

LIN 350: Experimental Phonetics

Scott Nelson

Contents

Overview of the course	5
I Review	9
1 Review of Articulatory Phonetics	11
1.1 The International Phonetic Alphabet	11
1.2 Articulatory Description	13
1.2.1 Place of Articulation	14
1.2.2 Manner of Articulation	17
1.2.3 Voicing	20
1.2.4 Putting it all Together for Consonants	23
1.3 Articulation and Description of Vowels	24
2 Review of Acoustic Phonetics	25
2.1 Spectrograms	25
2.2 Resonance	27
2.3 Silence	32
2.4 Noise	34
II Main Lessons	37
3 f0-pitch-tone-intonation	39
3.1 Sine Waves	39
3.2 Fundamental Frequency in Speech	42
3.3 Intonation	43

III Labs	49
4 Lab 0	51
5 Lab 1	53
Appendices	57
Appendix A Introduction to Praat	59
A.1 What is Praat?	59
A.2 Opening Praat/Sound Files	60
A.3 Text Grids	62
A.4 Measuring Things	64
A.5 Other Resources	66
Appendix B Syllabus	69

Overview of the course

These notes are from the Summer 2023 version of LIN 350: Experimental Phonetics at Stony Brook University. The official course description for LIN 350 is the following.

Introduction to common experimental methods for studying the sounds used in human language. Topics include basic speech acoustics, acoustic analysis, oral and nasal airflow, static palatography, linguography and electroglottography, as well as design of perception experiments. Students will learn the physical processes affecting each experimental variable and common methods of analyzing each kind of data. Students will get hands-on experience with each analysis method and will use two or more types of data to explore a hypothesis about sound structure in English or some other language of interest. Students will learn how to use software for making measurements and analyzing data. Students will learn to assess the validity of claims about language based on their understanding of the scientific method as applied to speech. The course will give students a solid foundation for further courses in laboratory skills relevant to assessment of normal and disordered speech and for pursuing research, either as undergraduate researchers, or in the early stages of graduate work.

In this version of the course we will primarily focus on acoustic and auditory analysis of speech. Articulatory analysis will be limited to discussion only. This is not because articulatory analysis is less important, but rather because it requires more resources and specialized equipment. On the other hand, acoustic and auditory analysis can be done (at least to a first approximation) with only a laptop and preferably some headphones.

The goal of this course is to teach students how to do their own analyses and the best way to learn is by doing. Therefore, this course is based around labs that will simultaneously guide students through an analysis, while also reinforcing the physical correlates of linguistic structure in the sound domain.

A schedule of topics and labs is listed on the following page. To conclude the course, each student will do an independent phonetic analysis on a topic of their choosing. The phonetic analysis must involve a comparison of some type between two speech communities. This can include speakers of different languages, speakers of the same language but different dialects, or disordered and non-disordered speakers. Each student will give a 15-20 minute presentation on their project as well as write a short paper explaining their findings. More details on the structure of the course can be found on the course syllabus.

Class Schedule

Week	Day	Time	Topic
1	7/11 (T)	Asynchronous	Lecture - review of Articulatory Phonetics
		Asynchronous	Lecture - review of Articulatory Phonetics
		Asynchronous	Lecture - review of basic Acoustic Phonetics
	7/13 (Th)	Asynchronous	Lecture - review of basic Acoustic Phonetics
		Asynchronous	Praat - Introduction
		Asynchronous	Lab 0: Analyzing Speech Data
2	7/18 (T)	9:30–10:30	Lecture - f0, pitch, tone, intonation
		10:40–11:40	Praat - measuring f0
		11:50–12:55	Lab 1: Intonation
	7/20 (Th)	9:30–10:30	Lecture - lexical tone + contrast
		10:40–11:40	Praat - automatic analysis + drawing
		11:50–12:55	Lab 2: Tone
3	7/25 (T)	9:30–10:30	Lecture - the acoustic theory of speech production
		10:40–11:40	Praat - measuring formants
		11:50–12:55	Lab 3: Formants
	7/27 (Th)	9:30–10:30	Lecture - acoustic correlates of stress
		10:40–11:40	Praat - measuring intensity + duration
		11:50–12:55	Lab 4: Stress
4	8/1 (T)	9:30–10:30	Lecture - voice onset time
		10:40–11:40	Praat - point tiers + measuring vot
		11:50–12:55	Lab 5: Voicing
	8/3 (Th)	9:30–10:30	Lecture - acoustic correlates of place of articulation
		10:40–11:40	Praat - spectral measurements
		11:50–12:55	Lab 6: Place of Articulation
5	8/8 (T)	9:30–10:30	Lecture - intro to speech perception
		10:40–11:40	Praat - running experiments
		11:50–12:55	Lab 7: Categorical Perception
	8/10 (Th)	9:30–10:30	Lecture - theories of speech perception
		10:40–11:40	Praat - ???
		11:50–12:55	Lab 8: Perceptual Similarity
6	8/15 (T)	9:30–10:30	Lecture - more experimental techniques
		10:40–11:40	Project Presentations
		11:50–12:55	Project Presentations
	8/17 (Th)	9:30–10:30	Project Presentations
		10:40–11:40	Project Presentations
		11:50–12:55	Lecture - course recap

Part I

Review

Lesson 1

Review of Articulatory Phonetics

In order to take this course, students are required to have received a grade of C or higher in LIN 201: Phonetics at Stony Brook University. The first two lessons in this course will provide a basic review of the material covered in LIN 201. It is expected that all students enrolled in this course have a strong familiarity with the [International Phonetic Alphabet](#). Being able to use IPA symbols in your typed documents is a useful skill. The [typeit](#) website is an easy way to copy and paste non-native symbols into your documents, but it better to install an IPA keyboard directly to your computer. I recommend an IPA Unicode keyboard. Information on how to install this type of keyboard layout on your computer can be found [here](#). If you need help with installation, please let me know.

1.1 The International Phonetic Alphabet

The International Phonetic Association was formed in Paris in 1866 (originally called *Dhi Fonètik Tîcerz' Asociécon*) and one of its main goals was to create a phonetic alphabet that could be used to help second language learners pronounce foreign words in a more native like way. Since then, the International Phonetic Alphabet (IPA) has become one of, if not the, most commonly used phonetic alphabets.

Phonetic alphabets are a useful tool for representing language because their use ensures that anyone reading a word will know exactly how to pronounce it. The reason that this can happen is due to a one to one mapping between symbol and sound. What this means is that every symbol of the

phonetic alphabet signifies one and only one sound. This is the opposite of the English alphabet where you may have many symbols representing a single sound, or one symbol representing many sounds.

For example, the following sounds all contain an identical sound, but it is spelled differently in each form:

sea
teeth
me
city
achieve
perceive

Notice that the sound corresponding to how we pronounce the letter 'e' is in all these words, but in each case, it is a different set of letters that represent that sound. In English spelling, 'ea', 'ee', 'e', 'y', 'ie', and 'ei' can all correspond to the same sound. This is a **many to one mapping**.

Now, let's look at the following set of words that contain the same sequence of letters that correspond to a different sound in each word:

ghost
rough
though

In these examples, we see the letter combination 'gh' in each word, but it corresponds to a different sound (or lack of any sound) in each word. In 'ghost' the 'gh' is similar to what we might expect for a 'g' sound. In 'rough' the 'gh' is similar to what we might expect for an 'f' sound. And in 'though' the 'gh' corresponds to the absence of any sound. This is a **one to many mapping**.

The IPA mediates this problem by providing a **one to one mapping** between each sound and a corresponding symbol. This means that when we see a word transcribed this way, we know exactly how to pronounce it. When transcribing words in IPA, it is standard to put the symbols inside square brackets like this: [həlou]. By doing so, we can differentiate between spelling and transcription.

Below you will find the symbols for English consonants and English vowels. Each symbol is paired with a set of words that contain the specific

sound. The sounds in the English word that contain the specific sound are bolded.

[p]	pin, lip, happy	[b]	bin, club, hobby
[t]	tin, mitt, atomic	[d]	den, hid, adorable
[k]	king, elk, tricky	[g]	game, flag, piggy
[tʃ]	chip, coach, itchy	[dʒ]	jam, huge, badger
[f]	fit, elf, stuffy	[v]	vine, glove, oven
[θ]	think, tooth, ethics	[ð]	this, smooth, brother
[s]	sale, mass, messy	[z]	zero, breeze, basil
[ʃ]	ship, fish, pushy	[ʒ]	(genre), measure, beige
[h]	hill, ahead	[ŋ]	ring, singer
[m]	mouse, gum, amish	[n]	net, win, honey
[l]	lost, pill, apollo	[ɹ]	round, core, curry
[w]	witch, awake	[j]	yellow, beyond

Table 1.1: English consonant symbols and corresponding words that contain each sound.

[i]	sheep	[ɪ]	ship	[eɪ]	game	[ɛ]	pen	[æ]	cat
[u]	food	[ʊ]	foot	[oʊ]	soap	[ɔ]	cough	[ɑ]	cot
[ʌ]	cut	[ɜ̄]	bird	[aɪ]	tie	[aʊ]	house	[ɔɪ]	boy

Table 1.2: English vowel symbols and corresponding words that contain each sound.

If we return to our first example from above, we now have a way to show that, despite the spelling differences, each word contains the same vowel sound [i].

1.2 Articulatory Description

Recall that articulatory phonetics is the study of how we produce linguistic sounds. One thing that we can do is describe the different ways we position our articulators to make the individual speech sounds. Articulators can be broken down into two types: **active articulators** which move and **passive articulators** which don't/cannot move.

The IPA chart for consonants is given below. There are three basic terms we can use to describe any linguistic sound. **Place (of articulation)** refers to where along the vocal tract a constriction is being made. The different places of articulation are represented in the columns of the IPA chart. **Manner (of articulation)** refers mainly to the size of the constriction being made, but also can refer to secondary properties such as whether or not air is flowing through the nose and if air flows through the center of your mouth or through the sides of your mouth. The different manners of articulation are represented in the rows of the IPA chart. **Voicing** refers to whether or not your vocal folds are vibrating while making a sound. Certain sounds vary only in whether or not the vocal folds are vibrating while they are being made. That is, certain sounds can have the same place and same manner, but vary along this voicing dimension. Sounds are said to be **voiced** if the vocal folds are vibrating and **voiceless** if they are not. This contrast is represented by orientation within cells in the IPA chart. Sounds on the left side of the cell are voiceless and sounds to the right side of the cell are voiced.

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2015)											
CONSONANTS (PULMONIC) © 2015 IPA											
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		t̪ d̪	c ʃ	k g	q χ	G	?
Nasal	m	n̪		n		n̪	n̪	n̪	n̪	N	
Trill	B			r						R	
Tap or Flap		v̄		r̄		t̄					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	s̪ z̪	ç j	x y	χ ʁ	h ʕ	h̄ f̄
Lateral fricative				ɬ	ɬ̄						
Approximant		v̄		ɹ̄		ɻ̄	j̄	w̄			
Lateral approximant				l̄		ɻ̄	ɻ̄	L̄			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

Figure 1.1: Official IPA consonant chart

1.2.1 Place of Articulation

Every place of articulation on the IPA chart is really a combination of **active/lower articulators** and **passive/upper articulators**. The active articulators are the lower lip, the tongue tip/blade, the front of the (body of the) tongue, the back of the (body of the) tongue, and the tongue root.

The passive articulators are the upper lip, the upper teeth, the alveolar ridge, the (hard) palate, the velum, the uvula, and the pharynx. Figure 1.2 shows where each of these articulators are located. It also includes some landmarks which are not used for articulation (such as the upper teeth), as well as the glottis which is used for making the glottal speech segments [h], [f], and [?] that don't really adhere to the same active/pассивive combination as all the other sounds.

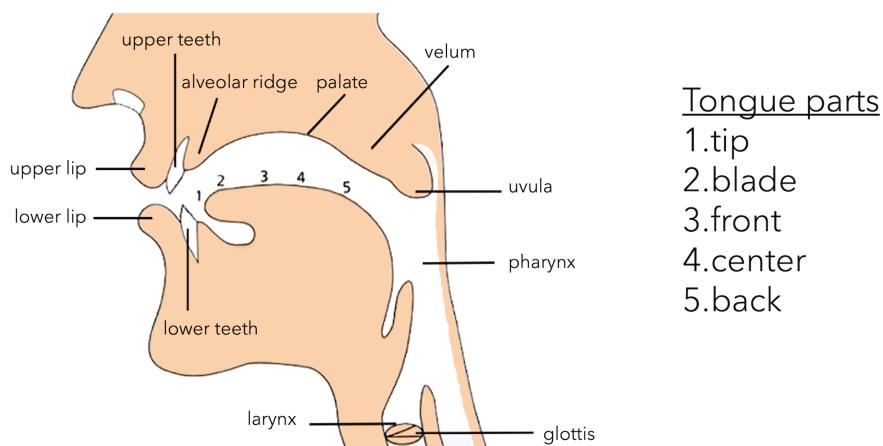


Figure 1.2: A list of most articulators used in speech production.

