# 2 Phonetic Variation in Dialects

*Rachael Tatman*

## 2.1 Introduction

Even if you've never studied linguistics, you already know that a lot of social information is reflected in the ways that we use language. You've almost certainly met someone for the first time and been able to immediately figure out something about their social background based on the way they speak or sign. Linguistic features associated with social or regional identity occur at every level of the grammar, from speech sounds to syntactic constructions to the cues that people use to show that their conversational turn is over. Such features can also encode a wide variety of social information, from social class to gender to how formal a speaker is being. Investigating the relationships between linguistics features and social identity is the focus of the field of sociolinguistics.

In this chapter, we will focus on a very small part of sociolinguistics: how variation in the sounds of speech[1] reflect the regional origin of the speaker.

The study of the production and perception of speech sounds is called *phonetics*, and the differences in speech sounds between language varieties is commonly called *phonetic variation*. Rather than focusing on the specific differences between a small handful of dialects we will discuss the ways speech sounds can vary across dialects and what causes these differences, illustrated with a handful of examples. This will give you the general framework and vocabulary you will need to begin reading the sociophonetic literature for the specific dialect(s) of interest to you and your work.

Before we began investigating specific reasons for differences between dialects, we will need to cover some introductory phonetic principles and vocabulary: phonemes, vowels, consonants, and suprasegmentals. Then we will briefly discuss why differences between dialects arise. After this introduction, the rest of the chapter will consist of a discussion of processes that affect

[1] While sign languages are beyond the scope of this chapter, they also show sociolinguistic variation. See Lucas (2001) for an overview.

different classes of phonemes, along with examples of dialect differences arising from each kind of process.

### 2.1.1 What Are Phonemes?

*Phonemes* are the smallest unit of sound in a language (Trubetzkoy, 1969). A change in a single phoneme can change an existing word into a different word or make it a non-word. For example, if you change the first sound of the word *shrimp* (*sh* as in *ship*) into *s* (as in *sip*) you are left with the non-word *srimp*. This is good evidence that *sh* and *s* are different phonemes in English. However, if you replaced the *r* sound in the Japanese アロエ(said *ah-row-ey*) with the American English *l* sound (so that it would sound more like *ah-low-ey*), a native Japanese listener would probably still accept is as アロエ. This is because the American English sounds *l* and *r* are not separate phonemes in Japanese. While not all linguists agree about the existence or nature of phonemes, phonemes are a useful shorthand for discussing clusters of similar sounds and most texts, including this one, will use them.

Phonemes can generally be divided into two large classes: vowels and consonants. They are distinguished by the shape of the vocal tract. The *vocal tract* is the term that linguists use to refer to the throat, mouth, and nasal cavity. In order to make speech sounds, we use our lungs to push air through our vocal chords, up through the throat and then out through either the mouth and/or the nasal cavity. By manipulating different parts of the vocal tract, we can change the quality of the sound produced.

Phoneme-level transcriptions of speech are almost always done using the specialized international phonetic alphabet, or IPA. However, as learning to read and write IPA is very time-consuming, it will not be used in this chapter. I would point interested learners to *A Course in Phonetics*, by Ladefoged and Johnson (2014), and the associated website.[2]

#### 2.1.1.1 *Vowels*

Vowels are the class of sound produced without a constriction or closing of the vocal tract. They are louder, longer, and easier to hear than consonants. In the vast majority of languages, each syllable (and thus each word) must contain at least one vowel. All languages have at least three distinct vowels and some have many more. Dutch, for instance, may have as many as thirteen (Mees and Collins, 1983). If a language does have only three vowels, it will generally be the vowels in the American English[3] words *beet*, *bot*, and *boot*.

Vowels are conventionally graphically represented in a two-dimensional space. The *x*-axis represents whether the tongue is more forward or more

[2] See www.phonetics.ucla.edu/course/contents.html.
[3] The author is a native speaker of American English.

backward in the mouth, with the left of the diagram representing the front of the mouth. The *y*-axis represents the tongue height or degree of opening of the jaw, with a *high* vowel having the jaw more closed and thus the tongue closer to the roof of the mouth. The vowel in *beet*, for example, is a high front vowel, while the vowel in *bat* is a low front vowel. If you place your hand under your chin and alternate between saying *beet* and *bat*, you will find that your jaw is more open while producing the *bat* vowel than the *beet* vowel. You can observe the difference in backness by gently holding a spoon in your mouth with the end resting on your tongue and repeating the vowel in *beet* and *boot*. The spoon will move toward the front of your mouth while producing the *beet* vowel and toward the back of your throat while producing the *boot* vowel.

