جامعة نيويورك أبوظبي

NYU | ABU DHABI

PSYCH-UH 2218: Language Science

Class 4: The acoustic properties of speech sounds (acoustic phonetics)

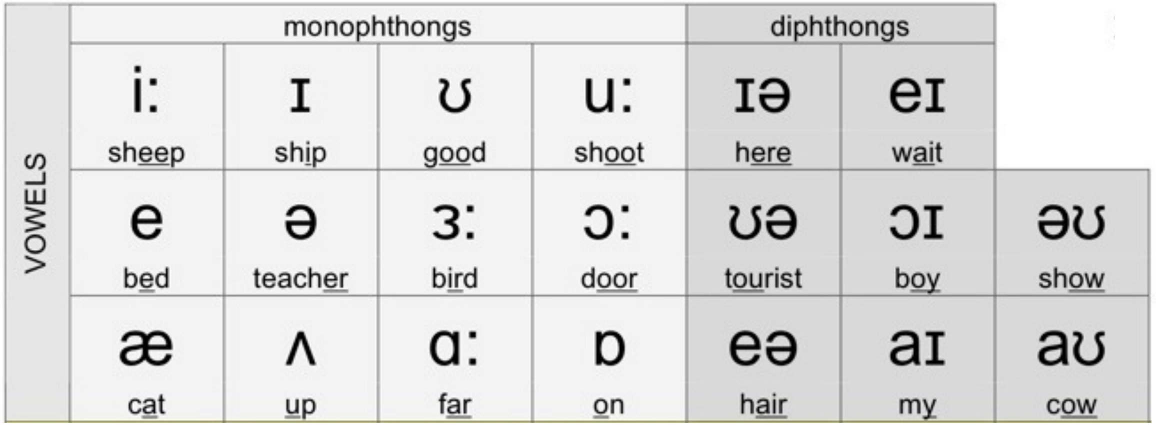
Prof. Jon Sprouse
Psychology

Acoustic Phonetics: What are the physical properties of speech sounds?

Acoustic Phonetics

Phonetics is the study of the physical properties of speech. There are at least two types of phonetics we could care about. The first looks at the physical properties of sound waves. This is called **Acoustic Phonetics**. This is what we will do today. (The second looks at the physical properties of the vocal tract, called **articulatory phonetics**. We will do that next time.)

Today we will focus on **vowels** to better understand the physical properties of segments.



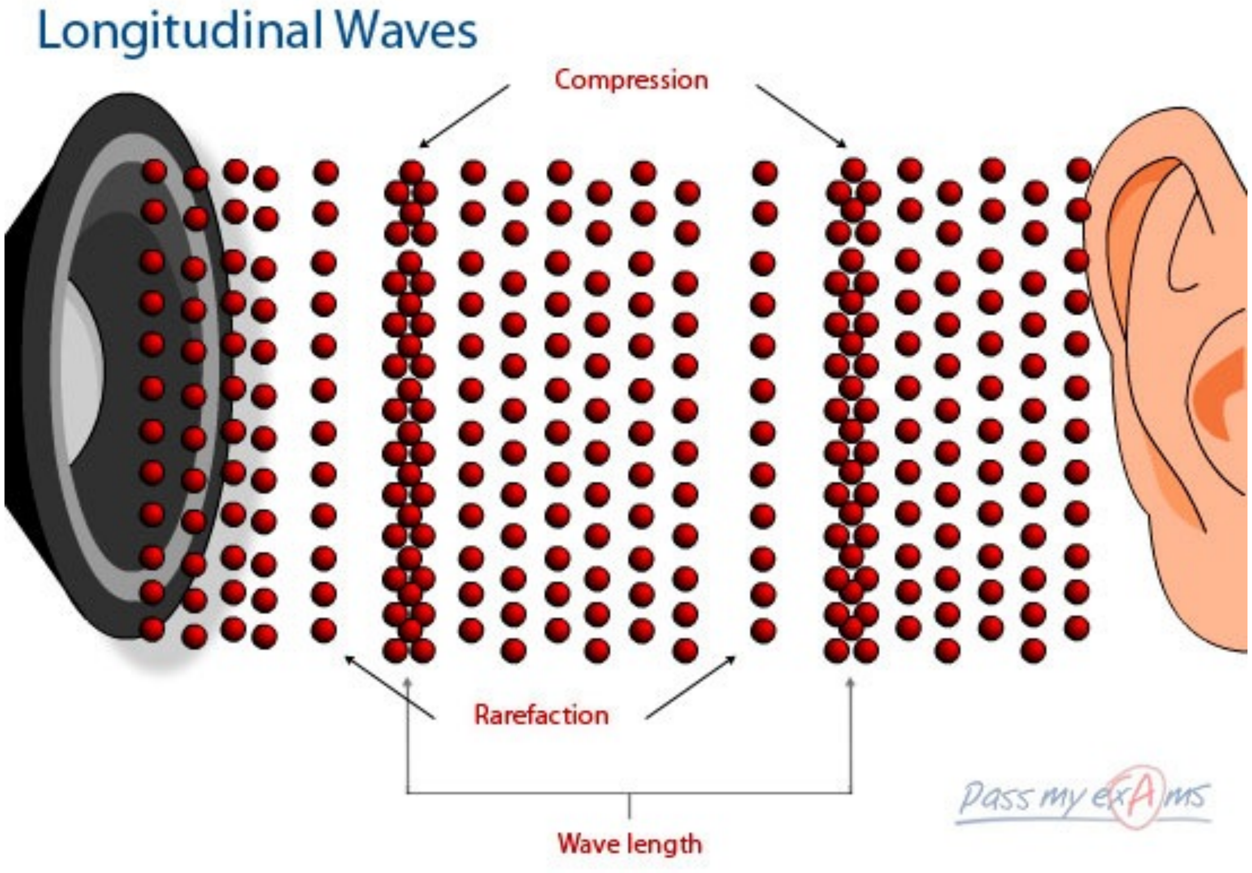
	monophthongs				diphthongs		
VOWELS	iː sheep	ɪ ship	ʊ good	uː shoot	ɪə here	eɪ wait	
	e bed	ə teacher	ɜː bird	ɔː door	ʊə tourist	ɔɪ boy	əʊ show
	æ cat	ʌ up	ɑː far	ɒ on	eə hair	aɪ my	aʊ cow

There are a number of reasons to start with vowels. But the primary reason is that we will only study acoustic phonetics for one day (today), and vowels give us a strong foundation in acoustic phonetics. (The acoustic phonetics of consonants is fairly complicated, so it would take several lectures to explore it in a satisfactory way.)

So here is our big question: What are the **physical properties** of each sound (we'll look at vowels) that makes them distinct from each other?

Sound is a distortion in air pressure

Sound is a wave that travels through air. This means that a sound wave is a disturbance in **air pressure**, or how closely packed the air molecules are.



Physics note: The compression-rarefaction cycle occurs because gas molecules try to fill the volume that they are in. When you push them together, they then try to spread apart:

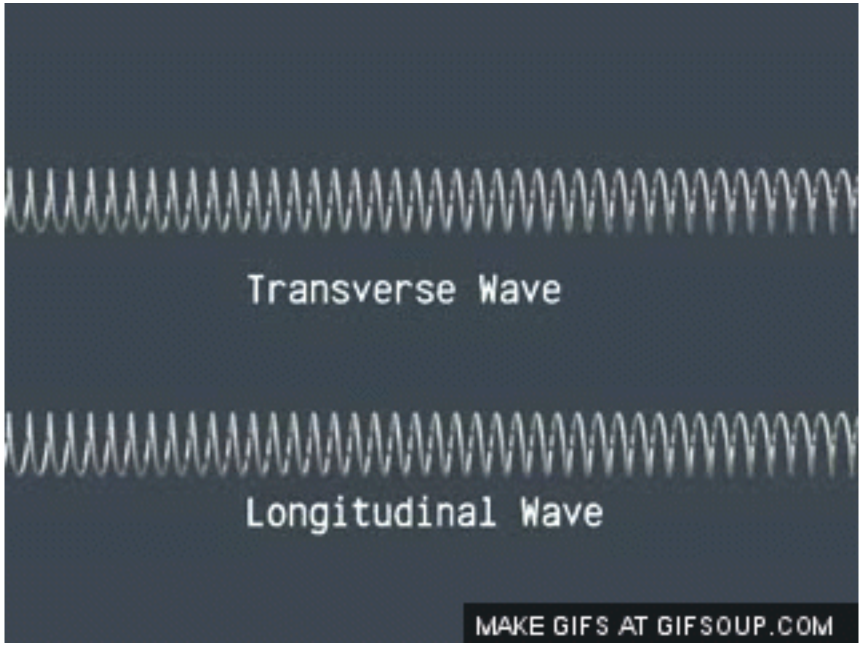
Some info on gas physics: https://www.livescience.com/53304-gases.html

Sound travels in waves

Everybody knows that sound is a wave, but what exactly does that mean? The first thing to realize is that there are two types of waves:

Waves in the ocean are transverse waves. This means the oscillation is perpendicular to the direction the disturbance is moving.

Sound waves are longitudinal waves. This means the oscillation moves in the same direction as the disturbance.



The pdf version of this slide won’t show the animation, so you can use this link to see the motion of transverse and longitudinal waves:

http://gifsoup.com/view/3529701/longitudinal-waves.html

Properties of waves

Waves have several properties. Here are two that have an impact on the way we experience sound, so you might think that they are relevant to segments:

1. **Amplitude** is a measure of the **force** applied to an area of air during compression. The perceptual effect of amplitude is a change in loudness. High amplitude sounds are perceived to be louder, low amplitude sounds are less loud.