Some articulator pairs can be deduced from the name alone. For example, a bilabial sound uses the lower lip and upper lip as its articulators (*bi*-meaning two and *labial* meaning lips). Other pairs cannot be deduced from the name alone since only the passive articulator is given. We know that a palatal sound is made at the palate which is a passive articulator (listed above), but what is the active articulator involved in making these types of sounds? The answer here is that it is the active articulator that is closest to it (in this case, the front of the tongue). So Dental/Alveolar/Postalveolar sounds are all made with the tongue tip/blade as the active articulator, Palatal sounds are made with the front of the tongue as the active articulator, Velar and Uvular sounds are made with the back of the tongue as the active articulator, and Pharyngeal sounds are made with the root of the tongue as the active articulator. You may be wondering why we did not include Retroflex sounds anywhere in this discussion. This is because Retroflex is not a place of articulation in the same way that these other

sounds are (despite being lumped in with it on the IPA chart). Instead, a Retroflex sound is made in the same general area as an alveolar sound, only the tongue tip curls backwards and the underside of the tongue is used as the active articulator.

[Seeing Speech](#) is an interactive website created by phoneticians in Scotland. It is an extremely useful resource for students and anyone trying to learn about the phonetic properties of speech. The link here takes you to a clickable IPA consonant chart. Here, you are able to click on a speech sound which will trigger a video that provides audio and video of that specific consonant in between two vowels. At the top of the page you are able to choose between three video types: ANIMATION, MRI, and ULTRASOUND. I suggest ignoring the ULTRASOUND setting as it is hard for newcomers to discern what is going on. As an activity, click on one of the alveolar plosive sounds ([t] or [d]). Notice how the tip of the tongue makes full contact with the alveolar ridge. Now, click on one of the retroflex plosive sounds ([ʈ] or [ɖ]). Here, you should see that the tongue tip curls backwards and it is the underside of the tongue that is making contact in the area between the alveolar ridge and the palate (it may be useful to use the ANIMATION setting here to get a clearer view of what is going on).

We end our discussion of place of articulation by talking about the tongue. The tongue is arguably the most important speech organ since it is involved in creating a majority of the consonant sounds in the world's languages and in creating all of the vowel sounds. Broadly, the tongue can be divided into three areas: the blade of the tongue, the body of the tongue, and the root of the tongue. The tip of the blade also gets its own special term. Many types of sounds can be articulated using either the tip or the blade. In LIN 201 they were lumped together which resulted in the tip/blade being referred to as almost a singular unit. Later in this course we will discuss the variation in tip/blade articulations in more detail.

The body of the tongue can also be divided into three parts: the front, the center, and the back. Only the front and the back are used for making speech sounds. This is a result of the way this part of the tongue moves. The back of the tongue can be thought to move diagonally back and up towards the velum and uvula, while the front of the tongue can be thought to move diagonally forward and up towards the palate. Finally, there is the root of the tongue, which is not pictured in the diagram above, but can be moved forward and back. For consonants, it is moved back to make pharyngeal fricatives such as [h] and [ʃ] (found in certain dialects

of Arabic and other languages).

1.2.2 Manner of Articulation

Manner of articulation is a catch-all term that refers to multiple things. It can refer to the **degree of constriction** such as **stop**, **fricative**, or **approximant** (you may also sometimes see **affricate** used though this is not included on the official consonant section of the IPA chart). Figure 1.3 below shows the dynamic trajectory of each of these degrees of constrictions. For stop consonants, the active and passive articulator completely touch, blocking all air from exiting the mouth. For fricative consonants, the active articulator gets extremely close to the passive articulator in order to generate what is called **turbulent airflow**. This manifests as a hissing like sound and is the same phenomenon as when a car window is just slightly cracked open. For approximant consonants, the active articulator moves into the airstream towards the passive articulator, but not enough to generate turbulent airflow. Affricate consonants are a combination of stops and fricatives. So when these sounds are made, the active articulator and passive articulator completely touch for a brief moment before the active articulator pulls away just enough to create turbulent airflow.

Manner of articulation can also refer to where the air flows through mouth. If it flows through the center of the mouth the term **central** is often used (though not listed on the IPA chart). If it flows through the sides of the mouth, the term **lateral** is used. Lateral sounds are made by curling the sides of your tongue up so that air is blocked from going down the center of your mouth and is instead forced around the sides. In English, we only have one lateral consonant: [l]. To get an idea of how the air is flowing through the side of your mouth we can do a quick activity. Make a [l] sound and hold it so your articulators stay in place. Now, breath in. What you should feel is cold air coming in along the sides of your mouth. This is just the opposite of what is happening while you are making the [l] sound.

The final dimension that manner of articulation can refer to is **nasality**. This refers to whether or not the air flows only through the mouth (called an **oral** sound) or through both the nose and mouth (called a **nasal** sound). Physically, this involves your velum raising and lowering to block air from escaping through the nose. If we look at the IPA chart, we notice that

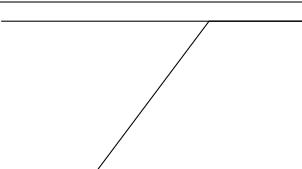
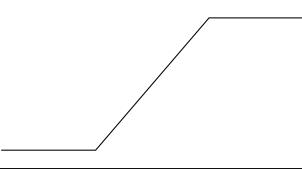
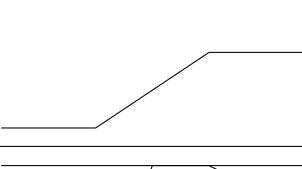
Type	Articulator Visualization		Description
Stop	upper/passive articulator		Lower/active articulator touches upper/passive articulator and blocks all air from escaping through the mouth
	lower/active articulator		
Fricative	upper/passive articulator		Lower/active articulator moves close enough to upper/passive articulator to generate turbulent airflow
	lower/active articulator		
Approximant	upper/passive articulator		Lower/active articulator moves into the airstream towards the upper/passive articulator, but not close enough to generate turbulent airflow
	lower/active articulator		
Affricate	upper/passive articulator		Lower/active articulator begins by moving to touch the upper/passive articulator but then retracts slightly away but remains close enough to generate turbulent airflow
	lower/active articulator		

Figure 1.3: Descriptions of the different degrees of constrictions used in speech

there is only one row for nasals. This suggests that they are all made in a similar way. Let's return to our discussion about degree of constriction from above. Can you guess what the degree of constriction for a nasal sound is? Make the [m] or [n] sound and think about what your active and passive articulators are doing. If you guessed "stop" as the degree of constriction then you are correct. Nasal sounds are all made by stopping air from exiting the mouth and lowering the velum so that it flows only through the nose. This is why you can hold a [m] sound out but not a [b] sound despite them both having the same active and passive articulators and degree of constriction.

Let's return now to the [Seeing Speech](#) consonant chart and look at nasal sounds and how they relate to oral stops. We will use the [m] and [b] example from the previous paragraph. First, click on the [b] and look at

what the articulators are doing. The lower lip comes up and touches the upper lip to block all air from exiting the mouth. Now, play the video again and notice that the velum raises up to block any air from going through the nose. Click on the [m] and do the same thing. Notice that the lower lip and upper lip are doing the exact same thing that they were doing during a [b] sound. Then, pay attention to the velum. Do you notice how it lowers to allow air to flow through the nose?

Before we move on, let's see how these three terms relate to the row headings of the IPA chart. The first row contains **plosive** sounds. All of these sounds have a stop degree of constriction, central airflow, and are oral. Colloquially, they are often just referred to as stops. The second row contains **nasal** sounds. All of these sounds have a stop degree of constriction, central airflow, and are nasal. Therefore, the only difference between nasals and plosives is the location of the velum and whether or not air is able to flow through the nose. Let's temporarily skip rows three and four.

The fifth row contains **fricative** sounds. All of these sounds have a fricative degree of constriction, central airflow, and are oral. Therefore, the only difference between fricatives and plosives is the degree of constriction. The sixth row contains **lateral fricative** sounds. As you may guess, all of these sounds have a fricative degree of constriction and are oral, but unlike the row above they have lateral as opposed to central airflow. The seventh row contains **approximant** sounds. All of these sounds have an approximant degree of constriction, central airflow, and are oral. Therefore, the only difference between plosives, fricatives, and approximants is the degree of constriction (or how close the articulators are to one another). The final row contains **lateral approximants** which once again are identical to the preceding row of **approximants** except for the fact that this row has lateral airflow as opposed to central airflow.

If we return to rows three and four we see rows for **trill** sound and **tap/flap** sounds. Trill sounds are made by jutting an articulator into the air stream in such a way that it naturally causes the airstream to open and close. There are only three trills known to exist in human language: the bilabial trill [b], the alveolar trill [r], and the uvular trill [r̪]. Tap sound are essentially really fast stop sounds in that the two articulators touch, but only for a brief second. The alveolar tap [ɾ] is found in English words such as *butter* ['bʌtəɾ] or *better* ['betəɾ]. Both trills and taps have central airflow and are oral. Descriptively, their degree of constriction

is their name: **trill** or **tap**. In actuality, it's a bit more complicated. In both cases, the oral cavity is completely cut off. The difference between these sounds and stops is the amount of time in which this occurs. For a tap, the lower/active articulator just briefly makes contact with the upper/pассив articulator. For a trill, the lower/active articulator is caused to make repeated brief contact with the upper/pассив articulator due to the bernoulli effect (discussed in relation to voicing below). Figure 1.4 shows a schematization of both of these types of sounds.

Type	Articulator Visualization		Description
Tap	upper/pассив articulator		Lower/active articulator touches upper/pассив articulator and immediately pulls away
Trill	upper/pассив articulator		Lower/active articulator moves into the airstream with an appropriate tenseness. This causes it to be pulled/pushed away from the upper/pассив articulator at semi-regular intervals

Figure 1.4: Descriptions of the different degrees of constrictions used in speech

To summarize, the rows of the IPA are labeled based on manner of articulation. Manner of articulation is broken up into three categories: degree of constriction, airflow location, and nasality. As the rows descend, the degree of constriction gets more and more open. If any degree of constriction allows for laterality, that is placed below the non-lateral row of the same degree of constriction. Nasality only appear in a single row where it is applicable.

1.2.3 Voicing

Voicing, at its most basic, refers to whether or not your vocal folds are vibrating while you are making a sound. In general, any type of sound can be voiced or voiceless, but it is usually the case that stops, fricatives, and affricates are the only types of sound that have a voicing **contrast**. When linguists use the term contrast, they mean that by changing a parameter

(voicing for our current case, but that's not the only parameter that can be used contrastively), the meaning of a word will change since there is a new, meaningful sound being made. Take the difference between [s] and [z] as an example. Both sounds have a fricative degree of constriction, central airflow, and are oral. Additionally, they both are made by bringing the tongue tip/blade (the active articulator) extremely close to the alveolar ridge (the passive articulator). The only way that they vary is based on whether or not the vocal folds are vibrating. But this is a meaningful difference since we can differentiate between words like *sit* [sit] and *zit* [zit] or *bus* [bʌs] and *buzz* [bʌz].

To get an idea of how only voicing changes between [s] and [z], try the following activity. Make an [s] sound and hold your hand over your vocal folds. Now switch to making a [z] sound. If you did it correctly, you should notice that the only change you can feel is the vibration of your vocal folds. The rest of your articulators should have stayed in place. You can try this exercise with other fricative pairs in English such as [f]/[v], [θ]/[ð], and [ʃ]/[ʒ].

In LIN 201 you were introduced to the basic articulatory property of voicing. It was based around the **bernoulli effect**. This is a general principle that states, given certain assumption, an increase in speed (of a fluid) occurs simultaneously with a decrease in pressure. When it comes to speech, the speeding up occurs when air tries to pass through the narrow space of the trachea that is the result of the vocal folds being placed closely together without a change in overall pressure. Since the speeding up occurs in between the vocal folds, this causes a decrease in pressure in that area and the vocal folds collapse in on each other. Pressure build up below the vocal folds causes them to be pushed back apart. By maintaining a satisfactory level of tenseness, the process will continue as the vocal folds spring back into their original position. Figure 1.5 schematizes this in an abstract way.

