.edu/course/chapter7/zulu/zulu.html

Appendix C Animations and MRI videos from the *Seeing Speech* project

The table below lists links to animation and MRI video clips from the *Seeing Speech* project that are used in this book. The full reference item for each video is of the form:

Animation. Janet Beck. <Segment>. Seeing Speech. Glasgow: University of Glasgow, 2018. Web. 5 March 2021. <Link>.

Where <Segment> should be substituted with a segment name and <Link> should be substituted with a link from the table below. For example, the reference to the voiced bilabial plosive video is:

Animation. Janet Beck. Voiced bilabial plosive. Seeing Speech. Glasgow: University of Glasgow, 2018. Web. 5 March 2021. https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=98.

Segment	Link
Voiced bilabial plosive	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=98
Voiced alveolar plosive	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=100
Voiced velar plosive	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=609
Voiced alveolar nasal	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=110
Voiced alveolar fricative	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=122
Voiced bilabial nasal	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=109
Voiced labiodental nasal	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=625
Voiceless alveolar fricative	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=115
Voiceless dental fricative	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=952
Voiced postalveolar fricative	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=658
Voiced retroflex approximant	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=635
Voiceless palatal plosive	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=99
Voiceless uvular plosive	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=113
Voiceless pharyngeal fricative	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=295
Voiced pharyngeal fricative	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=661

Segment	Link
Glottal plosive /a_a/ and /i_i/ environments.	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=1&speaker=1#location=660
Voiceless velar fricative	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=120
Voiced retroflex fricative	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=3&speaker=1#location=656
Voiced labial-velar approximant	https://seeingspeech.ac.uk/ipa-charts/?chart=3&datatype=1&speaker=1#location=119
Voiced bilabial fricative /a_a/ and /i_i/ environments	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=1&speaker=1#location=946
Voiceless bilabial plosive /a_a/ and /i_i/ environments	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=1&speaker=1#location=112
Front close unrounded vowel	https://seeingspeech.ac.uk/ipa-charts/?chart=4&datatype=1&speaker=1#location=105
Front open unrounded vowel	https://seeingspeech.ac.uk/ipa-charts/?chart=4&datatype=1&speaker=1#location=97
Front close rounded vowel	https://seeingspeech.ac.uk/ipa-charts/?chart=4&datatype=1&speaker=1#location=121
Back close rounded vowel	https://seeingspeech.ac.uk/ipa-charts/?chart=4&datatype=1&speaker=1#location=117
Front close-mid unrounded vowel	https://seeingspeech.ac.uk/ipa-charts/?chart=4&datatype=1&speaker=1#location=101
Front close-mid rounded vowel	https://seeingspeech.ac.uk/ipa-charts/?chart=4&datatype=1&speaker=1#location=248
Voiceless bilabial ejective	https://seeingspeech.ac.uk/ipa-charts/?chart=2&datatype=1&speaker=1#location=112700
Voiced bilabial implosive	https://seeingspeech.ac.uk/ipa-charts/?chart=2&datatype=1&speaker=1#location=595
Voiced palatal approximant /a_a/ and /i_i/ environments	https://seeingspeech.ac.uk/ipa-charts/?chart=1&datatype=1&speaker=1#location=106

Subject index

A

B

C

D

E

F

G

H

I

L

M

N

O

P

Q

R

S

T

U

V

W

Language index

References

Ararahih’urípih: A Dictionary and Text Corpus of the Karuk Language. (2023). https://linguistics.berkeley.edu/~karuk/

Asherov, D., Fishman, A., & Cohen, E.-G. (2016). Vowel Reduction in Israeli Heritage Russian. *Heritage Language Journal*, *2*, 113–133.

Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication*, *46*, 252–267. https://doi.org/10.1016/j.specom.2005.02.016

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2007). *Neuroscience: Exploring the brain* (3rd ed.). Lippincott Williams & Wilkins. https://doi.org/10.1007/BF02234670

Blau, J. (2010). Phonology and Morphology of Biblical Hebrew: An Introduction. In *Linguistic Studies in Ancient West Semitic*. Eisenbrauns.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9), 341–345.

Bright, W. (1957). *Karok language*. University of California Press.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. Harper & Row.

Collins, B., & Mees, I. M. (2003). *The Phonetics of English and Dutch* (5th ed., Vol. 58, Issues 1–2). Brill. https://doi.org/10.1016/0024-3841(82)90063-8

Dambriūnas, L., Klimas, A., & Schmalstieg, W. R. (1993). *Introduction to Modern Lithuanian* (Fifth). Franciscan Fathers.