The position of individual vowel is usually measured by the location of formants. Formants, or distinct bands of high acoustic energy in the time/frequency domain, are important cues to human vowel perception and fairly easy to annotate. The frontness of the tongue body is captured by the inverse of the first formant and the height is captured by the inverse of the second formant. (The inverses are used so that high vowels are at the top of the graph and back vowels are at the right.) The first formant is abbreviated *F1* and the second, *F2*. Some vowels may be produced with movement of the tongue body during articulation, such as the vowel in American English *buy*. These are called diphthongs and are generally show as vectors, rather than static points, in the F1 by F2 space.

Vowels may also be distinguished by other qualities, such as whether air escapes through the nose during articulation, how long the vowel production is held, and whether the lips are rounded. These differences cannot be easily seen in an F1 by F2 space.

*2.1.1.2 Consonants* Consonants are sounds that have at least one point of construction or closure of the vocal tract. Consonants are generally distinguished by the place and degree of restriction. The constriction can range from a complete closure (in which case the consonant is called a *stop*) to a slight movement of the tongue toward the side of the mouth (which is called a *lateral*). Consonants can be produced anywhere in the vocal tract from the vocal folds themselves (*glottal*) to the lips (*bilabial*).

In addition to the place and manner of articulation, many consonants can be produced either with or without sound being produced by forcing air through taut vocal folds. If sound is produced in this way while articulating the consonant, the resulting consonant is called *voiced*. If no sound is produced, the consonant is *voiceless*. You can feel the differences in voice quality by placing your hand on your throat. For voiced sounds, like the *z* in *zip*, you should feel a buzzing sensation in your throat. For voiceless sounds, like the *s* in *sip* this sensation will be absent.

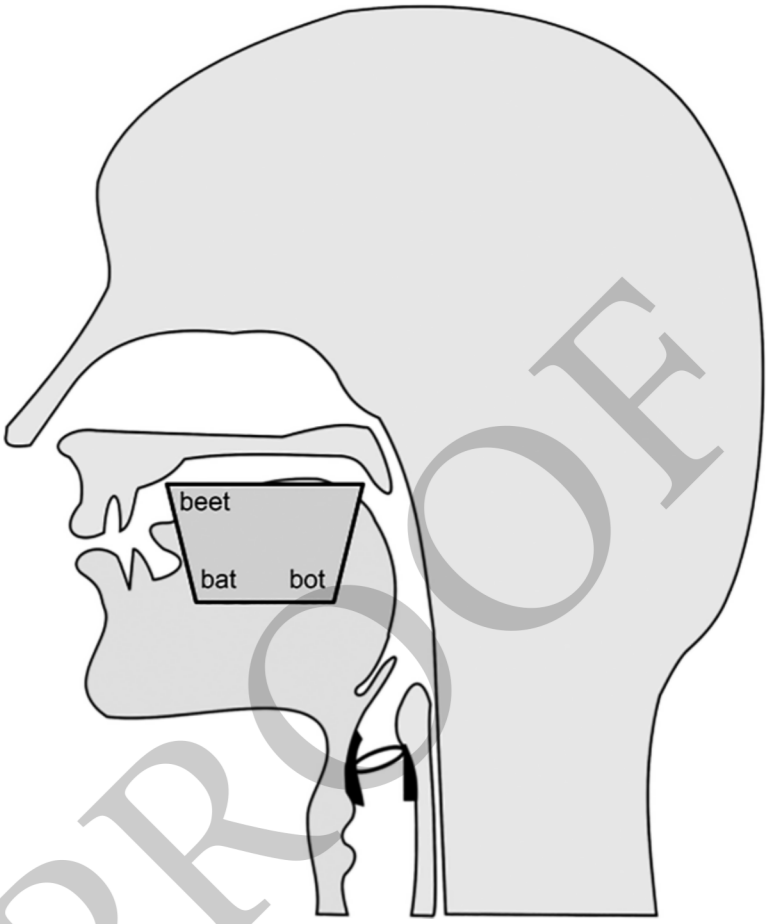


Figure 2.1 A vowel quadrilateral superimposed over a midsagittal diagram of the vocal tract. The position of each word corresponds to the position of the center of the tongue body while producing the vowel in that word.



We can combine these qualities: place and manner of articulation and whether the sound is voiced or voiceless, to precisely describe different consonants.

The sound *t* as in *tee*, for example, is a voiceless alveolar (produced at the alveolar ridge) stop. The sound *m* in *me*, on the other hand, is a voiced bilabial nasal (produced by forcing air through the nasal cavity). I will not discuss all the possible places and manners of articulation here, but interested readers are encouraged to refer to "The Sounds of the World's Languages" (Ladefoged and Maddieson, 1998) as an authoritative guide.