Generally, we can think of the process of voicing as a cycle that will only stop when one of the necessary conditions change:

1. Air flows between the vocal folds
2. Bernoulli effect pulls the vocal folds together
3. Pressure builds up beneath the closed folds

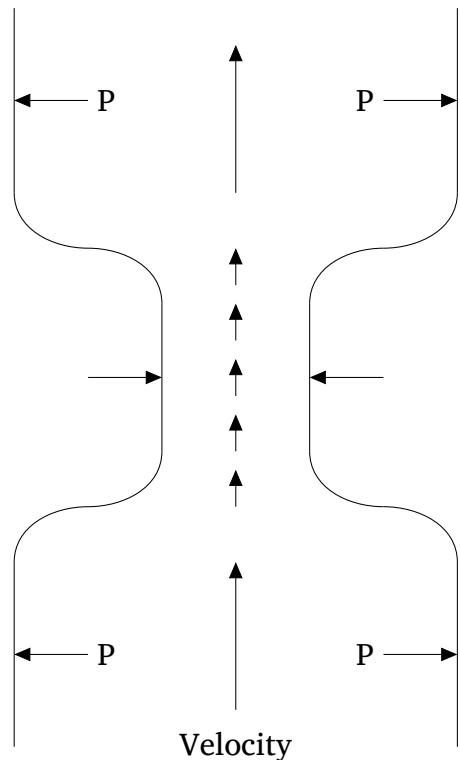


Figure 1.5: Schematization of the bernoulli effect in speech production. The increased speed as air travels through the vocal folds causes a decrease in pressure which in turn causes them to collapse in on themselves. The steady flow of air increases pressure below the vocal folds and causes them to be pushed back open. This process repeats and creates vocal fold vibration a.k.a. “voicing”.

4. Vocal folds are pushed apart by the high pressure
5. Due to anatomical reasons, the vocal folds return to their original position
6. The cycle continues

You may remember from LIN 201 that there were three necessary conditions for voicing: the vocal folds are close together and protruding into the airstream (to speed up air flow), the vocal folds are appropriately tensed (so that they automatically return to their original position), and there is sufficient flow of air (to cause the speed up/pressure drop). If any of these conditions are not met, the cycle will cease to run.

1.2.4 Putting it all Together for Consonants

Now that we've reviewed the core descriptive terms, let's put it all together. You may remember that a lot of time in LIN 201 was spent giving articulatory descriptions of consonants. This may have seemed tedious, but we wanted to make sure that you had a strong understanding of how the various speech articulators worked in conjunction to create speech sounds. It is especially important when thinking about non-native sounds. If you know how the articulators work to make certain types of sounds in your language, you should be able to configure them in new ways to figure out how to make similar sounds in another language. We fit consonants into the following template: VOICING, PLACE, AIRFLOW, NASALITY, DEGREE OF CONSTRICKTION. The options for airflow are *central* and *lateral*. The options for nasality are *oral* and *nasal*. Table 1.3 gives a few examples of consonant descriptions.

In addition to giving static definitions of consonants, we can also describe how the positions of the articulators change over time as they move from one consonant to the next. For example, we can describe how the articulators change in the word *helper* as they go from an [l] to a [p]: the vocal folds separate as it transitions from a voiced to a voiceless sound, the lower lip moves to touch the upper lip to prevent any air from escaping the mouth, the tongue tip moves away from the alveolar ridge and the sides of the tongue return to a position that allows for central airflow. You should be able to describe these types of movements for a transition between any two consonants.

	VOICING	PLACE	AIRFLOW	NASALITY	D.O.C
ɳ	voiced	retroflex	central	nasal	stop
χ	voiceless	uvular	central	oral	fricative
v	voiced	labiodental	central	oral	approximant
‡	voiceless	alveolar	lateral	oral	fricative
dʒ	voiced	palatoalveolar	central	oral	affricate
c	voiceless	palatal	central	oral	stop

Table 1.3: Full articulatory descriptions for a variety of consonants

1.3 Articulation and Description of Vowels

Our review of the articulatory properties of vowels will be much shorter. In the second half of LIN 201 you may remember that it was revealed that vowels are organized by the IPA in terms of their acoustic properties, specifically values for the first two formants. Nonetheless, there are some loose connections between placement on the vowel chart and articulatory properties. **Tense/lax** roughly corresponds with advanced/retracted tongue root. In English this isn't *really* the case, but in other languages it is. **Vowel height** roughly corresponds with the size of the cavity between the tongue dorsum which can be related to its "height". **Vowel front-backness** roughly corresponds with which part of the tongue body is highest while the vowel is being made. **Roundness** is the articulatory property that is most direct: it refers to whether or not the lips are rounded while the vowel is being made. There is no "degree of constriction" for vowels, but they are in some sense making a constriction. It is less than what it is for an approximant. Recall that in consonants, the final term is the degree of constriction. You may remember that in LIN 201 we had the final term when describing vowels always just be "vowel". You can think of that as the vowel's degree of constriction.

This gives us a parallel way to describe vowels in a similar way that we described consonants. The template for vowels is: TENSENESS, HEIGHT, FRONTNESS, ROUNDING, VOWEL. The options for tenseness are *tense* and *lax*. The options for height are *high*, *mid*, and *low*. The options for frontness are *front*, *central*, and *back*. The options for rounding are *round* and *unround*. Our "articulatory" descriptions really fall apart when it comes to both diphthongs, as well as the mid-central vowels. For a much more in depth explanation on the articulation of vowels, see Gick et al. (2013).

Lesson 2

Review of Acoustic Phonetics

The review of acoustic phonetics will be brief for two reason. One, not much is covered in LIN 201. Two, the remainder of this course will primarily focus on acoustic phonetics. Recall that linguistic phonetics is largely broken up into three subareas: articulatory phonetics, acoustic phonetics, and auditory phonetics. In this course, you will learn about ways that articulation is measured but will not get any hands on experience. Our primary focus will be acoustic measurements and you will get plenty of hands on experience in this area. The last couple lessons will focus on auditory phonetics and you will have a chance to run some simple speech perception experiments.

2.1 Spectrograms

Spectrograms are visual representations of sound waves that phoneticians and other speech scientists use to analyze the acoustic properties of speech. They are three dimensional graphs that represent time (x-axis), frequency (y-axis), and amplitude (z-axis). The z-axis is represented by the level of darkness in the spectrogram. So dark spots indicate frequencies in which there is a large concentration of energy. Above each spectrogram below is the corresponding sound wave. These graphs are two dimensional where the x-axis is once again time and the y-axis is (overall) amplitude.

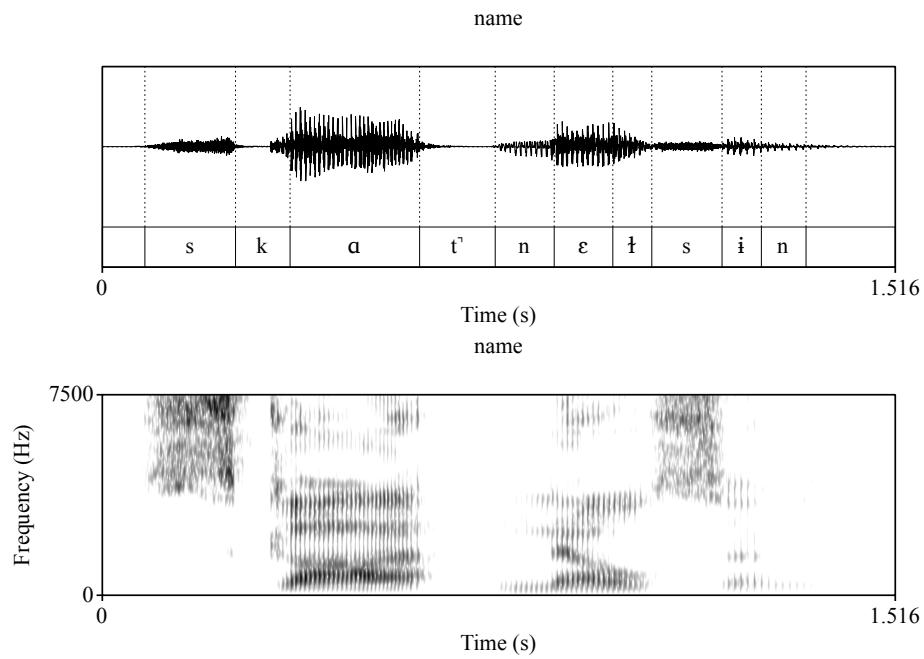


Figure 2.1: Waveform and spectrogram for [skat̚ nəts̚ɪn].

On a spectrogram, there are three basic types of information we are interested in: resonance, silence, and noise. In LIN 201 you were introduced to the **source-filter theory**. We will cover this more in depth in this class, but you may remember that the basic idea is that your vocal folds act as a source of sound (via vibration) and this sound is then shaped by the configuration of the articulators in the vocal tract. Of course, this can't be the entire picture because we have voiceless sounds as well! One crude way we can split the difference is to say that vocal fold vibration leads to **resonance** while the turbulent airflow seen in voiceless fricatives, the second half of approximants, and with aspiration leads to **noise**. These are both technical terms that we will get into in more depth later, but you hopefully remember that resonance shows up on a spectrogram as concentrated dark bands of energy while noise has no apparent pattern. **Silence** is the third thing we can look for in a spectrogram and appears as white nothingness. Below, we review these terms in slightly more detail and show examples.

2.2 Resonance

The basic idea of resonance is that there are certain frequencies of sound that are amplified when passing through an object. It is usually not a *specific* frequency even though it is often talked about that way. In general, it is a band of frequencies but we refer to the central frequency as the location of the resonance. In general, everything has a resonant frequency that will cause it to vibrate. If you have 45 minutes to spare and a lot of patience, you can listen to this piece of sound art by Alvin Lucier called *I am Sitting in a Room*. It demonstrates the resonant frequencies of a room by continuously playing back the same recording into a room over and over again. Each cycle, the frequencies outside the resonant areas are filtered out until all that is left is the resonant frequencies of the room itself!

When it comes to speech, these resonances are called **formants**. They are caused by placing the vocal tract into different configurations. They most clearly show up in vowels. On a spectrogram, they appear as dark horizontal bands. There are many of them, but primarily we are interested in the first three: F1, F2, F3. The numbering of formants begins from the bottom and moves up. If you're wondering why we don't include the dark band at the very bottom of the frequency range, this is the voicing bar which we will return to in a moment.

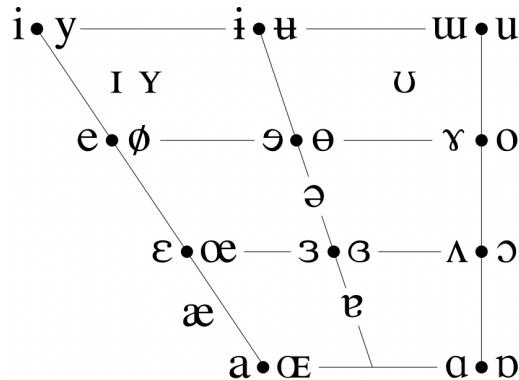


Figure 2.2: Official IPA vowel chart

For vowels, the IPA uses the values for F1 and F2 to chart their location. F2 is reversed on the x-axis and F1 is reversed on the y-axis. This means that vowels with a high F2 value are more front while vowels with a high

F1 value are more low. Figure 2.2 is the official IPA vowel chart. If we were to measure the F1 and F2 values for each of these vowels and plot those directly, we would get something that looks basically identical to this.

