Descriptive Phonetics and Project 3

Anjali Kantharuban and David R. Mortensen

March 12, 2024

Figure 1: Now we're talking about writing and sounds

Orthographies[1] are writing systems and, in the general case, each writing system has its own way of writing any particular sound. For example, the /ʃ/ sound in English *shoe* is written as ⟨sh⟩ (and sometimes ⟨ti⟩ or ⟨si⟩) in English but as ⟨sch⟩ in German, ⟨ch⟩ in French, ⟨s⟩ in Hmong, ⟨x⟩ or ⟨z⟩ in Portuguese, and so on[2]. And these examples are all from one script[3].

If we want to study the sounds in language systematically (and study orthographies systematically) we need a consisting way of writing sounds that is not **depedendent** on the writing system of any particular language. After all, no language includes all of the sounds that can occur in the world's languages. For this purpose, phoneticians developed the International Phonetic Alphabet, which is used throughout the world to represent the sounds of languages. **It is capable of representing all of the known sounds in the languages of the world with only limited ambiguity.**

Tables 1 and 2 show the IPA symbols for most of the common sounds in the kind of English that is taught in US schools. The vowels are quite different from those in prestige varieties of British English ("RP"), African American English, Southern English, Indian English, Australian English, and so on.

Phonemes and Phones

It may have surprised you, particular if your first language is American English, that the ⟨p⟩ in ⟨spit⟩ and that in ⟨pit⟩ are transcribed differently. These are functionally the same sound (they never distinguish one word from another in English). However, they are physically different. In Mandarin, the same pair of sounds (written in the Pinyin orthography as ⟨b⟩ and ⟨p⟩) distinguish words. For example, Mandarin 波 ⟨bō⟩ 'wave' has the [p] sound and 坡 ⟨pō⟩ 'slope' has the [pʰ] sound.

As you learn to speak many varieties of English, you learn to hear these

[1] Orthography comes from Greek *ortho-* 'correct' and *graph-* 'relating to writing' and is technical a prescribed way of writing a language. Naïve people will often confuse the orthography with the language and say things like "Chinese is the oldest language" or "Tamil is the oldest language of Southern India" whereas, in fact, all languages are equally old. Some just have a longer orthographic traditions that others.

[2] We will enclose orthographic sequences in angle brackets (⟨ and ⟩) and phonemic sequences (more later) in slashes (/ and /).

[3] A script is a writing symbol that uses roughly the same set of written symbols. English, German, French, Hmong, and Portuguese are all written with the Latin (or Roman) script. More about this in the next lecture

[pʰ]	**p**it	[b]	**b**it
[p]	s**p**it		
[tʰ]	**t**ock	[d]	**d**ock
[t]	s**t**ock		
[kʰ]	**c**ot	[g]	**g**ot
[k]	S**c**ot		
		[ɾ]	wa**t**er
[m]	**m**it	[m]	si**m**
[n]	**n**it	[n]	si**n**
[ŋ]	—	[ŋ]	si**ng**
[l]	**l**ip	[ɬ]	pi**ll**
[w]	**w**ack	[j]	**y**ack
[ɹ]	**r**ack		
[f]	**f**an	[v]	**v**an
[θ]	**th**igh	[ð]	**th**y
[s]	**s**ink	[z]	**z**inc
[ʃ]	a**ss**ure	[ʒ]	a**z**ure
[h]	**h**at		

Table 1: Some IPA consonant symbols for English as taught in US schools. Sounds written differently that in English are highlighted in purple.

as the same sound because they are not in CONTRAST in most Englishes. Even though they are physically distinct sounds (they are different PHONES) they are the same functional sound (the same PHONEME). How can they be physically distinct but not functionally distinct? Their DISTRIBUTIONS[4] are complementary: in the contexts where [pʰ] occurs (at the beginning of words and at the beginning of stress syllables) [p] never occurs (and vice versa). Thus, the distributions of the two sounds—the two phones—is predictable and replacing one with the other would convey no additional information. PHONEMICIZATION[5] is a kind of compression (similar to what occurs in human cognition) wherein this redundant predictable information is factored out, leaving only the CONTRASTIVE information.

Phones that belong to the same phoneme are called ALLOPHONES. Thus, [p] and [pʰ] are allophones of the phoneme /p/. Some phonemes, like English /f/, are REALIZED as only one phone (in this case, [f]) and we do not use the term ALLOPHONE in this case.

The distinction between PHONE and PHONEME may seems simple, but it is not one that all speech scientists have mastered. The Interspeech 2024 guidelines state "authors are encouraged to use the terms 'phoneme' and

[4] In linguistics, we refer to the total set (or multiset) of contexts in which an element like a phone can occur as its distribution.

[5] The process of converting strings of phones to strings of phonemes.