2. **Frequency** is a measure of the **number of oscillation cycles** (for sound, oscillation is **compression**) that a wave completes in a given unit of time. The perceptual effect of frequency is a change in pitch (or tone). High frequency sounds have high pitches, low frequency sounds have low pitches.

   Physics note about frequency: The sounds that we care about here involve multiple compression cycles (not just one) because speech sounds have a fairly long duration (so they need multiple cycles).

Can you think of a way to test whether amplitude and/or frequency are important to speech sounds?

Is amplitude important to segments?

Here is a simple experiment to determine if **amplitude** is critical to the perceptual difference between speech sounds.

Step 1: say “ah”

Step 2: say “ah” with high amplitude

Step 3: say “ah” with low amplitude

Remember, amplitude is a measure of the size of the distortion -- it is the force applied to the air to cause the disturbance

Question: Did varying the amplitude result in a different sound? (e.g., ‘ee’)

Alternative experiment: say ‘ah’ and ‘ee’ with the same amplitude...

Conclusion: Varying the amplitude does not result in changes in the perception of speech sounds, only changes in **loudness**, so amplitude is not critical to the difference between speech sounds.

Is frequency important to segments?

Here is a simple experiment to determine if **frequency** is critical to the perceptual difference between speech sounds.

Step 1: say “ah”

Step 2: say “ah” with high frequency

Step 3: say “ah” with low frequency

Remember, frequency is a measure of the number of cycles the wave completes in a given time. You might know it as the pitch of the sound.

Question: Did varying the frequency result in a different sound? (e.g., ‘ee’)

Alternative experiment: say ‘ah’ and ‘ee’ with the same frequency...

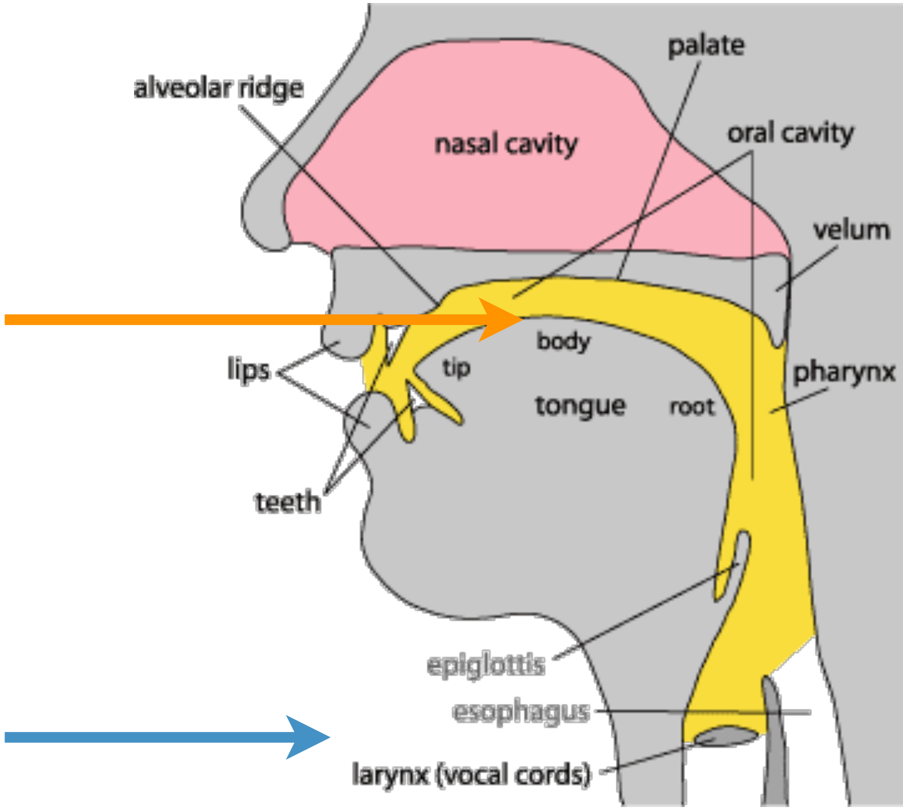
Conclusion: Varying the frequency does not result in changes in the perception of speech sounds, only changes in **pitch**, so frequency is not critical to the difference between speech sounds.

Properties of your voice

OK, so that was a bust. But it turns out that there are more complicated properties of your voice that do seem to matter for segments. But to see them, we must look closer at how your voice works.

There are two components of your voice: your vocal folds and your vocal tract.

Your vocal tract acts as a **filter** to the sound created by your vocal folds. Filters change the properties of a sound. The shape of your oral cavity and pharynx directly affect the properties of the speech sounds.



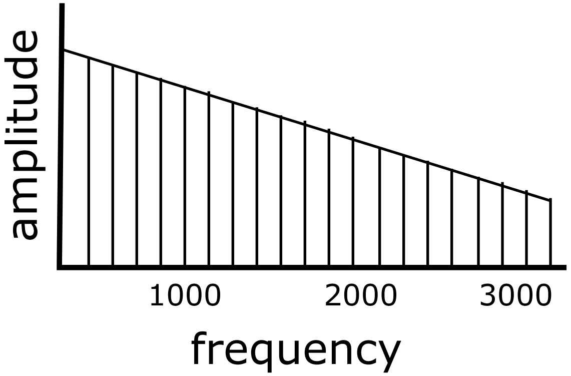
The **source** of your sound comes from your vocal folds.

Here is that video of the vocal folds of four people singing again!

http://www.youtube.com/watch?v=-XGds2GAvGQ

The frequencies of your voice

1. The first important property is that your vocal folds create a large set of frequencies simultaneously (thanks to the complex physics of vibrating objects, which we won’t go into here).



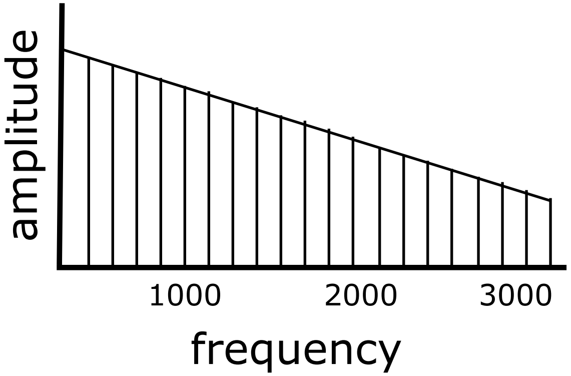
We can represent this with a graph like this: frequencies are on the x-axis and the amplitude of the frequencies is on the y-axis.

Fundamental Frequency: The lowest frequency generated by a sound source. We use the abbreviation F0 for the fundamental frequency. For your voice, this is the basic pitch that you hear.

Harmonics: The additional frequencies that are created by the source. There is one harmonic at each integer multiple of the F0. We label them H2, H3, H4, etc. Each higher harmonic is weaker in amplitude, hence the decreasing amplitude in the graph above.

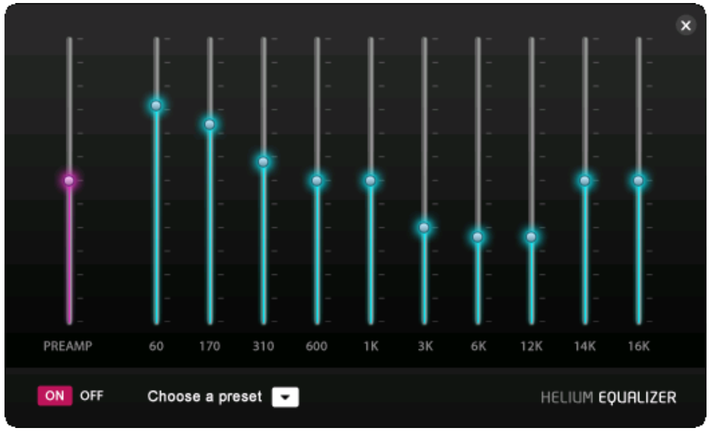
The frequencies of your voice

1. The first important property is that your vocal folds create a large set of frequencies simultaneously (thanks to the complex physics of vibrating objects, which we won’t go into here).



We can represent this with a graph like this: frequencies are on the x-axis and the amplitude of the frequencies is on the y-axis.

These graphs of frequencies might look strange, but you have seen them before in **music equalizers**. Equalizers let you set the amplitude of different frequencies, so you can get “more bass” or “more treble”. In this case, your vocal folds come with a default EQ setting like the one above.



Quick physics aside

For this class, you just need to know that your vocal folds create a fundamental frequency and a series of harmonics.