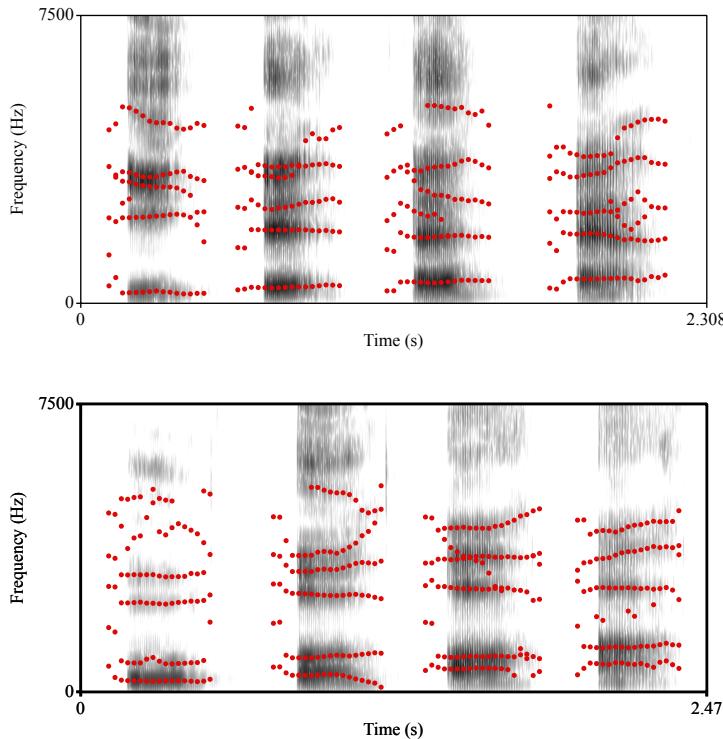


Figure 2.3: Spectrograms with automatic formant detection overlayed. Top row are the vowels [i], [ɪ], [ɛ], [æ]. Bottom row are the vowels [u], [ʊ], [ɔ], [ɑ].

We can look at the formant values for some American English vowels directly. Figure 2.3 includes spectrograms for four of the front and four of the back vowels and are overlayed with the location of the formants in red. The top row are the front vowels [i], [ɪ], [ɛ], and [æ]. Notice that we see a gradual rise in F1 as we go from left to right, indicating that the vowels are getting lower in the vowel space as expected. We also see F2 drop a little which may be unexpected since they are all “front vowels”, but if we look at the IPA chart we do notice a diagonal line along the

front dimension indicating that F2 should lower for the front vowels as F1 increases.

The bottom rows are the back vowels [u], [v], [ɔ], [ɑ]. Since we're going from a high vowel to a low vowel, we should also expect to see F1 increase from left to right which we do. In general, the back vowels F2 is not correlated with F1, but you may notice that my vowel [ɑ] has a higher F2 than the other three "back" vowels. This is a property of my dialect and is likely due to having a lower [ɔ]. As you may remember some people have these vowels merged into a single vowel. My fronting is likely to ensure that the two are not merged.

While vowels have the strongest formants (resonance), many consonants also have formants as well. The reason that vowels are the most resonant is related to the fact that the oral cavity is least restricted in the production of vowels. But approximants also have visible formants in the spectrogram. The approximants [j] and [w] are often called **glides** or **semi-vowels**. This is because they are quite similar to the vowels [i] and [u] in their spectral make up. The formants in [j] and [w] are similar to what you will find in [i] and [u], but they typically are not as strong and the segment is shorter in duration.

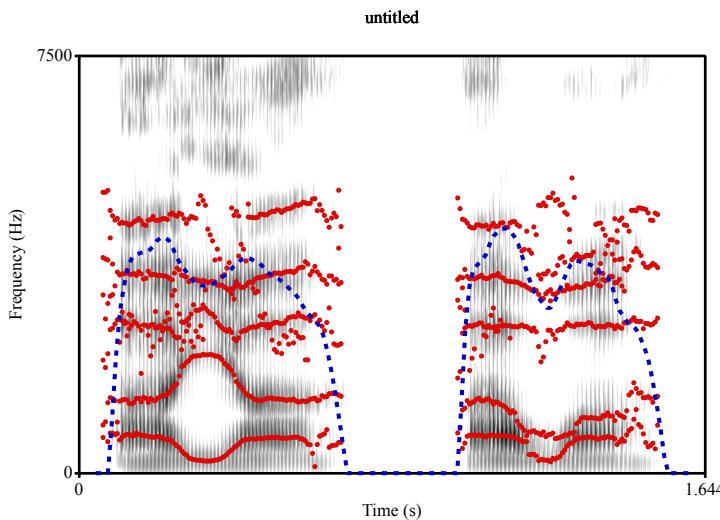


Figure 2.4: Spectrograms for [ajə] (left) and [awə] (right). Red dots are formant trajectories. Blue dashed line shows intensity profile.

Figure 2.4 shows spectrograms for [ajə] and [awə]. The red dotted

lines once again show the formant trajectories. These spectrograms also included a blue-dashed line showing the intensity profile. Notice that in both sequences, the intensity drops in the middle while the glide is being made, even though there are still formants. This is a characteristic of glide sounds. Also, notice that both glides match the formant profiles of their corresponding vowels: [j] on the left has a high F2 and low F1 just as [i] does while [w] on the right has a low F2 and low F1 just as [u] does.

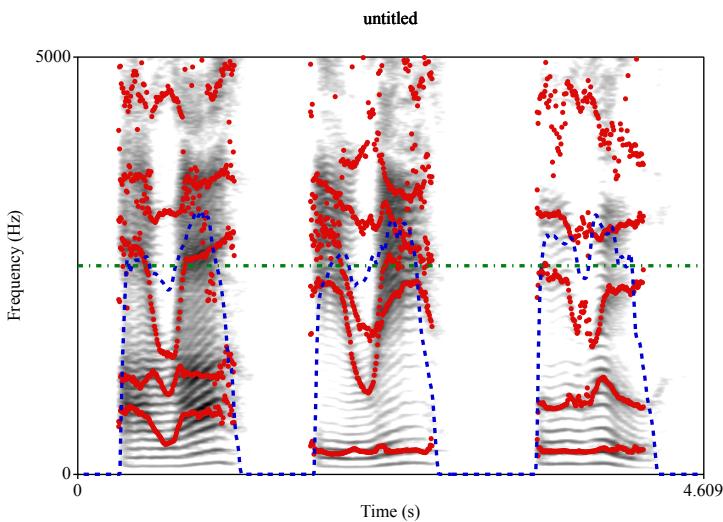


Figure 2.5: Spectrograms for [aɪə] (left), [iɪi] (center) and [uru] (right). Red dots are formant trajectories. Blue dashed line shows intensity profile. Green horizontal line at 2500 Hz.

The English non-glide approximants [l] and [ɹ] also have formant patterns that are noticeable. Starting with [ɹ], the main acoustic property of this sound is a lowered F3 (and to an extent a lower F2) as well. Figure 2.6 shows spectrograms for the English [ɹ] in between the low back vowel [ɑ] (left), the high front vowel [i] (center), and the high back vowel [u] (right). Once again, formant and intensity profiles are given in red and blue. All of these “words” were pronounced with stress on the second syllable. The intensity profile is therefore highest for the stressed vowel, second highest for the unstressed vowel, and lowest for the approximant. Again, remember that approximants have a more closed off oral cavity which reduces overall intensity, so this result is expected. The primary cue for English [ɹ] on a spectrogram is the lowering of F3 below 2500 Hz.

A green line indicating 2500 Hz is shown in Figure 2.6 as well to highlight the giant decrease in F3. Astute readers will also notice that F2 drops significantly as well, but this is only obvious in the middle spectrogram because [i] is the only vowel of the three that has a high F2 to begin with.

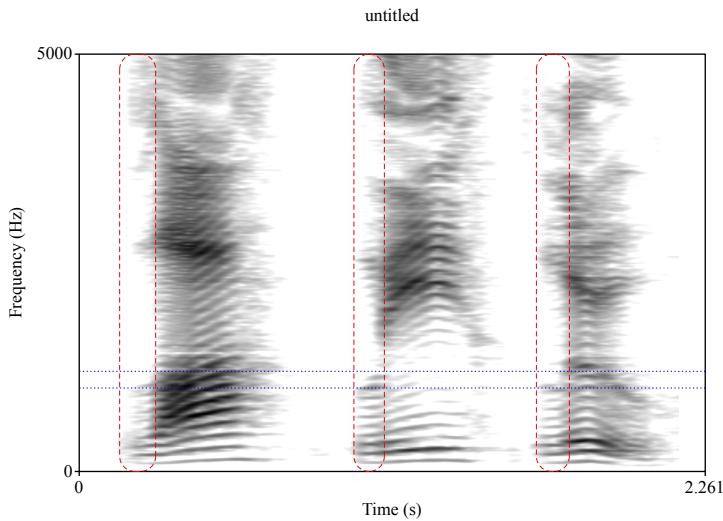


Figure 2.6: Spectrograms for [la] (left), [li] (center) and [lu] (right). Red dashed rectangles show the location of the onset [l] in each word. Blue lines indicate the range between 1000-1200 Hz.

The English onset [l] has a noticeable formant, usually around 1100 Hz. Figure 2.6 shows spectrograms for me pronouncing [la], [li], and [lu]. The location of the onset [l] is circled in red and the blue lines indicate a range of 1000-1200 Hz. Notice that for me, the formant shows up at the lower end around 1000 Hz. Coda [l] is less reliable to read on a spectrogram and we did not discuss it much in LIN 201. So we will leave it aside for now.

In lieu of a figure, I will just remind you that nasal stops also usually have a visible formant around 2500 Hz. The acoustic properties of nasals and laterals are even more interesting than you were lead to believe in LIN 201. The articulatory aspect of both sounds lead to similar acoustic outcomes, albeit at different locations of the frequency spectrum. When articulating both of these types of sounds, a speaker creates a **side-branch** where air can flow into, but not escape from. This creates all sorts of interesting effects that we did not discuss in LIN 201. I'm hoping we have

enough time to discuss both of these types of sounds in more detail in this course. If you’re interested, [Johnson \(2011\)](#) has a nice chapter that explains it in much more detail.

2.3 Silence

The idea of being able to “see” silence on the spectrogram is an interesting idea. Most of the recordings we work with, unless they were recorded in a special type of isolation booth, won’t ever be completely silent. Instead, silence is usually viewed as when the overall sound level is below some low threshold. On a spectrogram, this appears as a large area of whiteness. Recall that on a spectrogram, the amount of darkness in a given area corresponds to how much energy there is at those frequencies. If there is no energy at any frequency then we will get whiteness in the spectrogram. The only type of sound that completely blocks all air from escaping (thus any sound) are the oral stops. Affricates also have this property because the first half of their articulation is just an oral stop.

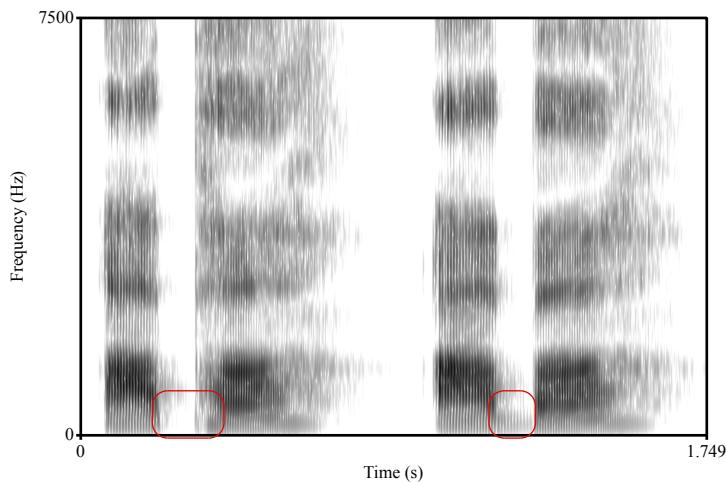


Figure 2.7: Spectrograms for [ɑ] and [ə]. Red circles indicate the lack or presence of a voicing bar.

There is one caveat of course, which is voiced stops in English between vowels are not completely “white” throughout the entire spectrogram dur-

ing the oral closure. You may remember from LIN 201 that voicing appears as a dark bar on in the low frequencies of the spectrogram. So we must take this into account when considering what is silence. Somewhat relatedly, the sound of your vocal fold vibrating emanates directly through your throat out into the world. This is how it is able to show up on the spectrogram even when no other sound is able to escape. Figure 2.7 shows spectrograms for [apa] and [aba]. The red circles show the lack and presence of a voicing bar. Notice that in both cases the frequencies above this area are all white, indicating silence.