Languages tend to have more constants than vowels, often around twenty-two or so. However the number of consonants in a language can range from six, like the Rotokas language spoken in Papua New Guinea, to over a hundred, like the !X language spoken in Botswana (Haspelmath et al., 2005).

### *2.1.2 What Are Suprasegmentals?*

So far we have discussed only vowels and consonants. However, there is a third category of sound that is used to indicate differences between words: suprasegmentals. Suprasegmentals are the acoustic aspects of speech associated with words or syllables rather than individual speech sounds. Examples include lexical tone, stress, and intonation.

Perhaps the simplest example of a suprasegmental is lexical tone. Lexical tone is the use of pitch (whether the fundamental frequency of the voice has a high or low frequency) to distinguish between words. In Mandarin Chinese, for example, the words for "horse," "mother," "hemp," and "to scold" are all pronounced *ma*. It is the pattern of the pitch used while producing the word that determines the meaning.

Pitch, particularly when looking at prosody or intonation, is sometimes annotated using the ToBI system, developed by Silverman et al. (1992). However, there are many different standards for annotating pitch and tone and they often depend on the language being annotated.

### *2.1.3 Why Does Dialectal Variation Exist?*

Language in use constantly varies. Even a single speaker producing the same word in the same environment will never produce it in exactly the same way twice. Two instances of the same speaker saying the same word will, however, probably be more alike than two different people saying the same word. And two individuals who share many social characteristics (gender, socioeconomic status, level of education, profession, friend network etc.) will probably produce language more similar to each other than two individuals who share none of those characteristics. These systems of variation that pattern with social identity tend to be systematic to the point that they can be summarized by systems of rules. They also occur for groups smaller than "all speakers of a particular language."

It is common to refer to a language as a single monolithic entity: French or Thai or Twi. However, this hides the complexity of language in use. It's more accurate to say that there are many versions of a language. For English, for example, we might refer to American English or British English. Or we might make even finer grain distinctions within American English and talk about Nevada English or Boston English. At an even finer grain, we might talk about the patterns of language use that occur in from a particular social network or even an individual speaker's idiolect. For this chapter, however, we will focus

mainly on differences in language use between geographical regions. This has long been a focus of sociolinguistic research, and many large-scale research projects, such as the *Dictionary of American Regional English* (Cassidy et al., 1985) or *The Atlas of North American English* (Labov et al., 2005), focus on finding and describing language differences between regions.

These different versions, or *dialects*, of a language come from a single historical language. For example, at one time French, Italian, and Spanish would all have once been considered dialects of Latin. However, all living languages are constantly changing at every level of the grammar. (We will discuss some of the specifics of phonetics change later in this chapter.) Since the French, Spanish, and Italian language communities were not in daily contact with each other, as they changed, they slowly diverged from each other to the point where they were no long mutually intelligible. When say that two language varieties are *mutually intelligible* when a speaker of variety A can understand a speaker of variety B and vice versa.

Other than historical changes, the other major source of differences between language varieties is migration. When speakers of a language variety move to a new area and either establish a new language community or integrate with the existing community, features of their language variety will often be incorporated into the dialect of that region. For example, the migration of Hindi-speaking indentured laborers to Fiji resulted in the emergence of new dialect of Hindi known as Fijian Hindi, which incorporates English and Fijian words (Kerswill, 2006).

## 2.2 Vowels

Dialectal variation in vowels has been extensively studied. There are two main ways that the vowel systems of two varieties may differ from one another: each variety may have a different number of vowels or they may have the same number of vowels but produce them in different ways.

Because dialects share a historical root with each other, it is common to compare an existing dialects to an earlier one. Thus, when we talk about something like a vowel *split*, we mean that where there was historically a single vowel in the vowel space there is now more than one. Similarly, a *merger* refers to a single vowel in the area of the vowel space that was previously occupied by more than one vowel. Finally, a vowel *shift* refers to the movement of vowels in the vowel space from some historical starting point, but without changing the total number of vowels.

### *2.2.1 Splits*

Vowel splits generally start as *conditioned* splits that appear only in certain environments. For example, in the Haroi language spoken in Vietnam, the vowel

*u* (as in American English *boot*) is pronounced as *o* (as in Canadian English *boat*) when it comes before a syllable that starts with a vioceless fricative, affricate, or stop (Lee, 1977). If this conditioning environment is lost over time, the result is words that are distinguished only by differences in vowels. This is the case in Swiss German, where the vowel *ai* as in *bait*, split into the *bait* vowel in some words and the *e* as in *bet* vowel in others (Moulton, 1961).