But for those of you who want to learn why your vocal folds create harmonics, you will want to look into the phenomenon of standing waves (particularly in the way that objects like strings vibrate):

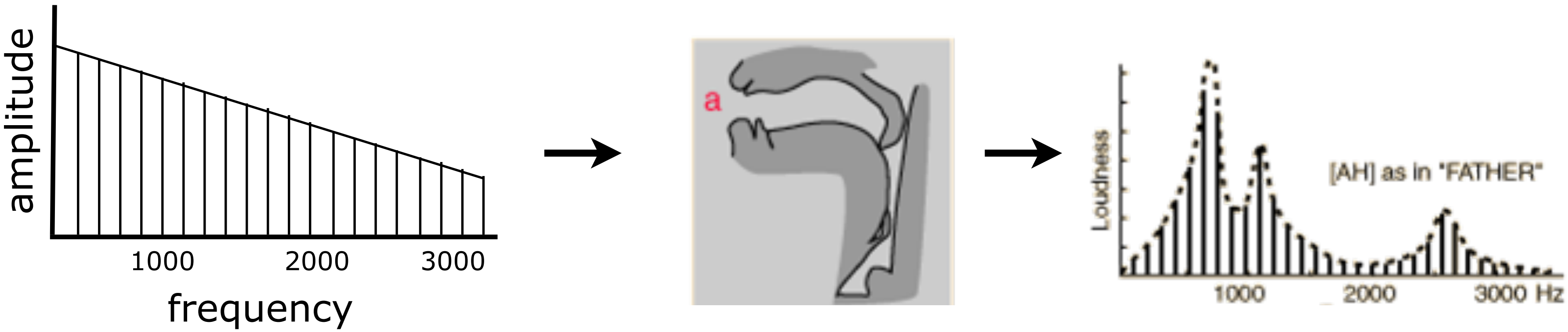
http://www.physicsclassroom.com/class/waves/Lesson-4/Harmonics-and-Patterns

http://www.physicsclassroom.com/class/sound/Lesson-4/Fundamental-Frequency-and-Harmonics

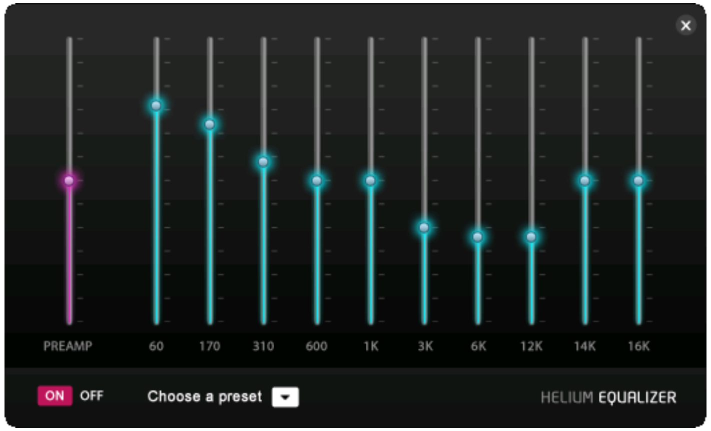
But you don't need to know that here. This is not a physics class. We just need to know that your vocal folds produce both a fundamental frequency and a series of harmonic frequencies (in decreasing amplitude) so that we can figure out how speech sounds work!

The filtering done by your vocal tract

2. The second important fact is that **the shape of your vocal tract** changes the amplitude of the frequencies created by your vocal folds. Some frequencies are increased, others decreased.



In essence, we can change the EQ settings by changing the shape of our vocal tracts. When we shape our vocal tracts to make an “ah” sound, we get a very specify EQ setting (shown above)



Quick physics aside

For this class, you just need to know that your vocal tract increases the amplitude of some frequencies, and decreases the amplitude of others. This process is called **filtering**.

But for those of you who want to learn why the shape of a cavity changes the frequencies, you will want to look into the phenomena of constructive and destructive interference (particularly for tubes, because the vocal tract is a tube!).

This link from Khan academy covers both standing waves on strings and constructive/destructive interference:

https://www.khanacademy.org/science/physics/mechanical-waves-and-sound/standing-waves/v/wave-interference-pulses

And here is the full unit on wave physics:

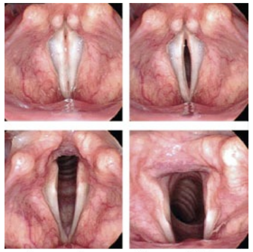
https://www.khanacademy.org/science/physics/mechanical-waves-and-sound

But you don’t need to know that here. This is not a physics class. We just need to know that your vocal tract filters the frequencies created by your vocal folds.

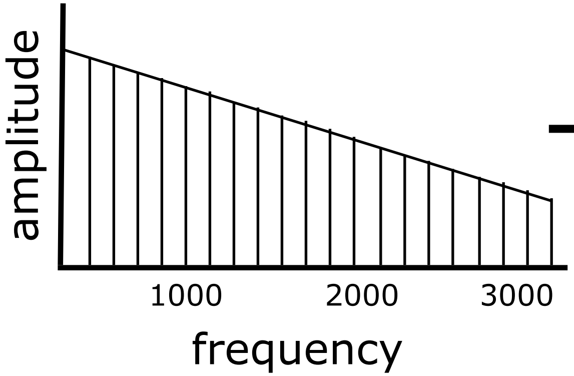
So to create an “ah”, we do this

source: vocal folds

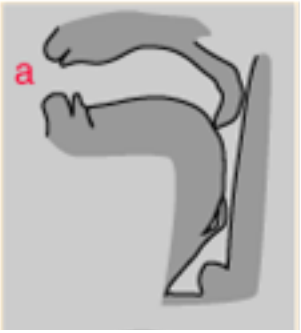
filter: vocal tract “ah”



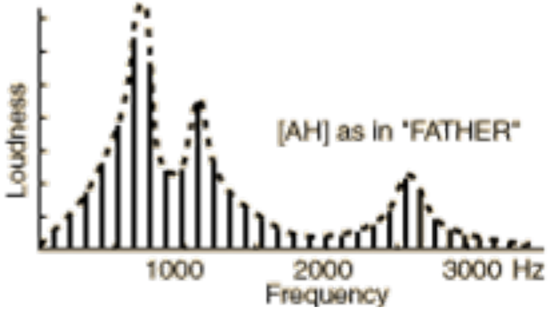
Stroboscopical imaging of the vocal fold movement using the LED laryngoscope



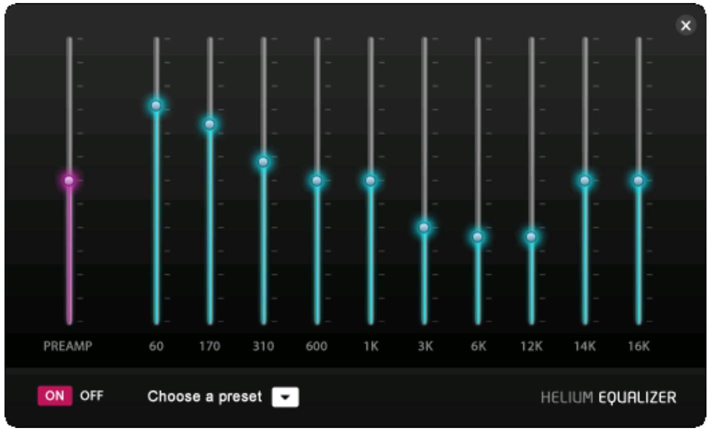
→



→

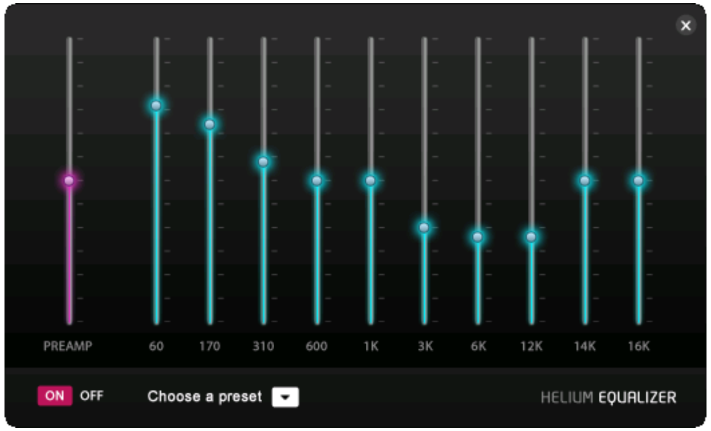


When we shape our oral cavity and pharynx to make an “ah” sound, we get a very specify EQ setting (shown above)



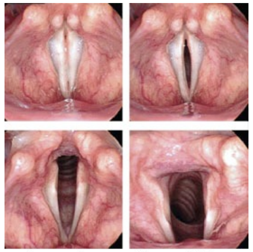
And to create an “oo”, we do this

And if we shape our oral cavity and pharynx differently to make an “oo” sound, we get a different EQ setting (shown below)!