[i]	beet			[u]	boot
[ɪ]	bit			[ʊ]	book
[ej]	bait	[ə]	but	[ow]	boat
[ɛ]	bet	[ʌ]	butt	[ɔ]	bought
[æ]	bat			[ɑ]	bot

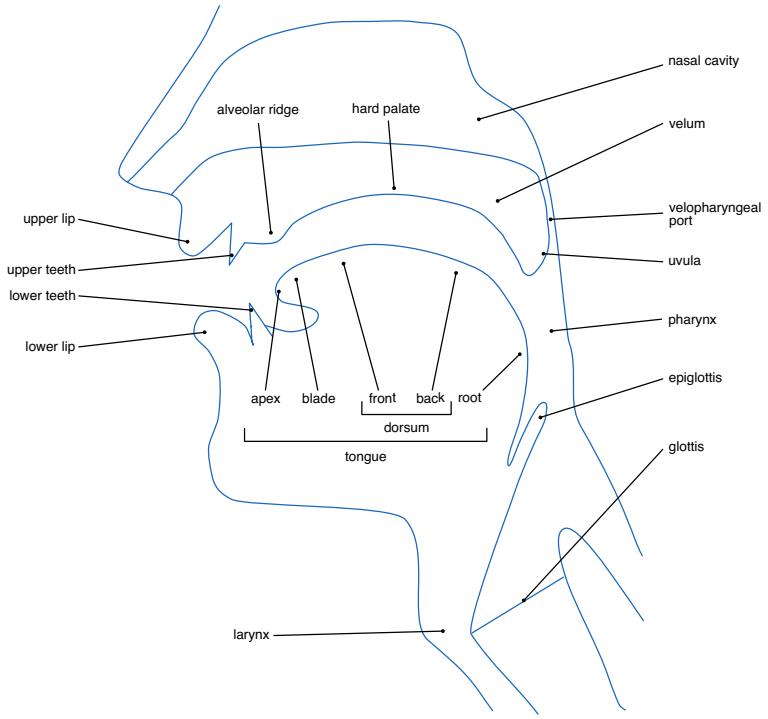
Table 2: Some IPA vowel symbols for English as taught in US schools. Sounds written differently that in English are highlighted in purple.

'phone' correctly," suggesting that they often do not.

Place of Articulation

Every vowel or consonant has its place. Specifically, it has a place of articulation—the anatomical location where there is the greatest constriction when it is made. These constrictions are within the VOCAL TRACT which extends from the larynx to the lips. A outline of the vocal anatomy is shown in Figure 2. This kind of diagram is what is called a MID-SAGITTAL SECTION.

Figure 2: Place of articulation by organ

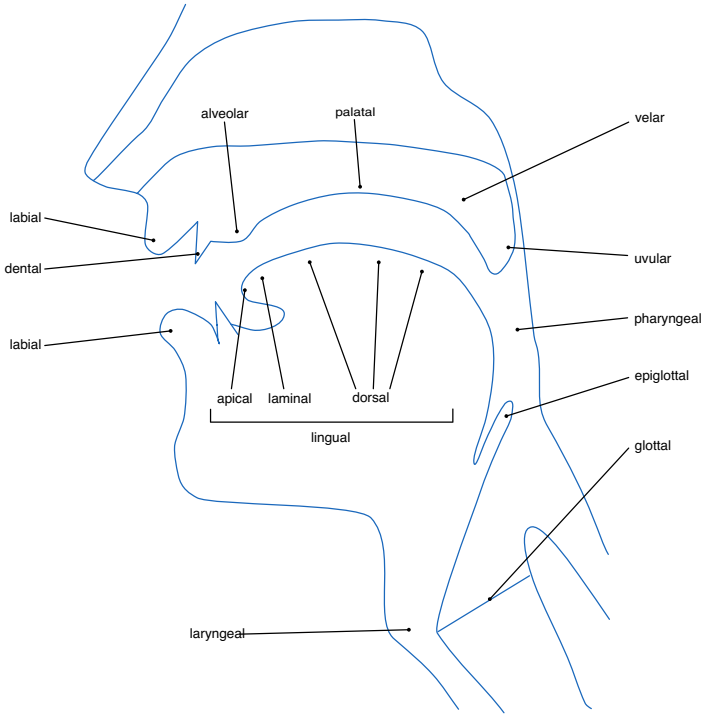


It is as if a phonetician has taken a human head and sliced it down the mid-

dle, starting with the nose and proceeding through the head to the back of the skull, severing the tongue up the middle. The labels show all of the major landmarks needed for identifying where sounds are made. For example, in American Englishes, the [t] sound in ⟨tuck⟩ is made by raising the apex of the tongue up to touch the alveolar ridge. The [i] sound in ⟨bead⟩ is made by gently raising the front of the tongue towards the front part of the hard palate.