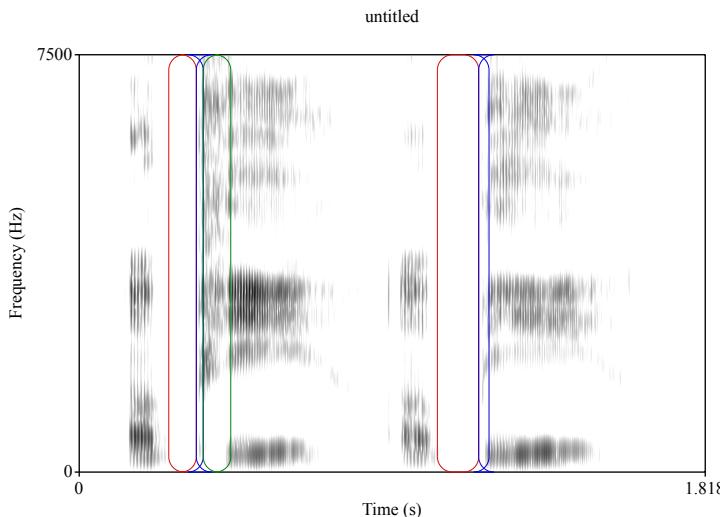


Figure 2.8: Spectrograms for [ə'p^hi] and [ə'bi]. Red circles indicate silence, blue circles indicate release burst, green circle indicates aspiration.

We end this section with a reminder that oral stops can be broken down into two parts: a closure and a release. The closure is the part where there is silence. There is always a release burst, but at the beginning of words there is variation between voiceless and voiced stops. Voiceless stops in English are aspirated, which means there is an additional period of noisy voiceless that appears on the spectrogram. Figure 2.8 shows the phrases *a pea* and *a bee*. The closure, release burst, and aspiration are circled in red, blue, and green. Notice, there is no voicing bar for the “voiced” [b]. You may recall from LIN 201 that word initial stops contrast in a property called **voice onset time** or VOT. We will do a deep dive into VOT later in this course.

2.4 Noise

The last acoustic information related to spectrograms we will look at is noise. Noise appears as (somewhat) equally dispersed energy throughout the spectrum. Therefore it is differentiated from silence by the fact that there is actually energy throughout the spectrum, and it is differentiated from resonance by the lack of clear formants and energy pooling around specific frequencies. Fricatives are the main sound type that has noise, but since the second half of an affricate is also like a fricative, affricates also have noise. Furthermore, the aspiration portion of voiceless stops also appears as noise on a spectrogram.

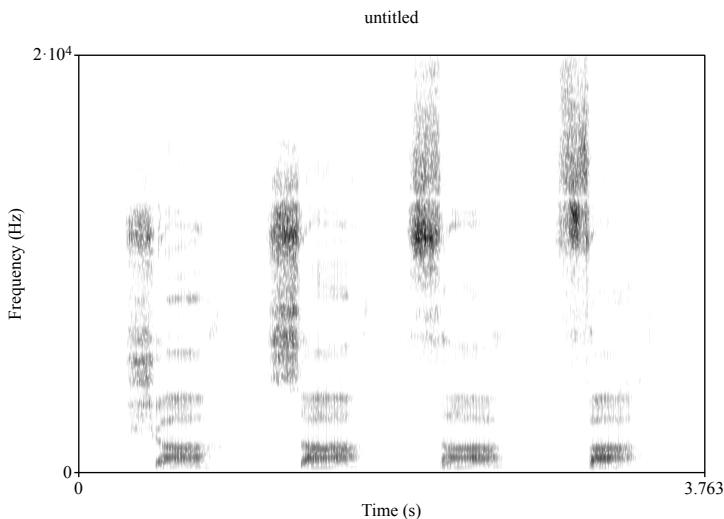


Figure 2.9: Spectrograms from left to right: [ʃa], [sa], [θa], [fa]

One thing we know about fricatives is that the location of their energy does provide cues to the place of articulation. Unlike vowel formants, we're more interested in a broad range of where the energy may peak. The location of the energy peak is directly related to how front in the mouth the constriction is made. Therefore, fricatives made more in the back of the vocal tract will have a lower peak while fricative made more in the front of the vocal tract will have a higher peak. This can be shown mathematically based on **tube models** of the vocal tract. These were covered briefly in LIN 201 and we will do a slightly deeper dive in this course. The general idea is that a constriction further back in the mouth causes there to be

a longer “tube” in front of it in the vocal tract which shapes the noise. The longer tube cause waves with longer wavelengths to be amplified and longer wavelength corresponds to a lower frequency. Hence, why the peak is lower for fricatives made further back in the mouth.

Figure 2.9 shows a spectrogram containing [ʃa], [sa], [θa], [fa]. As you should be able to say, energy is lowest on the left for [ʃa] and highest for [fa] on the right. That being said, there’s not much of a difference in the lowest peak for [θa] and [fa]. It turns out that in perception experiments these sounds often get confused. One reason for this is the small difference in “tube” sizes, but they also have much weaker energy across the board than the more strident fricatives [s] and [ʃ].

Part II

Main Lessons

Lesson 3

f0-pitch-tone-intonation

In this lesson we will discuss the way in which vocal fold vibration is used to encode linguistic information. The first part of this lesson will be a little math heavy, but stick with it. It's perfectly ok to only have an *intuition* as to what the math tells us rather than fully understanding how it all works. Providing a formalization of the underlying system is useful because it helps us understand why things are the way they are. Furthermore, it can help us identify when things deviate away from expectations.

3.1 Sine Waves

You may remember **sine waves** from trigonometry class. They are

There are certain aspects of sine waves that we are interested in for speech. The first is the **period** of the wave which is the duration of a single cycle of the wave. You can measure from any point in the wave as long as you measure to the same point in the following cycle. In figure 3.1 there is only a single cycle, but we can still measure its duration from where it crosses the origin on the x-axis. The x-axis scale is in seconds, so the **period** of this wave is 4 seconds. **Frequency** of a sine wave can be measured based on the number of times it repeats per some unit of time. The unit used in speech analysis is seconds. We measure cycles per second which is more commonly referred to as **Hertz (Hz)**. In figure 3.1, the wave has an incredibly low frequency because it has only finished $\frac{1}{4}$ of its cycle at the 1 second mark. This means the frequency is only 0.25 Hz which is

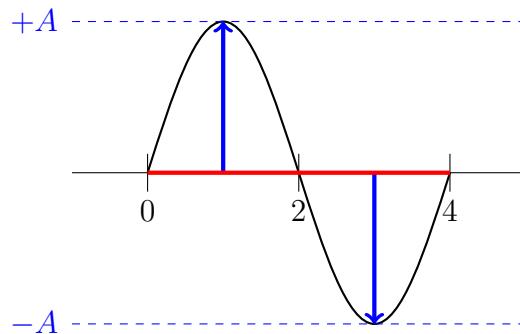


Figure 3.1: A simple sine wave with frequency $\frac{1}{4}$ Hz and amplitude A .

well below what any human listener can perceive.¹

A second aspect of sine waves we are interested in is the **amplitude**. This is how much the wave deviates from its neutral position (the x-axis). Mathematically, sine waves move equally in both directions. So the amplitude is really the absolute value of its largest displacement. This ensures that the amplitude is always positive. In figure 3.1 the amplitude is represented with blue dashed lines and blue arrows. Mathematically, the x-origin can be at any value, but for our purposes we can just assume it's 0 and the amplitude is some value A higher or lower than 0.

One important thing to note is that **frequency** and **amplitude** are not related. That is, you can change the values independently without affecting the other. Figure 3.2 displays this property. The leftmost wave has the same frequency as the wave in figure 3.1 but twice the amplitude. The center wave has the same amplitude as the wave in figure 3.1 but twice the frequency. The rightmost wave has the same frequency as the center wave but twice the frequency. A practical example is this, imagine you were singing the note corresponding to middle C on a piano (256 Hz). A low amplitude would correspond to trying to sing it very quietly while a high amplitude would correspond to trying to sing it very loudly. In both cases, the fact that its frequency is 256 Hz does not change. You can play the same game for amplitude.

Why do these aspects of sine waves matter? We know that any complex

¹Actually, it appears that this is lower than what any animal can perceive, as the [Guinness \(Book of\) World Records](#) lists the homing pigeon as the animal that can perceive the lowest frequency at 0.5 Hz

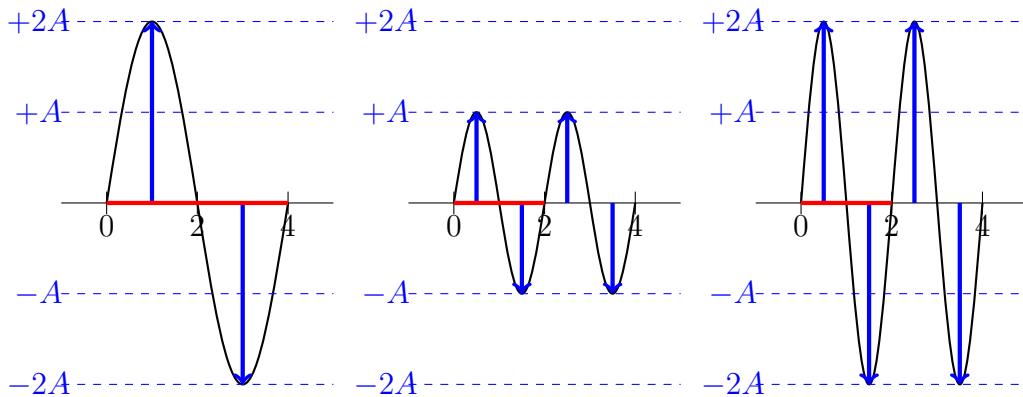


Figure 3.2: Sine

sound can be decomposed into the linear combination (sum) of simple sine waves. Pure sine waves don't really occur in nature but provide an intuitive way to talk about the complex waves that do occur. In this sense, sine waves do occur naturally, but only at a suitable level of abstraction: they are the component parts of the speech we produce and perceive every day.

There is a second aspect of this that is important for speech research. Given some assumptions, we can calculate a **fundamental frequency** (also referred to as **f0**) for a given complex wave. The fundamental frequency is the highest frequency which all the other waves share. Even if a sound is made up of a combination of many different frequencies, we will perceive it as the fundamental frequency. Furthermore, in speech research we can relate the fundamental frequency to the rate at which a speaker's vocal folds are vibrating while speaking a given utterance.

Before we completely dive into the speech aspect of fundamental frequency, let's try to get some intuition as to how amplitude and frequency interact when decomposing (and/or constructing) complex waves. Figures 3.3, 3.4, and 3.5 all contain sine waves with frequencies of 1 Hz, 2 Hz, 3 Hz, and 4 Hz in the first four rows. The fifth row is the combination of these waves. This is done by taking the value on the y-axis (amplitude) at every matching point along the x-axis and summing them all together. The outcome is a combination of each of the four sine waves.

In figure 3.3, each of the simple sine waves has the same amplitude, resulting in a complex wave that has equal properties of each wave. In

figure 3.4, the amplitude of the sine waves decrease as the frequency increases. This results in the complex wave having properties more like the waves with lower frequency. For figure 3.5 it is the reverse: the amplitude of the sine wave increases as the frequency increases, resulting in a complex waves with more of the properties of the waves with higher frequencies. In all cases, the fundamental frequency is the same: 1 Hz. This is because it is the greatest shared frequency of the four waves. You should be able to see in the figures how the wave repeats at this frequency even though it's not a perfect sine wave.

3.2 Fundamental Frequency in Speech

Given we are now somewhat familiar with what fundamental frequency (I'll use f0 from here on out) is at an abstract level, let's discuss how it is used in speech to encode linguistic information. To start, let's compare some terminology. When we use the term f0 we are giving a numeric description of the signal. We might say something like, "the f0 at the midpoint of the vowel is 112 Hz," or "the highest f0 in this sentence is 200 Hz and occurs during the final word." This is a **quantitative** description. The corresponding **qualitative** description for this phenomenon is **pitch**. Pitch relates to how speakers perceive the f0 of a given utterance. Here we might say something like, "that sounded high pitched," or "can you say that again with a lower pitch." This is a non-numeric description and therefore is always relative in some sense.

The next term we can talk about is **tone**. Tone is also measured based on its f0 value. Similarly to pitch, there are high tones and low tones (and all sorts of in betweens), but unlike pitch it is more than just the relative frequency of a given utterance. Tone specifically refers to implementational targets. We won't get into specifics now because the next lesson will dive deeper into tone specifically, but it's important to recognize that tone is another term used to describe f0, but in a very specific way. Relatedly there is **intonation**, which can be thought of as the use of f0 over a larger domain than a single word/syllable. Of course, this isn't a perfect explanation, but it's a good enough starting point. Intonation can be thought of as the melody over an entire phrase. You may remember from LIN 201 that we used values like M(id), H(igh), and L(ow) to describe intonational patterns. In the rest of this lesson we'll review the basics of

intonation and then look a little bit more at the acoustic and articulatory aspects of f0.