### *2.2.2 Mergers*

Like splits, mergers may be either conditioned or unconditioned. The *pin/pen* merger, for example, found in the Southern United States, means that the vowels in *pin* and *pen* are pronounced in the same way before nasals (Labov et al., 2005). This results in pairs of words like *Wendy* and *windy* or *djinn* and *Jenn* being pronounced the same. Some mergers occur in all contexts, however. The *cot-caught* merger, which is ubiquitous in the United States west of the Rocky Mountains, occurs in all contexts and results in words like *body* and *bawdy* or *hall* and *haul* being pronounced in the same way (Wolfram and Schilling, 2015).

### *2.2.3 Shifts*

In vowel shifts, all or part of the vowel system moves within the F1 by F2 space. Generally, this will be started by a single vowel moving. If it does not merge, it will either crowd another vowel or leave an empty space in the vowel space. This will cause a second vowel to move, either to scoot out of the way or to fill the vacuum, which in turn will push or pull a third, and so on. This process is known as a *chain shift*.

English's vowel system is particularly active. Perhaps the most notable vowel shift in English was the great vowel shift, which took place between the fourteenth and seventeenth centuries and affected all long vowels (Baugh and Cable, 1993). There are also many currently on-going vowel shifts in English, including the New Zealand vowel shift (Hay et al., 2008), the Canadian vowel shift (Clarke et al., 1995), the northern cities vowel shift, and the southern vowel shift (Labov et al., 2005).

## 2.3 Consonants

Historically, sociolinguistic variation in consonants has been the focus of less research than variation in vowels. This is due in part to the fact that consonants, which are generally shorter and quieter than vowels, are harder to measure.

Like vowels, consonantal sounds can undergo both splits and mergers between dialects of the same language. This may result in two related varieties that have a different number of consonants. In Spanish, for example, a historical distinction between the *s* and *th* sounds has been maintained in Spain, while

in American Spanish both sounds are now produced as *s* (Penny, 2002). These differences can also be conditioned: while most varieties of Mandarin have dropped all consonants at the ends of words, the Yue dialects (including Cantonese) have preserved most of them (Norman, 1988).

It is also common for the same underlying sound to be produced in different ways in different dialects. One example is the French *r* sound. In Parisian French, it is produced with the back of the tongue near the uvula (the pink "punching bag" visible at the back of the throat when yawning). In other varieties, such as that produced by older speakers of Midi French, the *r* sound is produced with the tip of the tongue against the alveolar ridge (behind the front teeth) (Klingler and Lyche, 2012).

## 2.4 Suprasegmentals

Dialectal variation in suprasegmentals is less well studied than variation in vowels or consonants. As with vowels and consonants, suprasegmentals may vary in the number used by differing dialects or in the way in which they are realized. As an example of the former, Mandarin spoken in Yantai has three lexical tones, rather than the four found in varieties such as Beijing Mandarin (Bao, 1999). As an example of the latter, while many dialects of Japanese make use of a pitch accent system where the pitch accent used is determined by the word, Japanese spoken in Miyakonoj assigns pitch accent based the position of the word in a sentence instead (Haraguchi, 1988). Vowels, consonants, and suprasegmentals can all vary betrences arise due to historical changes that start in and speech sounds. However, sociophonetic variation is reflected in text as well. Differences in how speech sounds are produced and perceived are often reflected in spelling, particularly in informal user-generated text, such as social media posts. Spelling variation sometimes reflects the user's own linguistic perception, such as a Twitter user from Kentucky tweeting "going to eat at windys." Since the *pin/pen* merger is common in Kentucky, it seems reasonable and *windy's* and *Wendy's* (an American fast-food restaurant) are homophones for this speaker, and this is reflected in their spelling.

The use of spelling to reflect regional phonetic variation is sometimes called *eye dialect* (Bowdre, 1964). In general, users are more likely to use an eye dialect to reflect their spoken variety if they have a strong regional identity (Shoemark, 2017). They are also more likely to write out phonetic markers that are stereotypically associated with the dialect they're attempting to show (Tatman, 2016).

The fact that these variant spellings appear only in user-generated text can lead to poor performance on user-generated text data for models trained on corpora of more formal text. One example of this is the fact that many identification algorithms have been shown to be unable to correctly identify

African American English as English (Blodgett et al., 2016). This may be due in part to spellings such as *dis* for *this* or *wit* for *with*, which reflect the merger of interdental fricatives (the *th* sound in English) with alveolar stops (the *d* and *t* sounds) (Mufwene et al., 1998).