Daskum, A. A. (2009). *Duwai-English-Hausa Dictionary* (R. G. Schuh (Ed.)). Ajami Press. http://aflang.humanities.ucla.edu/language-materials/chadic-languages/yobe/dictionaries-yobe-languages/

de Lacy, P. (Ed.). (2007). *The Cambridge Handbook of Phonology*. Cambridge University Press.

Dupoux, E., Hirose, Y., Kakehi, K., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568–1578. https://doi.org/10.1037/0096-1523.25.6.1568

Echeverría, M. S., & Contreras, H. (1965). Araucanian Phonemics. *International Journal of American Linguistics*, *31*(2), 132–135.

Ferragne, E., & Pellegrino, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. In *Journal of the International Phonetic Association* (Vol. 40, Issue 1). https://doi.org/10.1017/S0025100309990247

Freese, J., & Maynard, D. W. (1998). Prosodic features of bad news and good news in conversation. *Language in Society*, *27*(2), 195–219. https://doi.org/10.1017/S0047404598002024

French, K. M. (1988). *Insights into Tagalog: Reduplication, Infixation, and Stress from Nonlinear Phonology*. The Summer Institute of Linguistics.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, *22*(1), 179–228. https://doi.org/10.1023/B:NALA.0000005557.78535.3c

Gerken, L. A., & McGregor, K. (1998). An Overview of Prosody and Its Role in Normal and Disordered Child Language. *American Journal of Speech-Language Pathology*, *7*(2), 38–48. https://doi.org/10.1044/1058-0360.0702.38

Gordon, M. (2002). A Factorial Typology of Quantity-Insensitive Stress. *Natural Language & Linguistic Theory*, *20*(3), 491–552. https://doi.org/10.1023/A:1015810531699

Gussenhoven, C., & Jacobs, H. (2017). *Understanding Phonology* (4th ed.). Routledge.

Hall, T. A. (1997). *The Phonology of Coronals*. John Benjamins Publishing Company. https://doi.org/10.2307/417467

Halle, M., & Clements, G. N. (1983). *Problem Book in Phonology: a workbook for introductory courses in linguistics and in modern phonology*. MIT Press.

Hamann, S. (2004). Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Association*, *34*(1), 53–67. https://doi.org/10.1017/S0025100304001604

Harrington, J., Palethorpe, S., & Watson, C. (2000). Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, *30*(1–2), 63–78. https://doi.org/10.1017/S0025100300006666

Hayes, B., & Abad, M. (1989). Reduplication and syllabification in Ilokano. *Lingua*, *77*(3–4), 331–374. https://doi.org/10.1016/0024-3841(89)90044-2

Heffner, H. E., & Heffner, R. S. (2007). Hearing ranges of laboratory animals. *Journal of the American Association for Laboratory Animal Science*, *46*(1), 20–22.

Howard, I. (1972). *A directional theory of rule application in phonology*. Massachusetts Institute of Technology.

International Phonetic Association. (2018). *IPA Chart*. http://www.internationalphoneticassociation.org/content/ipa-chart

Johnson, K. (2012). *Acoustic and Auditory Phonetics* (3rd ed.). Wiley-Blackwell.

Jones, D. (1956). *The Pronunciation of English* (4th ed.). Cambridge University Press.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In Jinni A. Harrigan, R.

Rosenthal, & K. R. Scherer (Eds.), *The New Handobook of Methods in Nonverbal Behavior Research* (pp. 65–135). Oxford University Press.

Kager, R. (2007). Feet and metrical stress. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (p. 202). Cambridge University Press.

Kenstowicz, M. J., & Kisseberth, C. W. (1979). *Generative Phonology: Description and Theory*. Academic Press. https://doi.org/10.2307/413703

Kenstowicz, M. J., & Kisseberth, C. W. (Eds.). (1973). *Issues in Phonological Theory: Proceedings of the Urbana Conference on Phonology*. Mouton. https://doi.org/10.3406/mat.1992.410642

KissThisGuy: the archive of misheard lyrics. (2021). https://www.kissthisguy.com/

Kreitman, R. (2010). Attending to pronunciation issues in teaching Modern Israeli Hebrew. *Hebrew Higher Education*, *13*, 45–69.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, *255*(5044), 606–608.

Labov, W. (1972). The social stratification of (r) in New York City department stores. In W. Labov (Ed.), *Sociolinguistic Patterns* (pp. 43–54). University of Pennsylvania Press. https://doi.org/10.1017/cbo9780511618208.006

Ladefoged, P. (2005). *UCLA Phonetics Lab Data*. http://phonetics.ucla.edu/

Ladefoged, P. (2006). *A Course in Phonetics* (5th ed.). Thomson.