3.3 Intonation

Intonation is the the tune or melody of an utterance and it can signal beliefs and expectations, turn taking, among other things. In LIN 201 you may remember learning about two basic intonational melodies used in American English: the **Declarative contour** (MHL) and the **yes/no question contour** (MH). These intonational melodies were defined by the placement of the various pitches at different targets within intonational phrases.

The declarative contour places a high marker on the strongest phrasal stress, the low marker immediately after that, and the mid marker at the beginning of the utterance. The contour starts mid, then jumps up to high once it reaches the marker, then jumps down to low immediately after and extends to the end of the sentence as needed. Figure 3.6 shows the pitch contour for the sentence “I won a monkey”. You can see that at the beginning of the utterance the pitch value is in the middle region before raising up to its highest point during the first syllable of “monkey”, and then dropping to its lowest value for the second syllable of “monkey”.

The yes/no contour places a high marker at the very end of the utterance and a mid marker at the very beginning of the utterance. The contour starts mid and then rises from mid to high once it reaches the strongest phrasal stress. Figure 3.7 shows the wave form and pitch contour for the questions, “did you see that monkey?” The strongest phrasal stress is on the first syllable of “monkey” and that’s exactly where we see it rise. Notice, this is the same spot where the declarative sentence rises to high as well.

There are many other possible intonation patterns that speakers can use. These are just a couple basic examples and serve as useful examples rather than exhaustive coverage of the phenomena. What I hope you takeaway from this section is a newfound understanding of what exactly is goin on in the articulatory and acoustic space when speakers vary their intonation.

In the articulatory space, speakers are varying the tenseness of their vocal folds to alter the frequency at which they vibrate. When the vocal

folds are more tense, they spring back into position more quickly, thus resulting in more cycles per second. This is why it is harder to sing higher notes: it literally puts more strain on your vocal folds.

The acoustic consequence of changing the tenseness of the vocal folds is a change in f0 or the fundamental frequency. Think now about all the different things we can signal with intonation, both linguistics and non-linguistic. We do this all by varying the rate at which our vocal folds vibrate. Throughout the rest of the course we will continue to make connections between articulations and their acoustic consequences. I hope this mini-review of intonation has provided a simple on ramp for the rest of our phonetic journey.

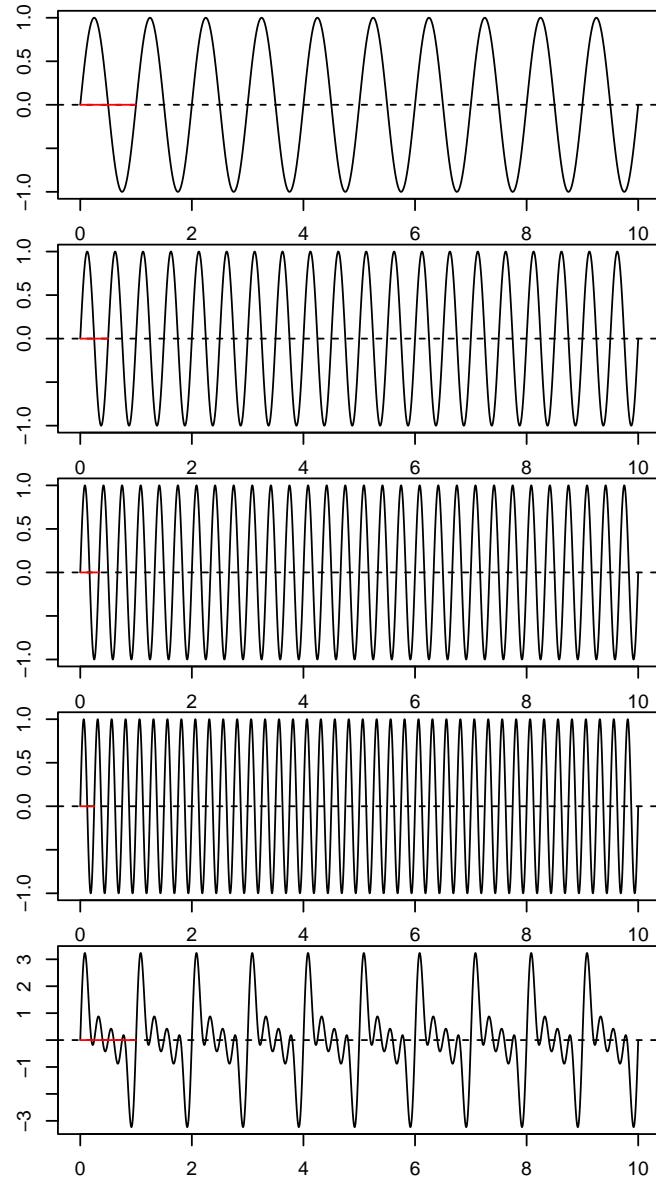


Figure 3.3: Rows 1-4 contain sine waves of increasing frequency with equal amplitude. Row 5 shows the combination of each wave which has the same frequency as the wave in row 1 but a higher amplitude. The red line indicates a single cycle of each wave.

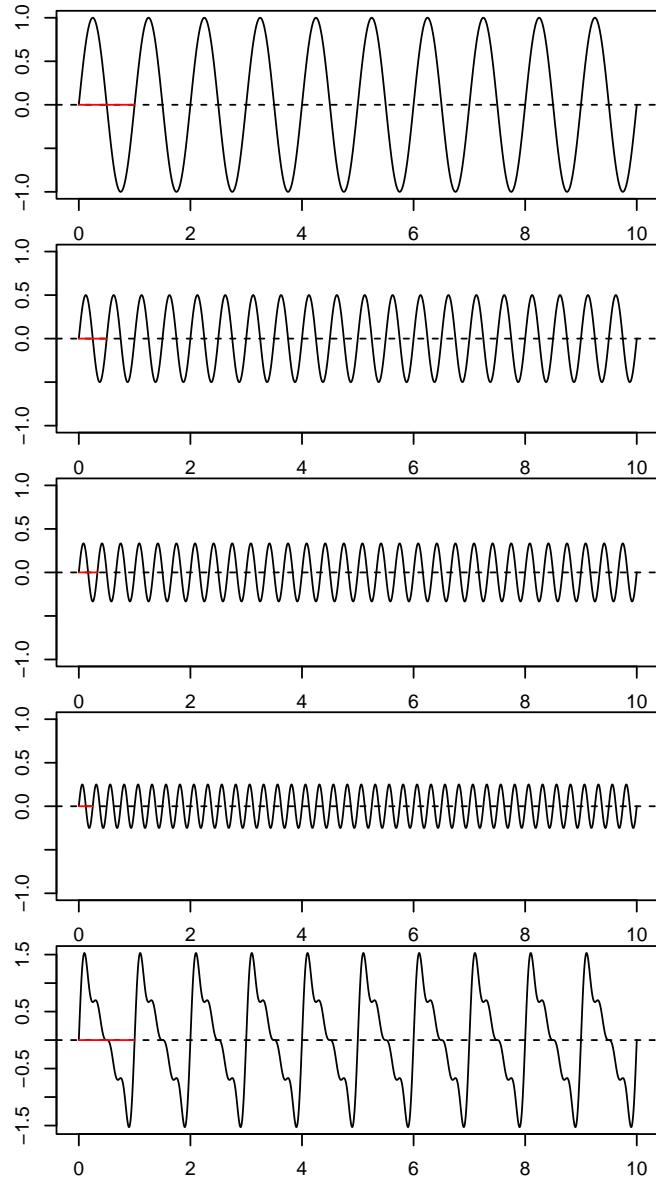


Figure 3.4: Rows 1-4 contain sine waves of increasing frequency with equal amplitude. Row 5 shows the combination of each wave which has the same frequency as the wave in row 1 but a higher amplitude. The red line indicates a single cycle of each wave.

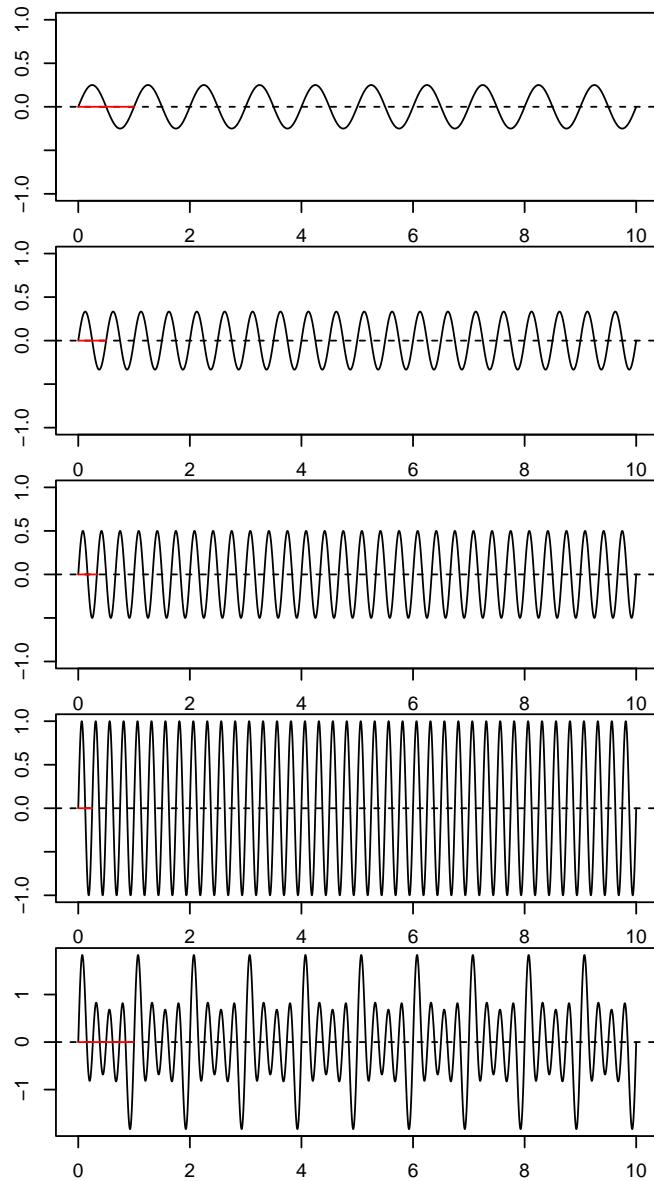


Figure 3.5: Rows 1-4 contain sine waves of increasing frequency with equal amplitude. Row 5 shows the combination of each wave which has the same frequency as the wave in row 1 but a higher amplitude. The red line indicates a single cycle of each wave.

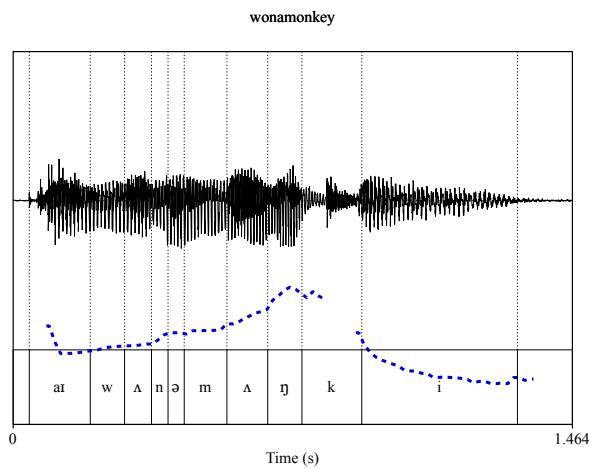


Figure 3.6: Wave form and pitch contour for the declarative phrase “I won a monkey”.

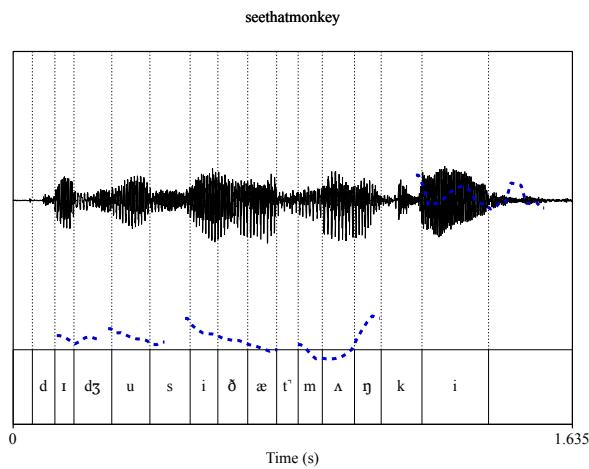


Figure 3.7: Wave form and pitch contour for the yes/no question “Did you see that monkey”.