In short, different dialects have systemic phonetic differences, and these differences are sometimes reflected in text. Natural language processing researchers who are aware of these differences and how and why they occur won't be surprised when they encounter them in data and are better prepared to build models to handle such variation.

## 2.5 Conclusion

Vowels, consonants, and suprasegmentals can all vary between dialects. These differences arise due to historical changes that occur in one dialect but for some reason do not spread to others. As a result, different dialects may have more, fewer, or different speech sounds in the same words or phonetic contexts.



## References

Bao, Zhiming. 1999. *The structure of tone*. Oxford University Press on Demand.

Baugh, Albert C., and Cable, Thomas. 1993. *A history of the English language*. Routledge.

Blodgett, Su Lin, Green, Lisa, and O'Connor, Brendan. 2016. Demographic dialectal variation in social media: A case study of African-American English. Pages 1119–1130 of: *Proceedings of the 2016 conference on empirical methods in natural language processing*.

Bowdre, Paul Hull. 1964. *A study of eye dialect*. Ph.D. thesis, University of Florida.

Cassidy, Frederic Gomes, et al. 1985. *Dictionary of American regional English*. Belknap Press of Harvard University Press.

Clarke, Sandra, Elms, Ford, and Youssef, Amani. 1995. The third dialect of English: Some Canadian evidence. *Language variation and change*, **7**(2), 209–228.

Haraguchi, Shoshuke. 1988. Pitch accent and intonation in Japanese. *van der Hulst & Smith*, 123–150.

Haspelmath, Martin, Dryer, Matthew S., Gil, David, and Comrie, Bernard. 2005. *The world atlas of linguistic structures*.

Hay, J., Maclagan, M., and Gordon, E. 2008. *New Zealand English*. Dialects of English. Edinburgh University Press.

Kerswill, Paul. 2006. Migration and language. *Sociolinguistics/Soziolinguistik. An international handbook of the science of language and society*, **3**, 1–27.

Klingler, Tom, and Lyche, Chantal. 2012. Cajun French in a non-Acadian community: A phonological study of the French of Ville Platte, Louisiana. *Phonological variation in French: Illustrations from three continents*, 274–312.

Labov, William, Ash, Sharon, and Boberg, Charles. 2005. *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.

Ladefoged, Peter, and Johnson, Keith. 2014. *A course in phonetics*. Nelson Education.

Ladefoged, Peter, and Maddieson, Ian. 1998. The sounds of the world's languages. *Language*, **74**(2), 374–376.

Lee, Ernest W. 1977. Devoicing, aspiration, and vowel split in Haroi: Evidence for register (contrastive tongue-root position). *Pacific linguistics. Series A. Occasional papers*, 87.

Lucas, Ceil. 2001. *The sociolinguistics of sign languages*. Cambridge University Press.

Mees, Inger, and Collins, Beverley. 1983. A phonetic description of the vowel system of standard Dutch (ABN). *Journal of the international phonetic association*, **13**(2), 64–75.

Moulton, William G. 1961. The dialect geography of hast, hat in Swiss German. *Language*, **37**(4), 497–508.

Mufwene, Salikoko S., Rickford, John R., Baugh, John, and Bailey, Guy. 1998. *African-American English: Structure, history, and use*. Psychology Press.

Norman, Jerry. 1988. *Chinese*. Cambridge University Press.

Penny, Ralph John. 2002. *A history of the Spanish language*. Cambridge University Press.

Shoemark, Philippa. 2017. Debnil Sur, Luke Shrimpton, Iain Murray, and Sharon Goldwater. 2017. Aye or naw, whit dae ye hink? Scottish independence and linguistic identity on social media. Pages 1239–1248 of: *Proceedings of the 15th conference of the European chapter of the Association for Computational Linguistics*. Vol. 1.

Silverman, Kim, Beckman, Mary, Pitrelli, John, Ostendorf, Mori, Wightman, Colin, Price, Patti, Pierrehumbert, Janet, and Hirschberg, Julia. 1992. ToBI: A standard for labeling English prosody. In: *Second international conference on spoken language processing*.

Tatman, Rachael. 2016. Ima spawts guay: Comparing the use of sociophonetic variables in speech and Twitter. *University of Pennsylvania Working Papers in Linguistics*, **22**(2), 18.

Trubetzkoy, Nikolai Sergeevich. 1969. Principles of phonology.

Wolfram, Walt, and Schilling, Natalie. 2015. *American English: Dialects and variation*. Vol. 25. John Wiley & Sons.