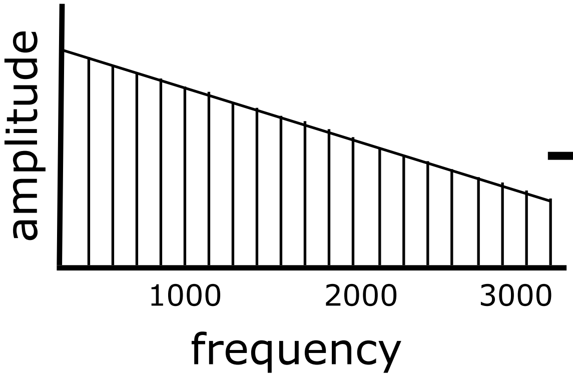


source: vocal folds

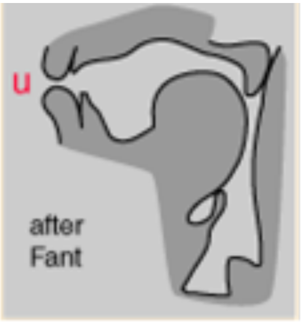
filter: vocal tract “oo”



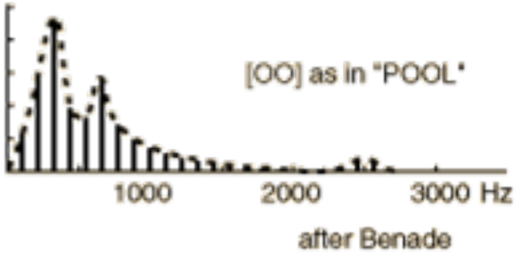
Stroboscopical imaging of the vocal fold movement using the LED laryngoscope



→



→

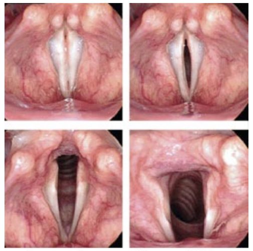


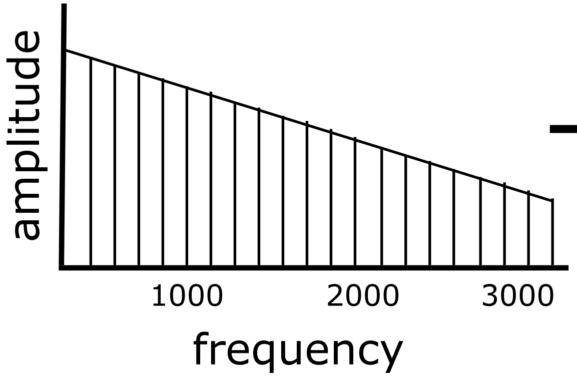
The difference between segments

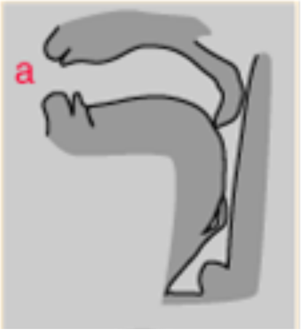
The idea is that the difference between speech sounds is a difference in the pattern of frequencies that are created by the filtering properties of the vocal tract (the different “EQ patterns”):

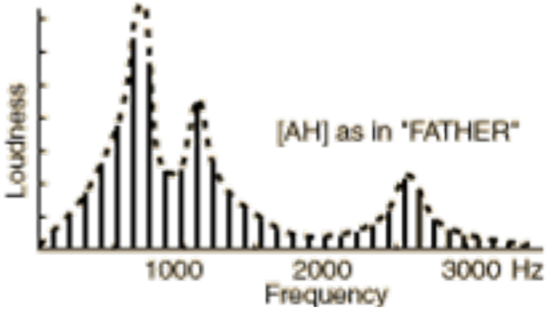
source: vocal folds

filter: vocal tract “ah”


Stroboscopical imaging of the vocal fold movement using the LED laryngoscope

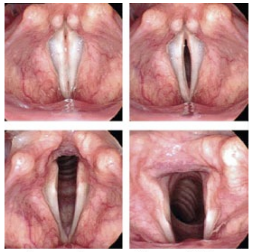


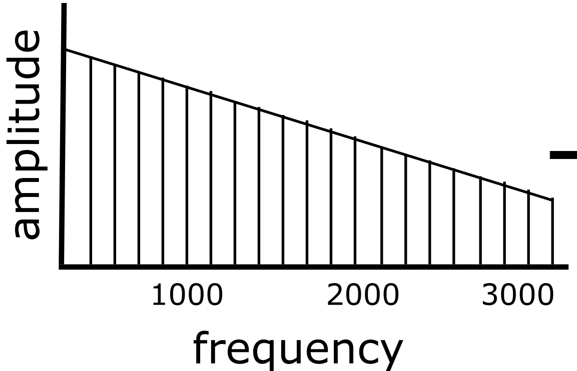


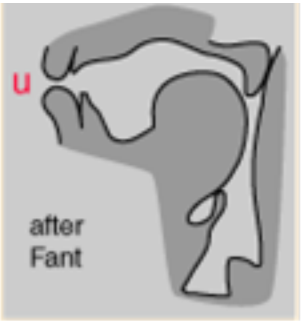


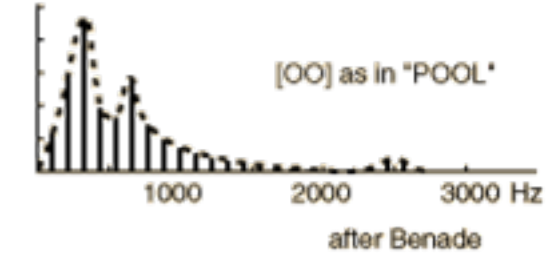
source: vocal folds

filter: vocal tract “oo”


Stroboscopical imaging of the vocal fold movement using the LED laryngoscope



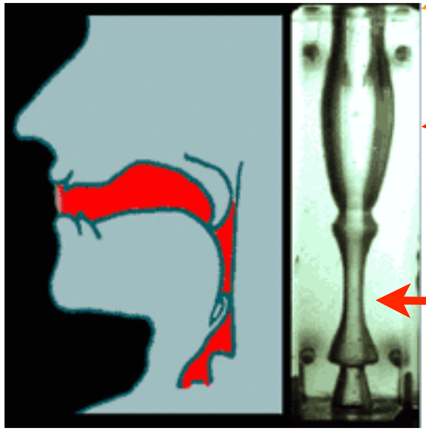




And here is a test to prove it!

In this demonstration, the frequencies created by the source stay the same each time (the same duck call). But the source is placed inside of different filters (the plastic tubes), which changes the frequency pattern. Listen to the result!

ah

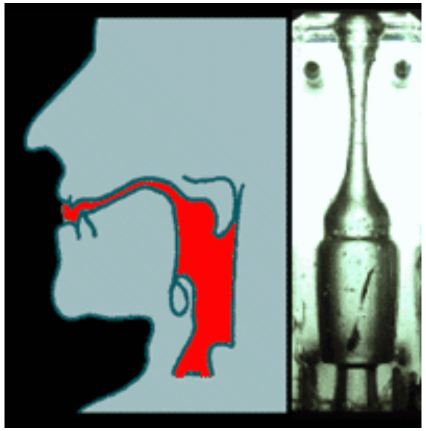


filter: plastic tubes

oral cavity

pharynx

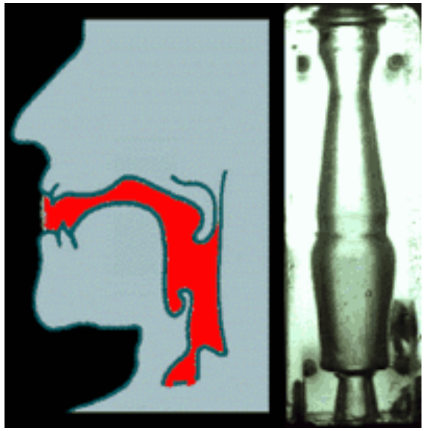
ee



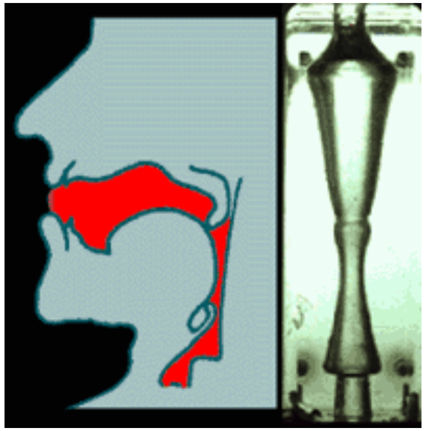
duck call

source: duck call

eh



oh

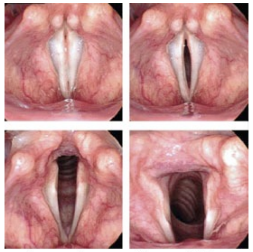


http://www.exploratorium.edu/exhibits/vocal_vowels/vocal_vowels.html

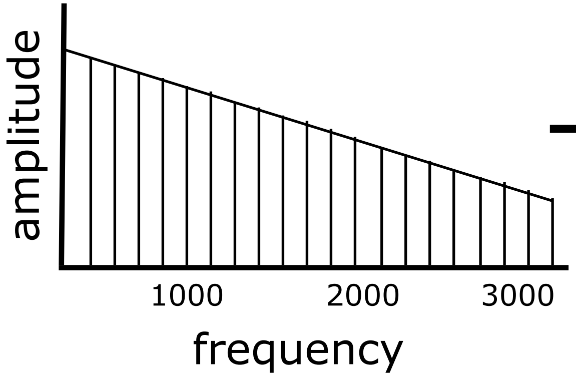
OK, let’s learn some new terms

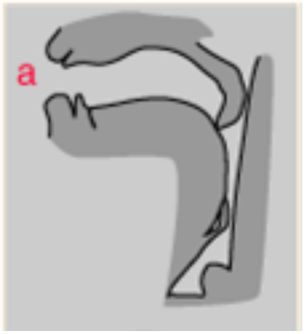
source: vocal folds

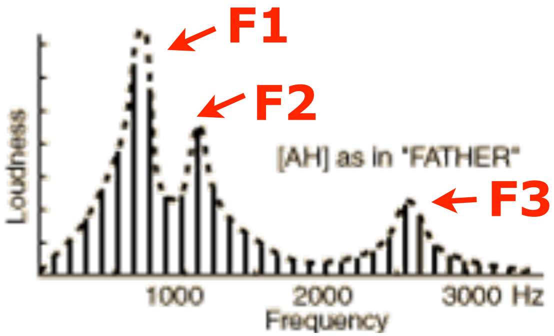
filter: vocal tract “ah”



Stroboscopical imaging of the vocal fold movement using the LED laryngoscope







For the human voice, we call the highest amplitude frequencies that occur after filtering (i.e., in the EQ settings) the **formants**.