Part III

Labs

Lesson 4

Lab 0

This lab is unlike the labs that will follow because it is strictly designed to have you practice using Praat before having to do the main labs that will count towards your grade. That does not mean this will not count towards your grade. If you fail to turn this lab in, it will affect your participation grade. You must type up your responses, but this does not have to be in the “lab report” style outlined in the syllabus.

There is an assignment called **Lab 0** on Brightspace where you will submit your work. A successful submission will include: a typed up (preferably pdf) of all questions below, 2 text grid files (*Lab0-a* and *Lab0-b*), and 1 wav file (*Lab0-c*).

Task 1

1. Download the Lab 0 files from the course website and import them into Praat.
2. *Lab0-a.{wav,TextGrid}*
 - * Intervals for each individual word have already been created for you. Don’t forget to save your TextGrid after annotating it.
 - (a) Using IPA symbols, provide a transcription on the “word” tier of the TextGrid.
 - (b) What is the duration of each word?

- (c) Why might determining the exact start point of word 4 be problematic?
- (d) Is the final stop in word 2 released or unreleased? Explain.

3. *Lab0-b.wav*

- (a) Create a TextGrid with two tiers: word and segment.
- (b) On the word tier, mark the boundaries for each word and provide an IPA transcription.
- (c) On the segment tier, try to mark the boundaries for each segment and provide an IPA transcription. Use your previous knowledge of phonetics to do the best you can.
- (d) What are the F1 and F2 values at the midpoint of each vowel?
- (e) Save the TextGrid as *Lab0-b.TextGrid*

Task 2

1. Record yourself saying 4 words that match the following patterns:
 - (a) A word that starts with a voiceless fricative and contains a low vowel.
 - (b) A word that ends with a bilabial nasal and starts with an affricate.
 - (c) A word that has a CCCV syllable structure.
 - (d) A word contains a major diphthong and a velarized lateral approximant.
2. Please list your words in your typed up pdf.
3. Save the file as *Lab0-c.wav*

Lesson 5

Lab 1

In this lab you will be measuring f0 as it is used to signify intonation in American English. For this lab, you must write it up in the “lab report” style outlined in the syllabus. There are two parts to this lab, but you can write them up as a single report.

There is an assignment called **Lab 1** on Brightspace where you will submit your work. A successful submission will include a typed up lab report (preferably pdf), and a zip file containing all the recordings you made. If you would like help with the latter portion, please contact me.

Task 1

1. Come up with 10 unique sentences that can be read as both a declarative statement and a yes/no question. For example, “We were invited to a party.” and “We were invited to a party?”.
2. Record both versions of the sentence one time each (20 total read sentences)
3. Measure the highest pitch value in each sentence. It should appear in the correct spot given the MHL and MH pitch contours we discussed in class.
4. In your lab report, include the mean and standard deviation for the H’s in the declarative sentences and the H’s in the yes/no questions.

Discuss anything you find interesting. Figures/plots are helpful as well.

Task 2

1. Come up with a sentence of the form: Pronoun - AuxVerb - Verb - Determiner - Adjective - Noun. An example is: You might buy the old couch.
2. Record yourself saying your sentence 20 times total which will be 5 times each with emphasis on the Pronoun, Verb, Adjective, and Noun.
 - (a) YOU might buy the old couch
 - (b) You might BUY the old couch
 - (c) You might buy the OLD couch
 - (d) You might buy the old COUCH
3. Measure the highest f0 value in each recording of the emphasized word.
4. In your lab report, include the mean and standard deviation for these values, summarized by emphasized word. Also include the mean and standard deviation for all of these values lumped together. Compare these values to what you measured in Task 1. Discuss anything you find interesting. Figures/plots are helpful as well

Bibliography

Gick, B., Wilson, I., and Derrick, D. (2013). *Articulatory phonetics*. John Wiley & Sons.

Johnson, K. (2011). *Acoustic and Auditory Phonetics*. John Wiley & Sons.

Appendices

Appendix A

Introduction to Praat



A.1 What is Praat?

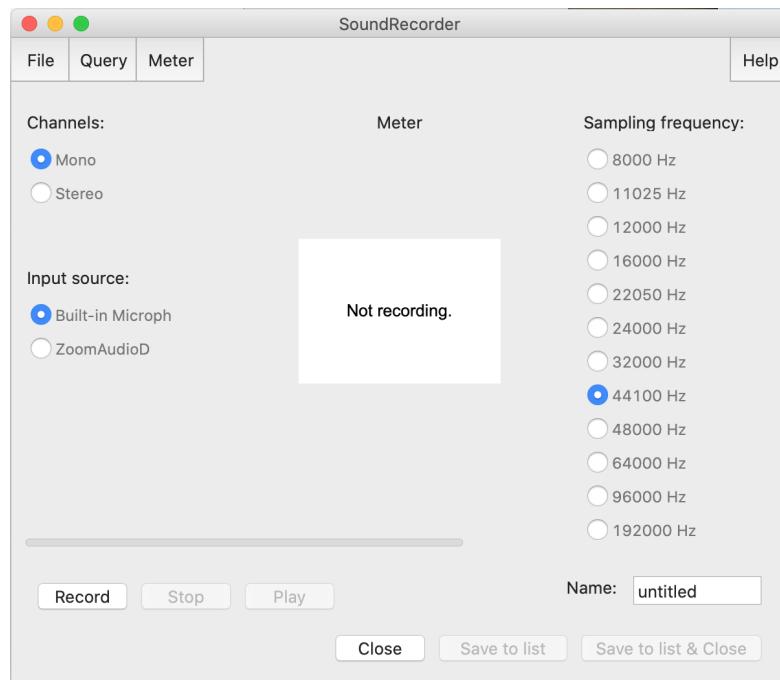
Praat (pronounced [pri:t]) is a free, open-source software created by Paul Boersma and David Weenink at the University of Amsterdam. It has many functions, but is primarily used to analyze speech. In this course, we will be using it to learn about different speech sounds and how we can analyze them manually by hand or automatically with the help of a computer. You can download Praat by going to <http://www.praat.org> and following the links for your specific platform.

***** Activity 1 *****

Download and install Praat. If you run into any problems, send me an e-mail at scott.nelson@stonybrook.edu or message me on Slack and I will help you troubleshoot.

A.2 Opening Praat/Sound Files

When you first open Praat you should have two windows pop up. One should be called PRAAT OBJECTS and the other should be PRAAT PICTURES. You can close out of PRAAT PICTURES as we won't be using that for anything that we do. In order to do anything with the PRAAT OBJECTS window, we need an object to work with. So let's create one now. At the top of the Objects window you should be able to click New → Record mono Sound.... This should cause a new window called SOUNDRECORDER to pop up.



You may have to change your Input source, but otherwise you should be ready to record at this point. To record, simply click Record and speak into whatever input source you chose. If your computer does not have a built in microphone and you are unable to record, that is ok. I will show you how to import sound files momentarily. Once you are done speaking, click Stop. You can listen back to the recording you just made by clicking Play. If you are satisfied with the recording you made, name your file in the text box next to Name: and then click Save to list & Close. If you are unsatisfied with your recording, you can click Record again. This will essentially record over your original recording and start the cycle anew.

In the Objects window you should now see an object named “1. Sound <name of sound>”. When this sound is selected, it should bring up a bunch of buttons you can click on the right side of the Objects window. We’ll ignore those for now. Instead, let’s save the recording we just made. At the top of the objects window click Save → Save as WAV file... and then save the file onto your computer. Now that we’ve saved the file, let’s see how we can import sound files. The first step will be to select the sound in the Objects window. After you’ve done this, there is a button at the bottom of the window labeled Remove. Once you clicked the remove button, the sound file should disappear. Let’s add it back to the Objects window now by going to the top and clicking Open → Read from file.... Find the file that you just saved to your computer and select it then click Open. You should see it reappear in the Objects window as “2. Sound <name of sound>”. Don’t worry about the numbering. Praat adds 1 to every new object number added to the window regardless if it has been there before. This resets every time you open the program.

If you were unable to record a sound file directly into your computer, but had other ways of recording sound (e.g. - with your phone), you can upload those into Praat this same way. I’d suggest saving it as a .WAV file when transferring from source to source as Praat can be a little picky with audio file types.

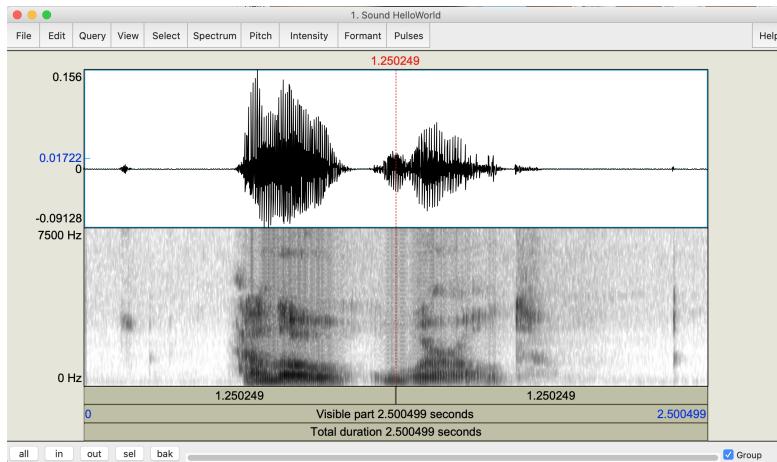
*** Activity 2 ***

Make sure you understand how to record new sounds in Praat. Record yourself saying the phrase “Hello World” and save it to your computer as HelloWorld.wav. If you recorded it within Praat, exit out of Praat, reopen it and load the saved file into your Object window. If you did not

record within Praat, open Praat now and load the saved file into the Object window.

A.3 Text Grids

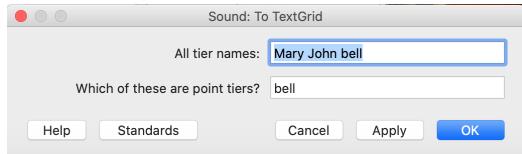
Now that we have a sound to work with, let's see what it looks like. If you highlight your sound file in the Objects window, the buttons you saw earlier should reappear. Let's take a look at what a sound object looks like in Praat. Click on View & Edit now. A window called SOUND <NAME OF SOUND> should pop up now. This is a spectrogram. It should look something like the figure below.



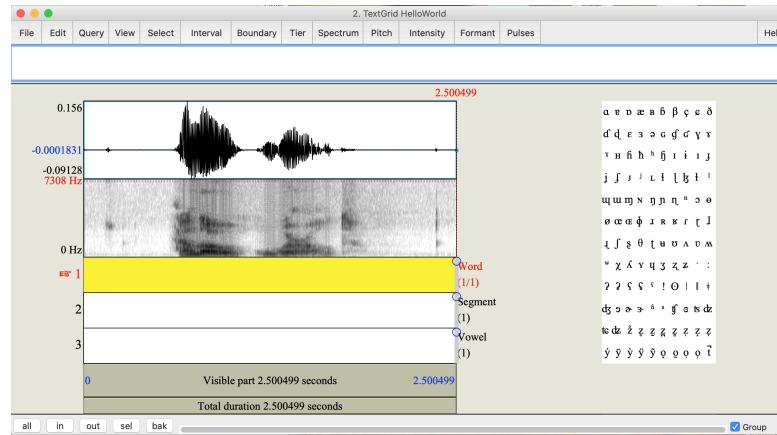
A spectrogram is a visual representation of sound. The x-axis is time, the y-axis is frequency, and the z-axis (represented by how dark a spot is) is amplitude. A higher amplitude means that there is a lot of energy at that specific point. If you click anywhere on the spectrogram, you will be presented with two values: the time at which you clicked (in seconds) and the frequency value at which you clicked. These are useful for making measurements by hand.

Another useful thing to do is divide the speech signal up into different segments. We can do this by creating what is called a TextGrid. Let's make one of those now. First you need to close out of the Sound Viewer

window. This should bring you back to the Objects window. Make sure your sound is highlighted. On the right side of the Objects window click Annotate → To TextGrid.... You should have a window called SOUND: TO TEXTGRID pop up that looks like the one below.



You can delete all the values in both cells. We will leave the cell for Which of these are point tiers? blank. In the cell for All tier names: we can supply a list of different tiers we want to create. This is done by putting spaces in between each tier name. Let's add two tiers. The first tier will be called "Word", the second tier will be called "Segment", and the third tier will be called "Vowel". Once you have done that click OK. You should know have two objects in the Objects window. Your original sound and "TextGrid <name of sound>". Highlight both in the Objects window. When you click View & Edit now your SoundViewer window should look slightly different than before.