Because they want to appear to be more like physicians than other humanists/social scientists, they love to show off complicated terminology. They have adjectives to describe each of the places of articulation. These are shown in Figure 3.

Figure 3: Place of articulation by adjective



These terms would actually be adequate for describing both consonants and vowels, but for historical reasons, there is a separate set of terms for vowel place of articulation. Vowel place is described in three dimensions: height (or openness) backness (or frontness), and rounding (labialization). These define a three dimensional space. The height and backness dimensions are illustrated (for schoolbook US English) in Figure 4.

Say each of these words in your best Central Ohio accent. Note that your

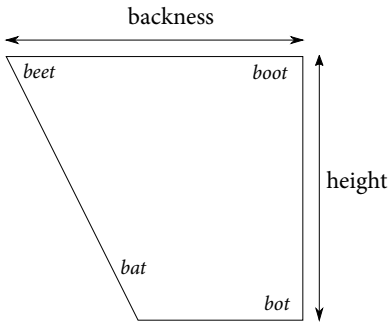


Figure 4: Vowel place in two dimensions

tongue is closer to the roof of your mouth when you say *beet* than when you say *bat* and when you say *boot* than when you say *bot.* The backness dimension may be more difficult to sense via PROPRIOCEPTION[6]. Try saying *beet*, then—while still saying it—slide into *boot* and notice what your tongue is doing. You should feel it sliding back somewhat.

You should feel something else—your lips rounding and extending. This is the third dimension of vowel place, namely LIP ROUNDING. This is visualized in Figure 5.

[6] Your body's ability to sense its own position.

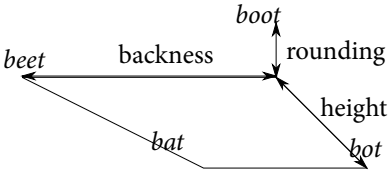


Figure 5: Vowel place in three dimensions

Manner of Articulation

The phones /t/ and /s/ have the same place of articulation, in most American Englishes, at least, but they are different. This difference is MANNER OF ARTICULATION, the way in which the tightest constriction is made when producing a sound. The highest level distinction between manners of articulation is between consonants and vowels, and vowels—to a first approximation—have only one manner[7] There are several consonant manners of articulation. Here are the main ones:

[7] An exception might be the fricative vowels that are found both in some languages of Africa as well as some varieties of Chinese, for example the vowel written as ⟨i⟩ in Mandarin 四 sì 'four', which sounds like something between the vowel [i] and the consonant [z].

- **plosives** or **oral stops** Characterized by the complete obstruction of the vocal tract and the closure of the velopharyngeal port; like the ⟨p⟩ in *porpoise*
- **nasal stops** or NASALS Characterized by the complete obstruction of the vocal tract but with the velopharyngeal port open; like the ⟨m⟩ in *muddle*
- **trills** Produced with a "loose" closure so that the passage of air produces an oscillation
- **flap** or **tap** essentially a momentary plosive produced when an ACTIVE ARTICULATOR strikes a PASSIVE ARTICULATOR; like the ⟨t⟩ in *writer* or the ⟨d⟩ in *rider*
- **fricatives** Characterized by a tight constriction that produced turbulence when air is blown through it; like the ⟨s⟩ in *slither*
- **lateral fricative** A special kind of fricative in which the opening is on one or both sides of the tongue; common in exotic languages like Hmong and Welsh
- **approximant** Characterized by a loose constriction; includes glides like the ⟨w⟩ in *wand* and other sounds like the ⟨r⟩ in *raven*
- **lateral approximant** A special type of approximant in which there is an opening on one or both sides of the tongue; like the ⟨l⟩ in *leprechaun*

Airsteam Mechanism

AIRSTREAM MECHANISM refers to how air is made to move into order to produce a sound. There is a five-way distinction:

Pulmonic egressive Lungs force air out. Used in most speech sounds.

Pulmonic ingressive Lungs pull air in. Used only in special styles of speech.

Velaric ingressive Air is pulled into a vacuum created between the tongue and velum. Click sounds occurring only in languages of southern Africa (like Xhosa) and one variety of Nahuatl in Mexico.

Glottalic egressive The VOCAL FOLDS are closed and the larynx is raised in order to force air out. Ejectives.

Glottalic ingressive The vocal folds are closed and the larynx is lowered in order to create a vacuum. Implosives like Vietnamese [ɗ]. Also common in West Africa.

Voicing

Finally, the state of the GLOTTIS or vocal folds is important. Produce an [s], then produce a [z]. Alternate back and forth with your finger on your throat. You should notice a buzzing when you produce [z] but no buzzing when you produce [s]. That buzzing is the vibration of the glottis. Because of this vibration, [z] is called VOICED. [s] is, in contrast, called VOICELESS.