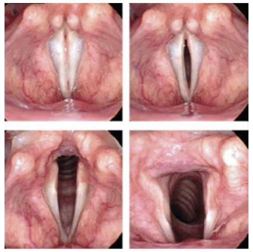
Formants are the highest amplitude peaks in the frequency spectrum created by the human vocal tract.

Much like harmonics, we label the formants in order beginning with the lowest frequency (F1, F2, etc.).

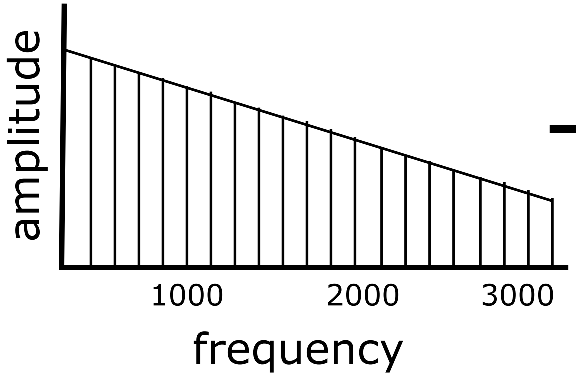
Here are the formants for two speech sounds

source: vocal folds

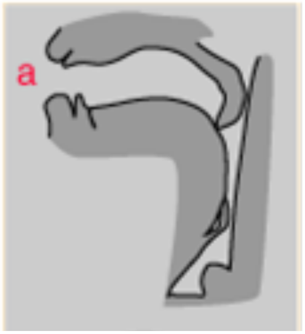
filter: vocal tract “ah”



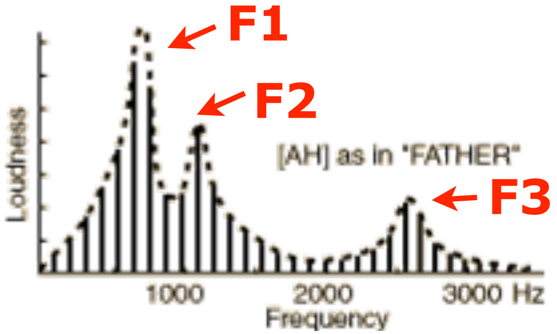
Stroboscopical imaging of the vocal fold movement using the LED laryngoscope



→

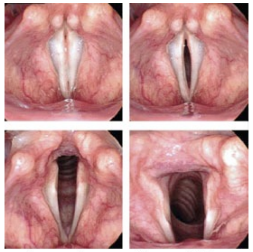


→

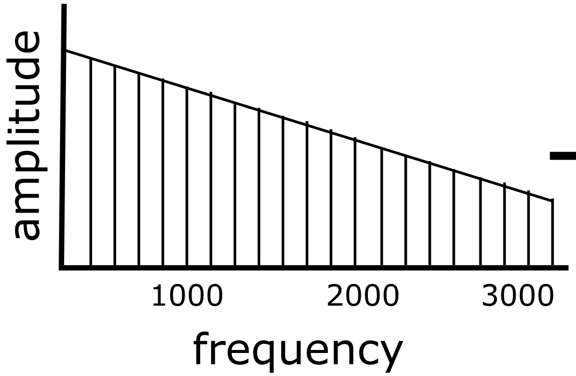


source: vocal folds

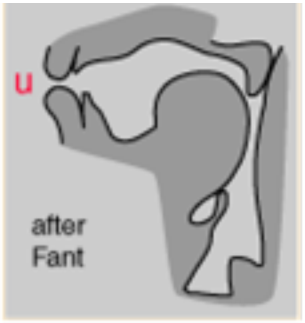
filter: vocal tract “oo”



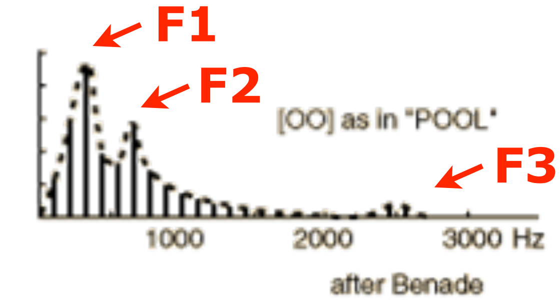
Stroboscopical imaging of the vocal fold movement using the LED laryngoscope



→



→



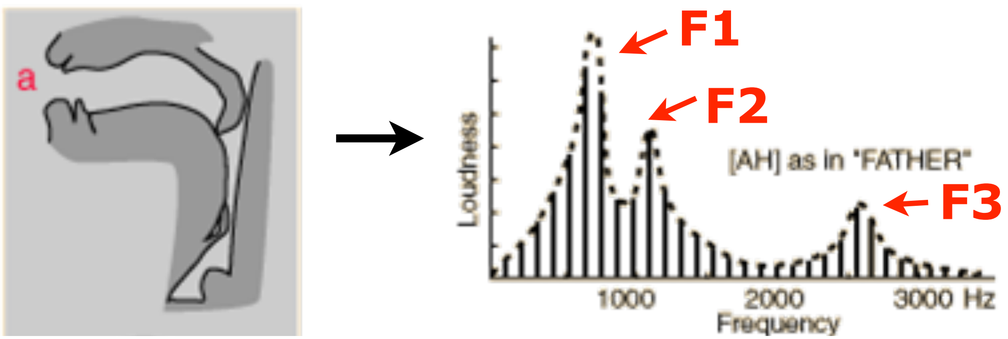
The unity of segment percepts is a great example of structure in the mind!

When you hear “ah”, you think it is a single sound. In cognitive science, we say that you **perceive it as a single percept** (where percept just means “thing that is perceived”).

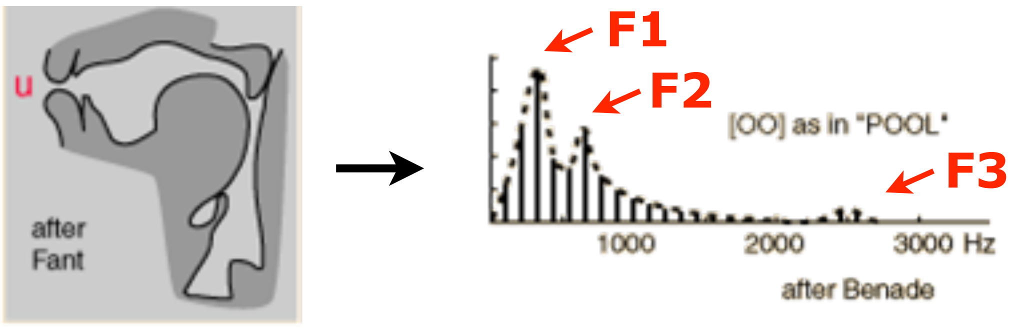
But physically, that phoneme is really **a combination of three frequencies**, the three formants, put together.

The really cool thing is that **you can’t hear the three frequencies**, no matter what you try. You just can’t. That is because the mind is structured to perceive speech as a single percept!

filter: vocal tract “ah”



filter: vocal tract “oo”



But you can hear formants in music!

Even though we can't hear formants in speech, it is possible to hear them in music. They are called overtones in music. Overtones are always present in music - they are the reason two instruments sound different!

Jaw Harps are a metal reed that you pluck in between your teeth in order to make a tone. Although the fundamental frequency never changes (the reed is always the same length), its vibrations resonate in your mouth giving rise to different formants based on the shape of your mouth:

www.youtube.com/watch?v=yx0nnZZVnd8

www.youtube.com/watch?v=VDnio2axqNI

Overtone signing (overtones are what musicians call formants) is a singing technique in which performers change the shape of their vocal tract to create different formants. In effect this creates two notes at once: the low fundamental frequency and a higher frequency formant. In many traditions, the low frequency fundamental is kept relatively constant, such that the melody is actually carried by the higher frequency formant! Listen for the high-pitched melody in these clips:

https://www.youtube.com/watch?v=qhSEKxQjOpY&ab_channel=KUULAR

And you can see that speech is just tones

This person took songs that include lyrics, and fed them into a midi-synthesizer that synthesizes piano sounds. What this means is that everything is being played on the piano, even the singing part.

The fact that speech is just a series of tones means that even after converting it to piano music, you can still (mostly) understand the lyrics (and in particular the vowels):

https://www.youtube.com/watch?v=ZY6h3pKqYI0

Speech with just formants: Sine Wave Speech

The importance of formants to speech can be seen by deconstructing normal human speech. Here is a typical English sentence deconstructed into the following parts:

1. Its first formant (F1)

2. Its second formant (F2)

3. Its third formant (F3)

4. The sounds that are not formants (pops, cracks, shushes)

OK, now let's put the first three formants together (without the pops and cracks), and see what it sounds like!

Could you get it? If not, here is the original recording of the sentence.

And here are the three formants again.