Ladefoged, P., & Johnson, K. (2010). *A Course in Phonetics* (6th ed.). Cengage learning.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Blackwell.

Laks, L. (2011). *Morpho-phonological and Morpho-thematic Relations in Hebrew and Arabic Verb Formation*. Doctoral dissertation. Tel-Aviv University.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge University Press.

Lawson, E., Stuart-Smith, J., Scobbie, J., & Nakai, S. (2018). *Seeing Speech: an articulatory web resource for the study of Phonetics*. https://seeingspeech.ac.uk

Levrero, F., Mathevon, N., Pisanski, K., Gustafsson, E., & Reby, D. (2018). The pitch of babies' cries predicts their voice pitch at age 5. *Biology Letters*, *14*, 20180065. https://doi.org/10.1098/rsbl.2018.0065

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. WORD, *20*(3), 384–422.

https://doi.org/10.1080/00437956.1964.11659830

Lockwood, W. B. (1955). *An Introduction to Modern Pharoese*. Føroya Skúlabókagrunnur.

Mathiassen, T. (1996). *A Short Grammar of Lithuanian* (Vol. 41, Issue 3). Slavica Publishers. https://doi.org/10.2307/310215

McCarthy, J. J. (1981). A Prosodic Theory of Nonconcatenative Morphology. *Linguistic Inquiry*, *12*(3), 373–418.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Moran, S., & McCloy, D. (2019). *PHOIBLE 2.0* (S. Moran & D. McCloy (Eds.)). Max Planck Institute for the Science of Human History. http://phoible.org

Newlin-Łukowicz, L. (2012). Polish stress: Looking for phonetic evidence of a bidirectional system. *Phonology*, *29*(2), 271–329. https://doi.org/10.1017/S0952675712000139

Odden, D. (2013). Introducing Phonology. In *Cambridge Introduction to Language and Linguistics* (2nd ed.). Cambridge University Press.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*(1), 1–16. https://doi.org/10.1159/000261706

Paradis, C., & Lacharité, D. (1997). Preservation and minimality in loanword adaptation. *Journal of Linguistics*, *33*(2), 379–430. https://doi.org/10.1017/S0022226797006786

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175–184.

Pullum, G. K., & Ladusaw, W. A. (1996). *Phonetic symbol guide*. University of Chicago Press.

Ramachandran, S., & Hubbard, E. M. (2001). Synaesthesia—A Window Into Perception, Thought and Language. *Journal of Consciousness Studies*, *8*(12), 3–34. https://doi.org/10.1111/1468-0068.00363

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*(4), 407–429. https://doi.org/10.1006/jpho.2001.0145

Rice, K. (1995). On Vowel Place Features. *Toronto Working Papers in Linguistics*, *14*, 73–116.

Roach, P. (2011). *Glossary - a little English encyclopaedia of phonetics*. http://www.sgha.net/library/PeterRoach_Glossary.pdf

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, *80*(3), 475–531. https://doi.org/10.1353/lan.2004.0144

Schuh, R. G. (2002). Palatalization in West Chadic. *Studies in African Linguistics*, *31*(1), 97–128.

Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge University Press.

Suzuki, K. (1998). *A Typological Investigation of Dissimilation*. Univeristy of Arizona.

Trask, R. L. (1996). *A dictionary of phonetics and phonology*. Routledge.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, *88*(1), 97–100. https://doi.org/10.1121/1.399849

Treiman, R., & Zukowski, A. (1990). Toward an understanding of English syllabification. *Journal of Memory and Language*, *29*(1), 66–85. https://doi.org/10.1016/0749-596X(90)90010-W

Trudgill, P. (2016). *Dialect matters: Respecting vernacular language*. Cambridge University Press.

Tsur, R. (1992). *What Makes Sound Patterns Expressive: The Poetic Mode of Speech-Perception*. Duke University Press.

van den Berg, R., & Busenitz, R. L. (2012). *A grammar of Balantak: A language of Eastern Sulawesi*. SIL International. https://doi.org/10.1353/ol.2014.0012

Weenink, D. (2018). *Speech Signal Processing with Praat*. https://www.fon.hum.uva.nl/david/

Wright, S. (1954). The death of Lady Mondegreen. *Harper's Magazine*, *209*(1254), 48–51.

Xu, K., Yang, Y., Stone, M., Jaumard-Hakoun, A., Leboullenger, C., Dreyfus, G., Roussel, P., & Denby, B. (2016). Robust contour tracking in ultrasound tongue image sequences. *Clinical Linguistics and Phonetics*, *30*(3–5), 313–327. https://doi.org/10.3109/02699206.2015.1110714