That may seem simple enough, but the reality is more complicated. There is actually a continuum of voicing. Take [d], [t], and [tʰ]. Of these, [d] is the "most voiced." The vocal folds, in a "real" [d], are vibrating all the way through the sound. In [t], the vocal folds start vibrating just as the closure of the consonant (a plosive) is released. In [tʰ], the voicing does not start until long after that. This distinction is called VOICE ONSET TIME. "Voiced" sounds have negative VOT; "voiceless" sounds have zero VOT; and aspirated sounds have positive VOT.

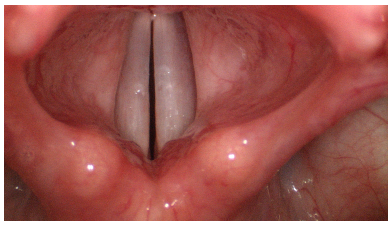


Figure 6: The vocal folds

Putting it All Together

Taking the cross product of all of these vectors, then excluding the combinations that do not appear to show up in any language, we can construct a complete set of categories that can be mapped onto a set of symbols. The based pulmonic egressive symbols are shown in Table 3 and Figure 7.

There are other symbols for non-pulmonic consonants and for specific places of articulation that do not fit well on the consonant chart. There are also many diacritics that are used to represent tone, secondary places of articulation, and to modify sounds in various ways.

	Bilabial		Labiodental		Dental		Alveolar		Postalveolar		Retroflex		Palatal		Velar		Uvular		Pharyngeal		Glottal	
Plosive	p	b					t	d			ʈ	ɖ	c	ɟ	k	g	q	ɢ			ʔ	
Nasal		m		ɱ				n				ɳ		ɲ		ŋ		ɴ				
Trill		ʙ						r										ʀ				
Tap or Flap								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
Lat. Fric.							ɬ	ɮ														
Approximant				ʋ				ɹ				ɻ		j		ɰ						
Lat. Approx.								l				ɭ		ʎ		ʟ						

Table 3: The IPA pulmonic consonants

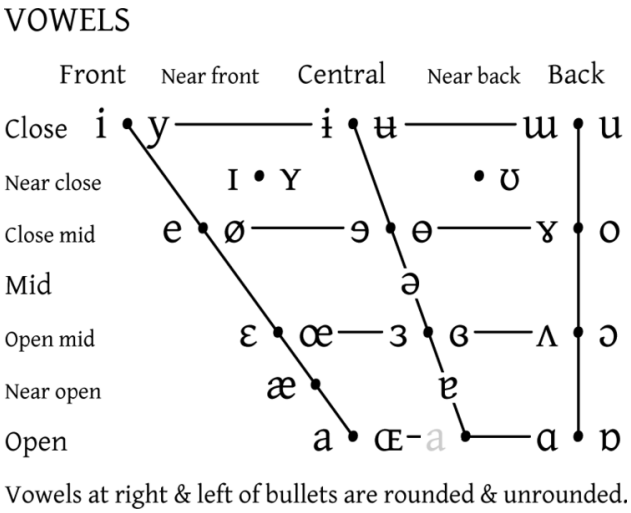


Figure 7: The IPA vowels

Mini-Project 3

In the third mini-project, you will use Epitran [8] (or a tool of your choice) to construct a GRAPHEME-TO-PHONEME module[9] for a language that is not currently supported by Epitran (see `https://github.com/dmort27/epitran`). If you choose to make a pull-request and it is accepted, you will be included as a coauthor on the forthcoming Epitran 2 (a rewrite of Epitran in Rust with a number of new features and improvements) paper. For more information, see the course website.

For a more interesting Mini-Project 3, you might consider building a G2P model for a language with a complex orthography, where the written form underdetermines the phonemic form. One example is Arabic, for which lots of training data exists. Another example is English (for which we have CMUDict). A much more challenging example is Urdu, for which their is not a pronouncing dictionary. Urdu is written in a Perso-Arabic script and short vowels are not indicated. A solution to this could be to build a language model over phonemic transcriptions of Hindi, another variety of Hindustani, and use it to decode noisy and ambiguous outputs from Urdu. This could

[8] David R. Mortensen, Siddharth Dalmia, and Patrick Littell. Epitran: Precision G2P for many languages. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA). URL `https://aclanthology.org/L18-1429`

[9] Grapheme-to-phoneme (or G2P) systems convert orthography into phonemes (or, sometimes, phones).

make a nice paper for a workshop like SIGDIAL.

Exercises

1. Transcribe the following words phonetically, as they would be pronounced in US schoolbook English:

   (a) cold

   (b) scold

   (c) anthrax

   (d) young

   (e) leisure

   (f) talk

References

David R. Mortensen, Siddharth Dalmia, and Patrick Littell. Epitran: Precision G2P for many languages. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA). URL `https://aclanthology.org/L18-1429`.