Speech with just formants: Sine Wave Speech

For those of you reading this as a pdf, here is a link to the sounds on the previous slide:

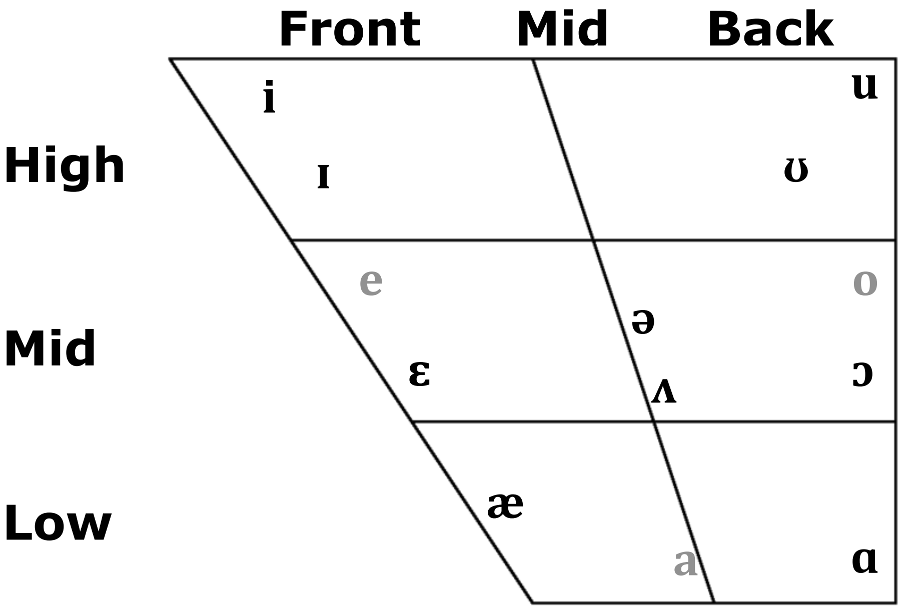
http://www-oedt.kfunigraz.ac.at/hlt/content/04LV4.../01-LADEFOGED-VOWELS-KONS-CD/vowels/chapter7/abirdinthehand.html

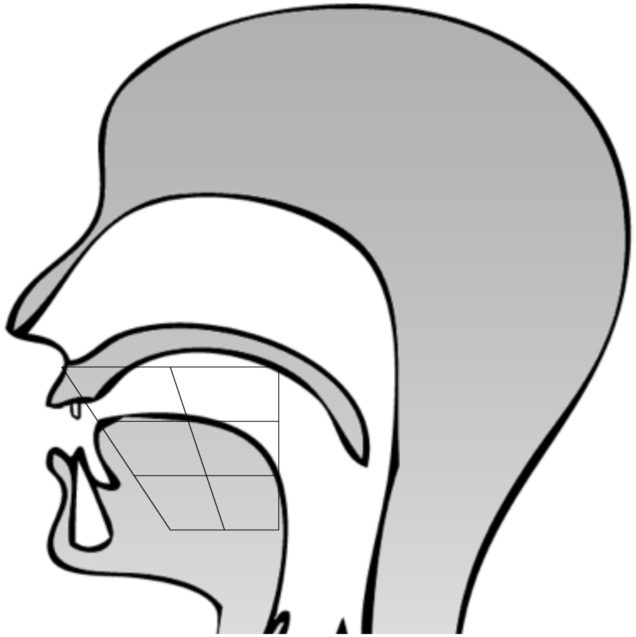
Disclaimer:

Sine wave speech isn’t perfectly intelligible. For most people, it is much easier to recognize as speech if you already know what is being said. This shows that the formants aren’t quite enough for normal speech perception (we will come back to this soon). But the fact that you can understand SWS does show that formants play a large role in segments!

The relationship between articulatory features and phonetics

Remember our friend the vowel chart?



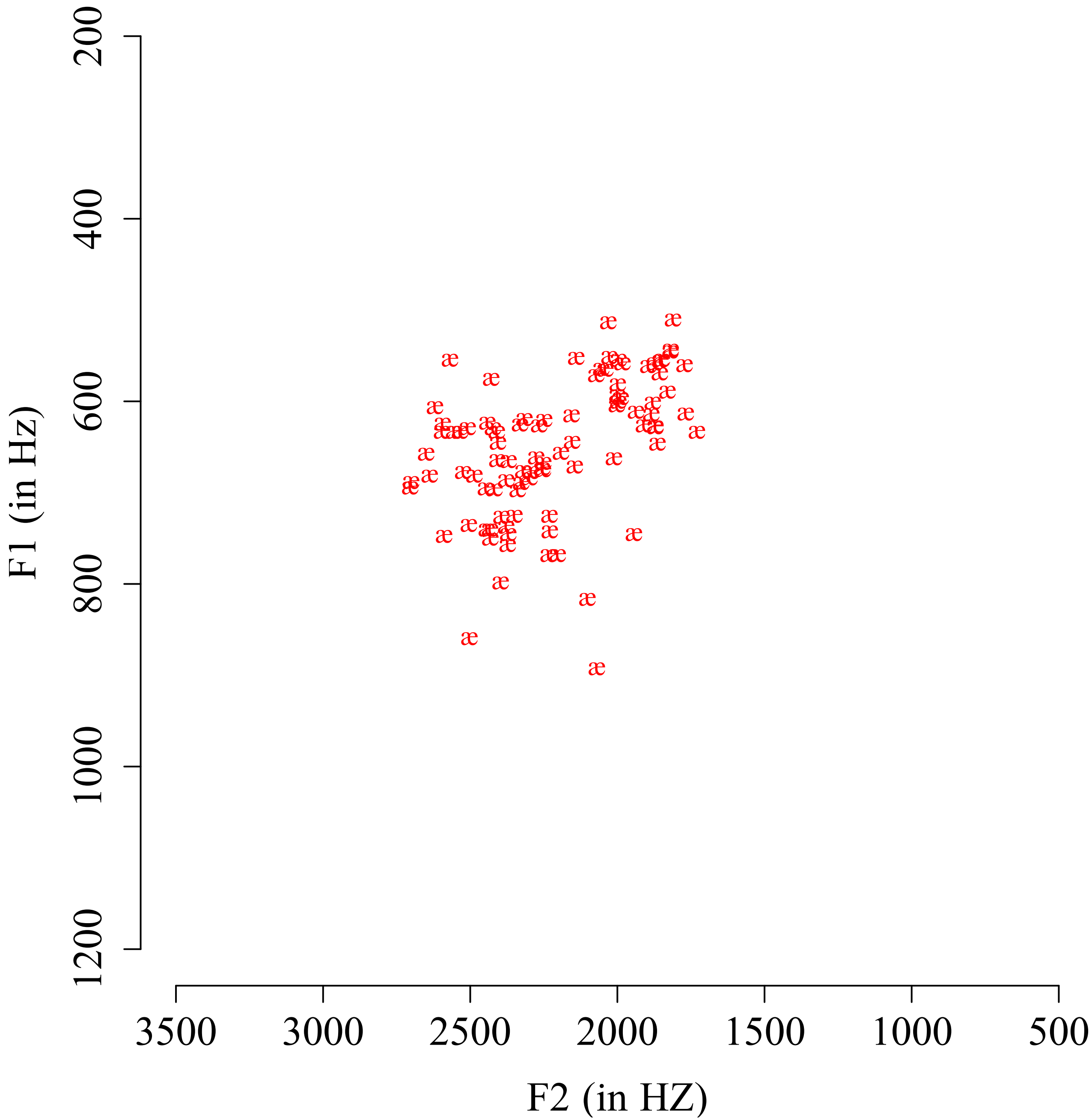


The idea behind a vowel chart is that it represents a human mouth pointing to the left. We then write the IPA symbol for a vowel in the location that the tongue would be in the mouth when that vowel is produced!

So, if the tongue would be high in the mouth, the symbol is written high in the chart; if the tongue would be front in the mouth, then the symbol is written toward the left of the chart.

Finally, the space is divided into sections so that the features can have values: front/mid/back for the backness feature, and high/mid/low for the height feature.