The window now includes the three tiers we just created underneath the spectrogram, and a set of IPA symbols on the right. The IPA symbols can be used to label different segments on the tiers. Let's make a few

segments now. If you click on the first tier (Word), it should turn yellow. This tier is for segmenting the speech sound into different words. If you click on the spectrogram at the start of a word you should see a line appear on all the tiers. If you click Enter on your keyboard this will create a boundary marker on the selected tier. You can then do the same thing for the end of the word. Once you're done segmenting all the worlds, be sure to label them with English spelling or IPA. Once you have different segments created on a tier, clicking on a specific area will highlight only that interval and you will be able to enter text there.

*** Activity 3 ***

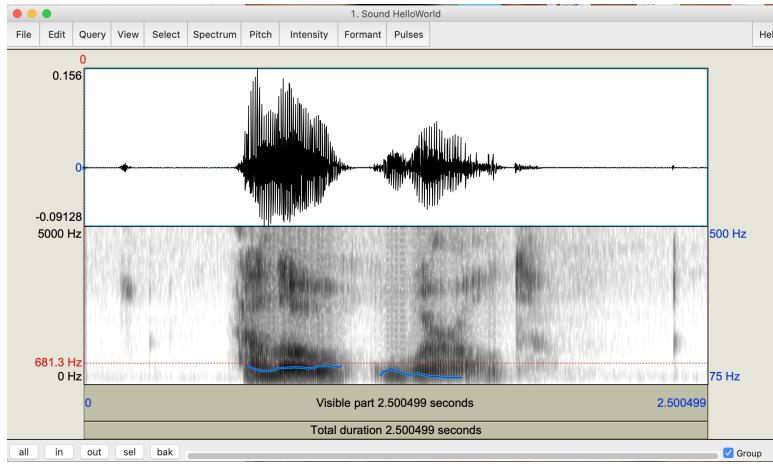
Create a TextGrid for “Sound HelloWorld” with one tier labeled “Word”. Create an interval on the word tier for each word of the phrase. Use the IPA letters on the right to label each segment with the correct IPA transcription.

A.4 Measuring Things

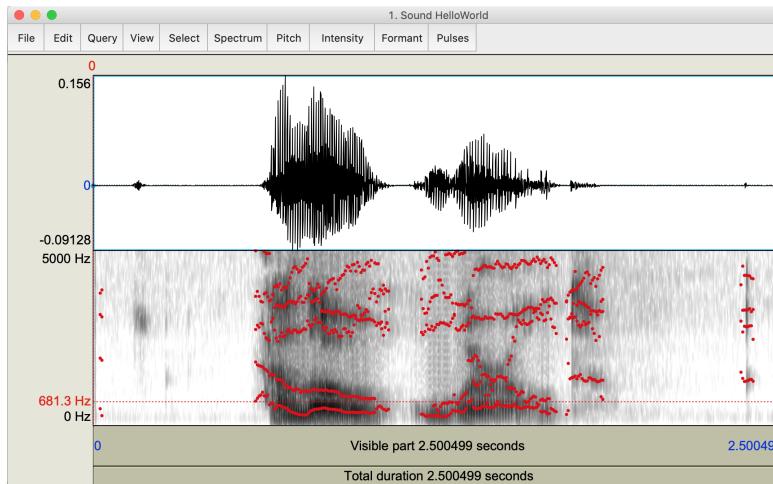
There are different parts of the speech signal that phoneticians and other speech researchers use for analysis. I will mention two of them here today: pitch and formants. Let's start by looking at this in the sound viewer. For this, we do not need the TextGrids, so you can open just the sound in the SoundViewer.

At the top of the window you can click Pitch → Show pitch. This should do two things to the spectrogram. First, it should add a blue line overtop the spectrogram. This is the pitch path. Second, it should add a blue numbers (500) to the top right side of the spectrogram. This is the range for the pitch values. Notice now when you click on the spectrogram it shows a value in red on the left, but also at the bottom right it shows a value in blue (given there is a pitch at the certain time selected). This allows us to manually get pitch values at any time we'd like.

If you click Pitch → Show pitch again, the blue line and numbers will go away. Let's show the formants now. If you click Formant → Show formants you will see a bunch of red dots appear. Unlike pitch, there are no new numbers on the right hand side as formants are really just



measurements of the normal energy distribution on the spectrogram. Even though they are dots, you should notice trend lines appearing. While there is one single pitch value, there are multiple formants within the speech signal. We are primarily interested in the bottom two which are called F1 and F2. If you click on one of the lowest of the red lines at any given point on the spectrogram, the red number on the left is the F1 value at that point. If you click on the second lowest point, then that is the F2 value.



We can also get values for pitch or formants without using the Sound-Viewer. Let's go back to the Objects window to see how to do this. The

first thing we have to do is create a Pitch object and a Formant object. To create the Pitch object, highlight the sound you want to analyze and click Analyse periodicity → To Pitch.... This will open up a window name SOUND: TO PITCH. You can just click OK as we don't need to adjust any of the settings. You should know see a new object in the Objects window called "Pitch <name of sound>".

Let's do the same thing for formants. Highlight the sound then click Analyse spectrum → To Formant (burg) This will open a window called SOUND: TO FORMANT (BURG METHOD). We can once again avoid changing any of the parameters and just click OK. You should now have four objects in the Objects window with the newest being called "Formant <name of sound>".

Now that we have the Pitch object and the Formant object, we can query values at a specified time. To do this, highlight either of these objects and click Query → Get value at time.... Once you've done this, you can specify the time (in seconds) that you would like to get a value at. If you are querying formant values, you will also have to choose which formant you would like information about. Once you enter these parameters, click OK and a new text window will pop up with the values given the entered parameters.

*** Activity 4 ***

Open up "Sound HelloWorld" and analyze its pitch. Try to measure its pitch at five different points. Record the time and pitch value for each one. Now, go back to the Objects window and create a Pitch Object for "Sound HelloWorld". Using the Pitch object, query the pitch values at the times you recorded. Do they match up with the values you got from the SoundViewer?

A.5 Other Resources

In this course we will go more in depth using Praat, but If you'd like to learn even more about it, I suggest looking at [Will Styler's manual](#).

Praat also has a built in scripting language. We may use this a little bit throughout the course. Scripting allows for automatic analysis of speech

and is very useful when conducting research. If you have a coding background and are interested in learning more about this, there is a tutorial within Praat. In the Objects window click Help → Praat Intro and then scroll down under the General section and click on Scripting. There are additional third party tutorials [here](#) and [here](#).

Appendix B

Syllabus

LIN 350 – Experimental Phonetics

Summer 2023

Instructor: Scott Nelson

Office: SBS N235

✉ scott.nelson@stonybrook.edu

Class Time: T/Th 9:30am – 12:55pm

Class Location: SBS N310

Office Hours: by appointment

Office Hours/Contact

I will be holding both in person and virtual office hours via Zoom by appointment. If you would like to talk with me outside of class, please do not hesitate to send me an e-mail so we can work out a time to meet up. You may access my “virtual office” using the meeting link provided here: <https://stonybrook.zoom.us/j/4091340647>. *Note: Zoom does not require you to download their software to join a meeting.*

If you are contacting me via e-mail, please put [LIN 350] in the subject line. I will respond within two hours from 9am-5pm on Monday-Friday. All other times I will respond sporadically, but at least within half a day.

You may also communicate with me via our class Slack channel (details below). In theory I should be able to respond to Slack posts quicker. Please

limit this type of communication to quick or general questions. Any communication about class performance or grades needs to be done via e-mail.

Materials and Resources

Brightspace - The official course Brightspace page will be used to host readings, accept assignments, and post grades. A supplemental course website will be used for everything else.

Class website - A website for the class can be found at [this link](#). Additional class material will be posted here including a course log and links to Zoom/Google Forms/etc.... Please check this page regularly as you are responsible for all information posted there for this class.

Slack - You can join the class Slack channel [here](#). Slack is a messaging app designed to be used for workplace communication. You can find an introduction on how to use it [here](#). It is also good for class communication as it allows for easier and less formal communication between myself and the students. This is good for quick questions or off topic thoughts you want to share with me or other students. It also allows for sharing files/pictures/etc... so it can be used for all different types of communication. I recommend downloading the app for either your phone or desktop.

Textbook - There is no required textbook for this class. Readings will come from a variety of sources and will be posted on the class Brightspace page.

Course Description

Introduction to common experimental methods for studying the sounds used in human language. Topics include basic speech acoustics, acoustic analysis, oral and nasal airflow, static palatography, linguography and electroglottography, as well as design of perception experiments. Students will learn the physical processes affecting each experimental variable and common methods of analyzing each kind of data. Students will get hands-on experience with each analysis method and will use two or more types

of data to explore a hypothesis about sound structure in English or some other language of interest. Students will learn how to use software for making measurements and analyzing data. Students will learn to assess the validity of claims about language based on their understanding of the scientific method as applied to speech. The course will give students a solid foundation for further courses in laboratory skills relevant to assessment of normal and disordered speech and for pursuing research, either as undergraduate researchers, or in the early stages of graduate work.

Grading/Structure of Course

Every class meeting will be broken up into three sections: **Lecture**, **Praat**, and **Lab**. The **Praat** and **Lab** portions will be interactive and will require students to bring a laptop that is able to run Praat to class. If you do not have access to a laptop, please consult with the [Laptop Loaner Program](#).

A.	Final Project	40%
B.	Reflection Essay	10%
C.	Lab Reports (8)	40%
D.	General Participation	10%

A. Final Project

Each student will be required to do an acoustic analysis on a topic of their choosing. The analysis must involve a comparison of some type between two speech communities. This can include speakers of different languages, speakers of the same language but different dialects, or disordered and non-disordered speakers. The main goal of this project is for students to apply the technical skills learned in this class to new data. A secondary goal is to provide practice for students presenting their work to others. There are three parts to the final project: a short proposal, an in-class presentation, and a written report.

You must submit a project topic proposal to me for approval by **Friday, August 4 at 5pm**. The proposal can be a short paragraph outlining the analysis you plan to perform. The goal of the proposal is to make sure

you have a plan in place. I may request changes to your plan if necessary. Each student will also be required to give an approximately 10-15 minute presentation on their topic. After each presentation, there will be a 3-5 minute question period where students will ask the presenter further questions about their project. Presentations will take place on **August 15** and **August 17**. The final report should be 3-5 pages (single spaced, 1in margin, 12pt font) and is due **Friday, August 18 at 5pm**.

B. Reflection Essay

Alongside the final draft of your project report you will turn in a reflection essay. This will be a one page document (single spaced, 1in margin, 12pt font) where you reflect on your final project and the class as a whole. It should minimally include what you consider to be the strengths and weaknesses of your final project and what you may have done differently if you had more time. Other than that, the topic of your reflection essay is open ended. Some possible directions to take it include: what you liked and didn't like about the course, how you will use what you learned in this course moving forward, or general suggestions for the structure of this course in the future.

C. Lab Reports

In weeks 2-5, each class period will end with a lab. These labs are designed to give students hands on experience measuring and analyzing acoustic data. Each lab will be completed individually and involves writing up a short lab report. The lab report is due by the start of the class period one week after the assignment of the lab (i.e. - if the lab is on a Tuesday, it is due by the start of the following Tuesday's class).

Every lab report should contain the following four sections: *Introduction, Material and methods, Results, Discussion and conclusion*. The introduction provides a brief overview of what is being measured and why it is of interest. The materials and methods sections should give an explicit description of how things were measured. This includes both descriptions of materials used (software, microphone) as well as techniques used (e.g. - measured

the vowel at its midpoint). Someone should be able to replicate your process exactly after reading this section. The results section contains the results. This should be a combination of both prose and figures. The discussion and conclusion section summarizes all of the results and discusses interesting and significant findings.

D. General Participation

Students are expected to be active participants in class. If you are unable to make a class, please try to let me know ahead of time.

Student Accessibility Support Center Statement

If you have a physical, psychological, medical, or learning disability that may impact your course work, please contact the Student Accessibility Support Center, Stony Brook Union Suite 107, (631) 632-6748, or at sasc@stonybrook.edu. They will determine with you what accommodations are necessary and appropriate. All information and documentation is confidential.

Academic Integrity Statement

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty is required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Professions, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at https://www.stonybrook.edu/commcms/academic_integrity/index.html.

Critical Incident Management

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Student Conduct and Community Standards any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures. Further information about most academic matters can be found in the Undergraduate Bulletin, the Undergraduate Class Schedule, and the Faculty-Employee Handbook.