We can plot F1 and F2 for vowels

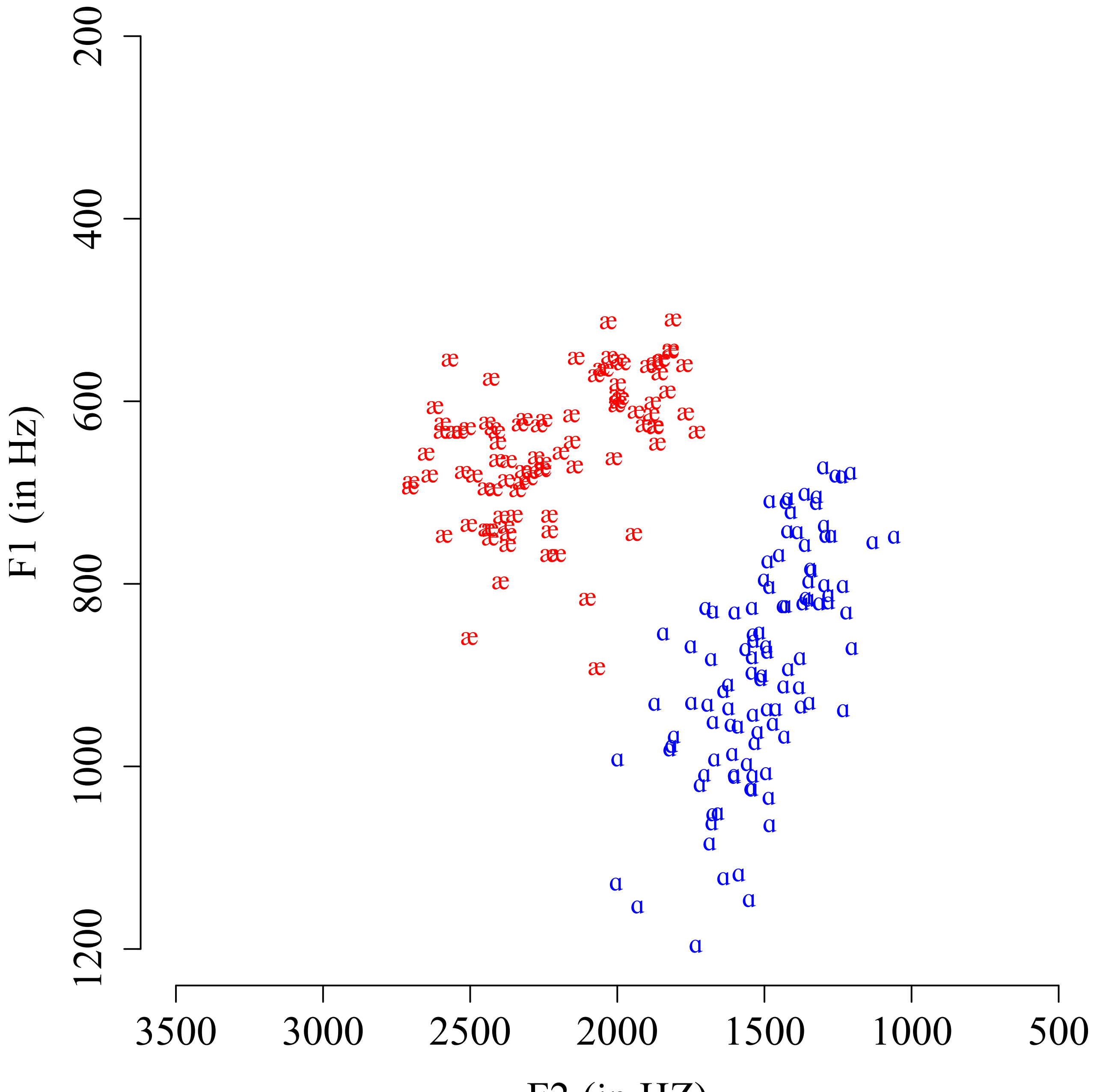


Hillenbrand 1995 recorded a large number of English speakers saying each vowel.

We can plot the **first formant** and the **second formant** of each vowel for each speaker.

As you can see, each speaker produces a slightly different æ (as in "cat"). Speaker variability is an issue to explore in speech perception, but for today, we can set it aside to see how the formants map to articulation.

We can plot F1 and F2 for vowels

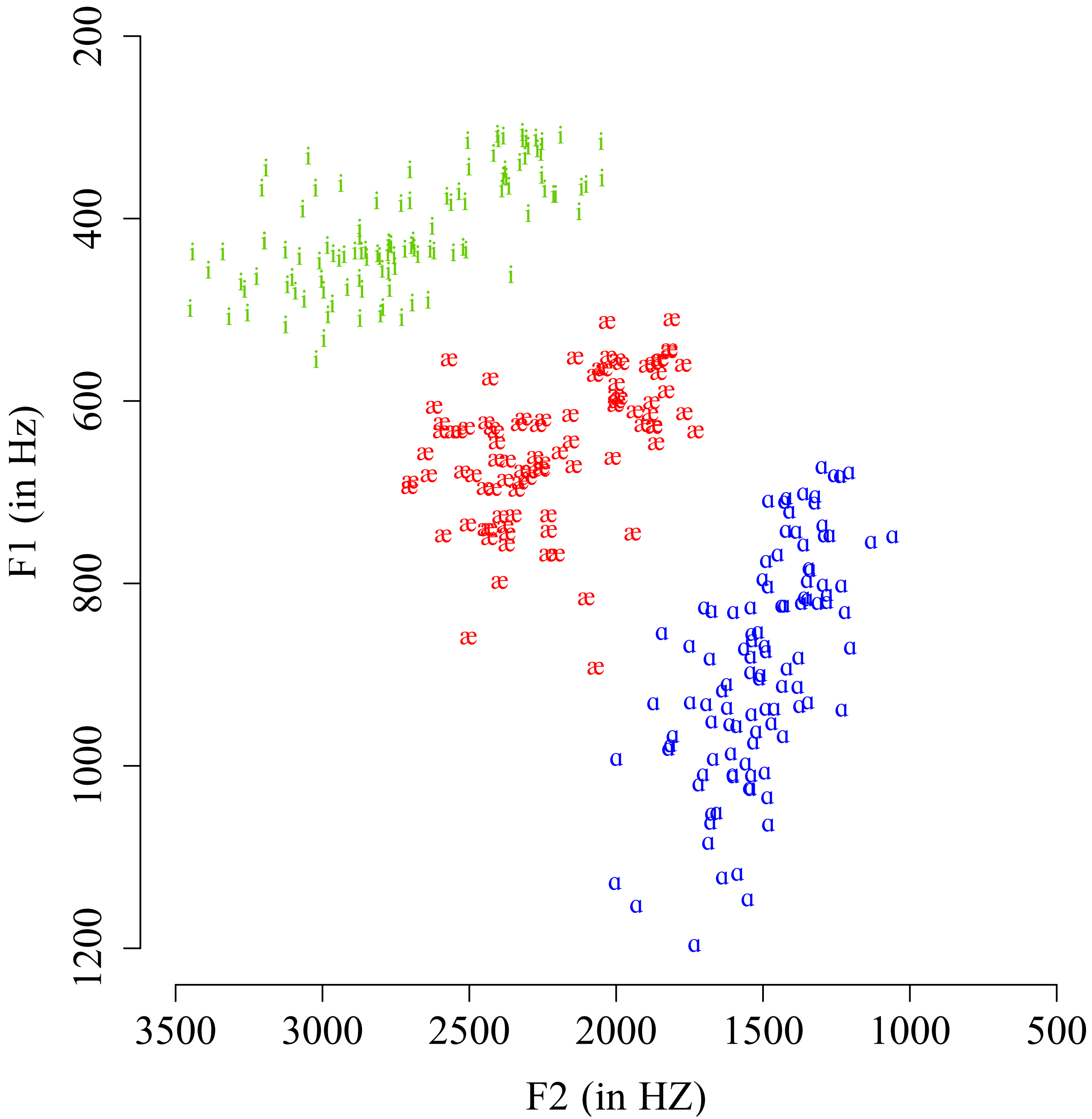


One thing to note is that I have plotted the F1 and F2 frequencies in descending order.

This is not standard for plots. We would typically plot them ascending from low to high on the y-axis and from left to right on the x-axis.

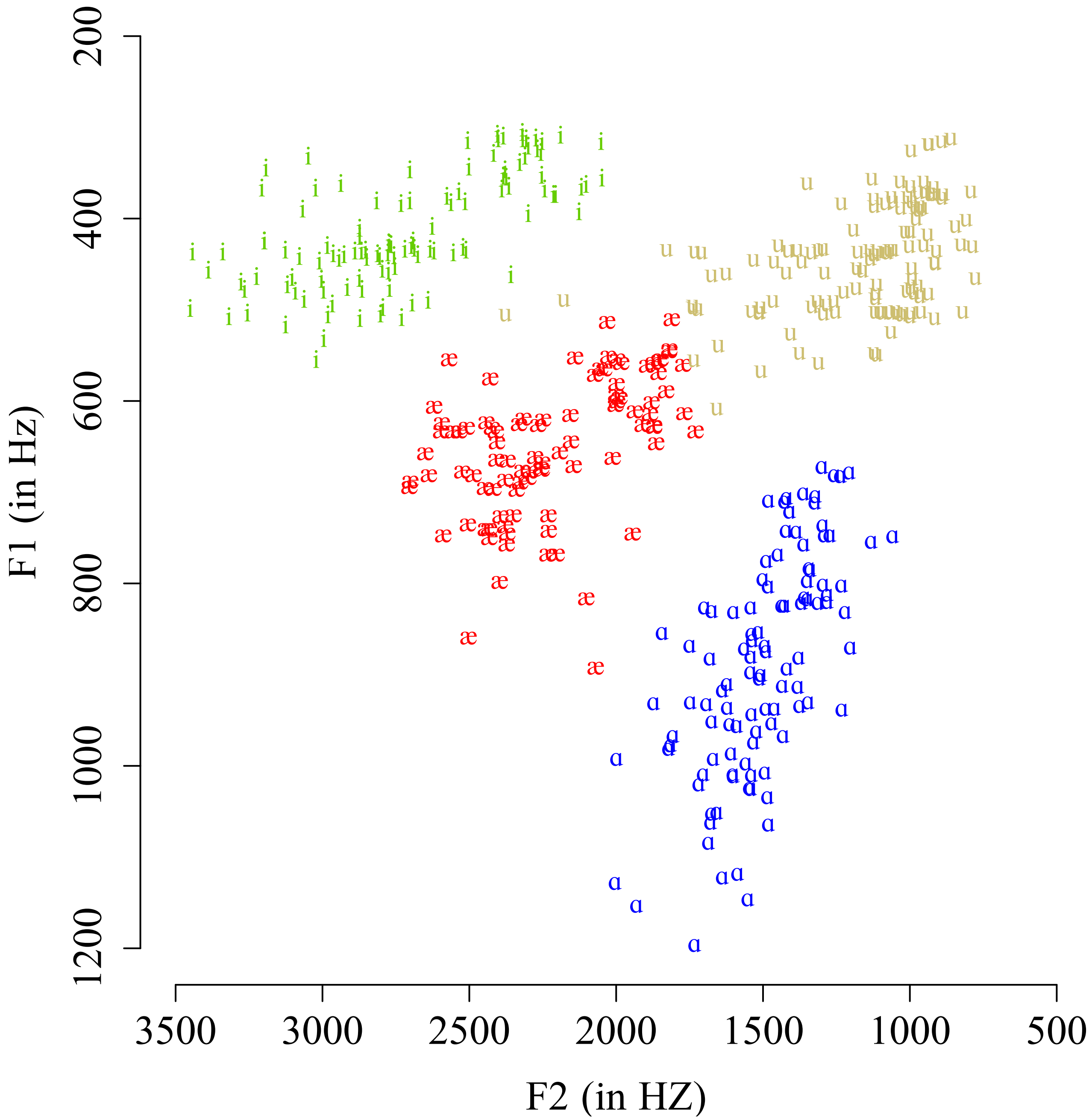
But you will see why I have done this once we add several of the vowels.

We can plot F1 and F2 for vowels



You can begin to see that each of the vowels tends to be produced in a slightly different part of the F1/F2 space!

We can plot F1 and F2 for vowels

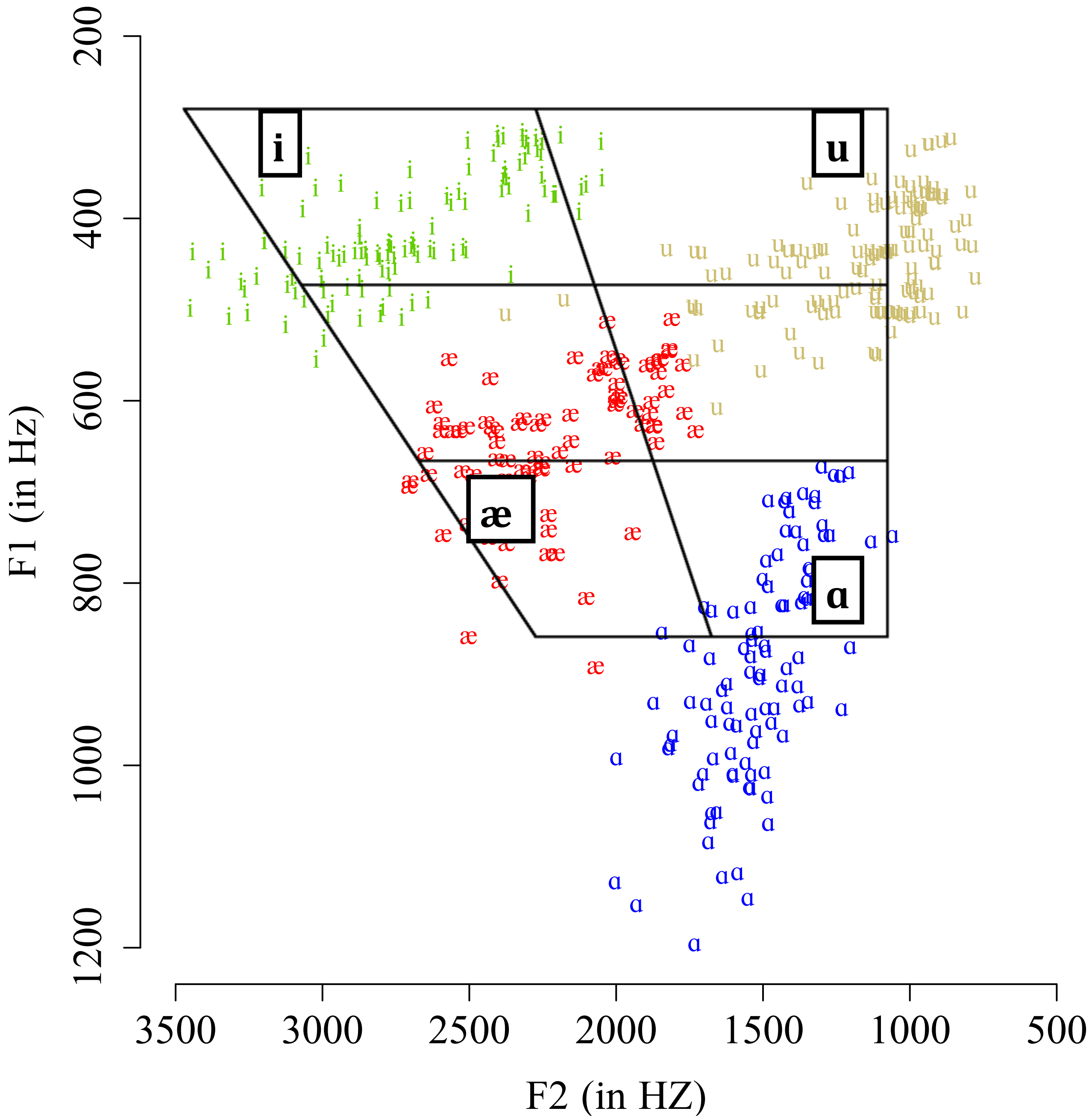


Here I have plotted 4 vowels, so that we can begin to see the locations in F1/F2 space.

I don’t want to plot all of the English vowels yet because the speaker variation leads to some overlap in location (which is an issue that is studied in speech perception).

For now, I just want to give you an idea of the relative location of these vowels in the F1/F2 space.

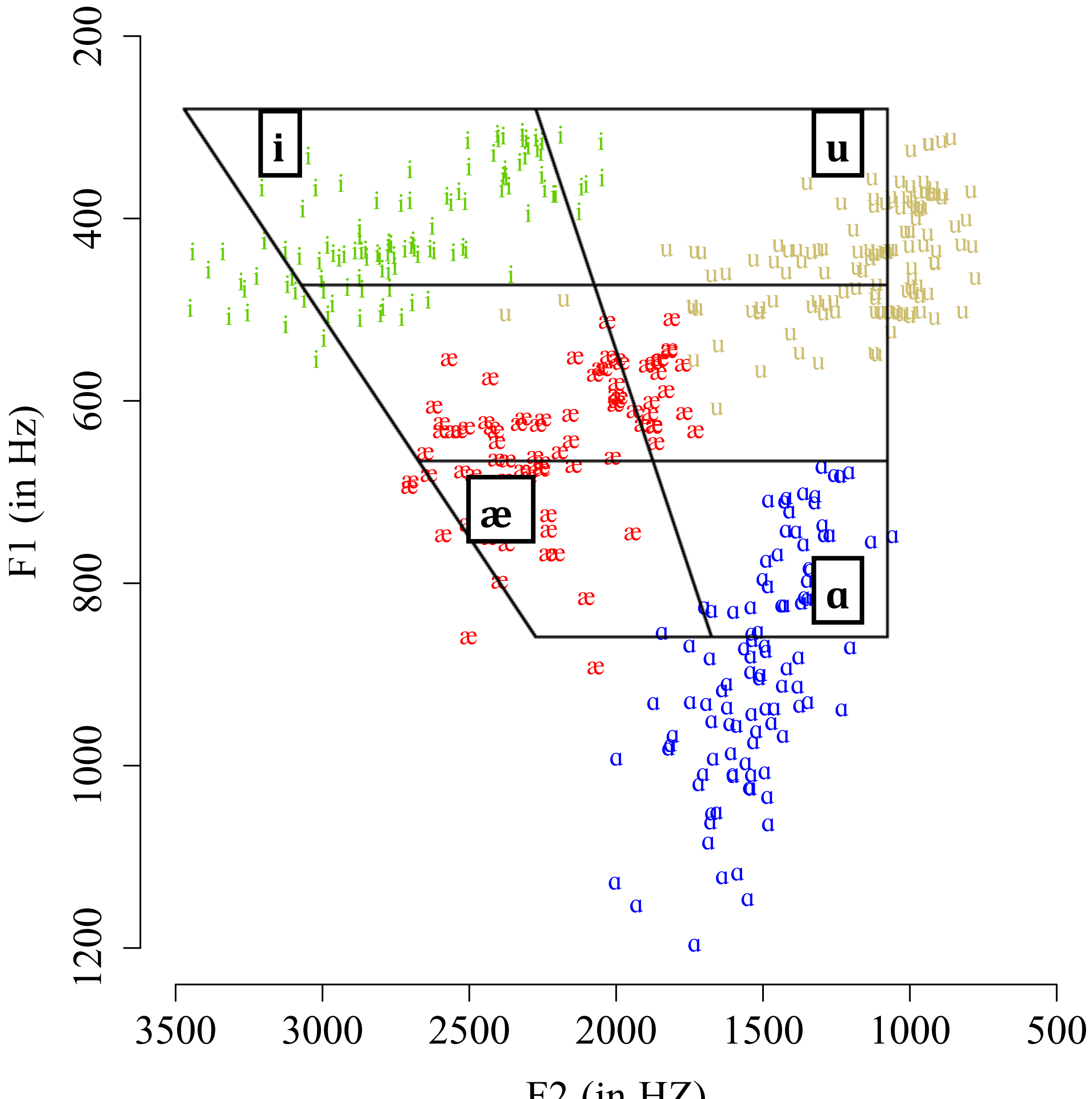
We can plot F1 and F2 for vowels



And here we can overlay a vowel chart (which, remember, is based on articulatory features!) to begin to the see the relationship between F1, F2, and articulation!

As you can see, setting aside speaker variation, the vowels appear in roughly the same spatial layout that they appear in the articulatory feature vowel chart!

We can plot F1 and F2 for vowels



This shows how tongue position directly affects the formants! Exactly as we would expect if the articulatory features of height and backness are responsible for the changes in vowel quality that we perceive!

Height affects F1
High = lower F1
Low = higher F1

Backness affects F2
Back = lower F2
Front = higher F2