

MITOCW | watch?v=V3N-Cwz-pC8

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high-quality educational resources for free. To make a donation or to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

PROFESSOR: So the cochlea is-- I'm just going to close this. Is that OK? The cochlea is long and skinny-- and maybe not? Maybe not. And the different parts of it are sensitive to different frequencies. Ringing any bells? Wow. Thanks.

AUDIENCE: That's where the higher [INAUDIBLE] I think, Abby said yesterday that for really low tones, it's basically an increase in the frequency [INAUDIBLE] but that at a certain point if that becomes impractical then-- and then certain areas are [INAUDIBLE]

PROFESSOR: So I think mainly for the frequencies that speech are in-- I'm here to talk about language, by the way-- the frequencies of speech are more like the different frequencies of the cochlea. Different parts are sensitive to different frequencies. So I'm just going to go with that story for now. But I'm sure Abby knows more about this than I do.

The different frequencies of the cochlea are sensitive to different frequencies and sound. Do you guys know much about waves? Basically, any signal, like any sound signal can be analyzed as a composite of different frequencies of waves. So it's going to be maybe a lot of 400 Hertz and a little bit of 600 Hertz or what have you, and you can basically entirely understand the sound, more or less, based on this analysis. So that's what the cochlea does. Do you understand this?

So cochlea will tell you, at any given moment, how present each of the frequencies are. Nod if you understand what I'm saying. Go like this. Do like this, if you don't understand. No? Yes? The cochlea is in your ear.

AUDIENCE: [INAUDIBLE]

That was nice of you guys to tell us.

Hey, Sara--

PROFESSOR: --sensitive to different frequencies. Also sensitive to different frequencies is my handy dandy computer program, which I will now show you. So this is what it sounds--

COMPUTER: b.

[LAUGHTER]

PROFESSOR: I don't know why it shakes. It says bee. So we have a bee sound going on and then an e vowel. So what's up here is just plain and simple, just the wave form. And what's going on down here is what I was talking about before with the frequency analysis.

So basically, along the x-axis, we have time. And then on the y-axis, we have many different frequencies. And then, the darkness of the dot is how strong that frequency is at that time. Does this make sense? So what you can see, for example, is that during this e vowel--

COMPUTER: e.

PROFESSOR: During the e vowel, we have many striations, like up and down stripes. Do you see this? Up and down stripes. And that's from, just plain and simple, the vocal chords-- just the sound going up and down. So sound is a periodic function, sometimes high pressure, sometimes low pressure. And that's what that's all about. So does this make sense? So this is voice. And you can tell that it's voice because of all these striations, especially down here. And this is a pitch track, which we should probably turn off.

So does everyone understand what's going on here? You're not at all confused?

So let me explain in more detail. So let's say we were to turn on the lights. Then, the vocal track looks kind of like this. This would be the mouth opening. These are the opening parts. This is the inside of your mouth. And then you have here your windpipe. And somewhere below here are some lungs. And somewhere down here you have your vocal folds, and those are going to vibrate create voicing.

So anytime you say a voice vowel, like b or ah or whatever, you're going to have these guys vibrating. And they're going to create those up and down striations. So the question is how do you make different sounds? I'll ask you. How do you make different sounds?

AUDIENCE: Obstructing the air flow from the windpipe use your mouth, different parts of your mouth.

PROFESSOR: Yeah, exactly. So basically, the difference between different sounds is going to have something to do with the way this whole thing is shaped. Does that make sense? So for example, if you were to make a buh, you would close your lips. So then this would be blocked off here, and then suddenly opened. Go like this if you understand.

And then the way your tongue is, is going to make this shape different. And then the way you're glottis is, is going to make the back different and all this sort of stuff. And this is how you make the different sounds.

So for example, the way you make different vowels-- if you try saying like awe, e, awe, e, or something like that, or pretend like you're doing it, then it's mainly your tongue moving. You can kind of-- if you want to try it, it would be OK. awe, e-- it's basically your tongue going one way or another.

And your tongue is going to create some sort of obstruction somewhere in here for each one. If you think about it, obstructing any part of your--of this whole pipe is going to make different cavities of different sizes. Does that make sense? So if you were to have-- this is just purely theoretical-- if you were to have an obstruction like here, then you would have a cavity of this size and a cavity of that size.

If you were to make one here or something like that-- I'm not sure you could-- but then you would have you-- you could have multiple ones. If you were to do this with your lips, you would make this one a little bit bigger by a few centimeters, two centimeters or so. Does that make sense?

So what do you guys know about resonance? Yeah?

AUDIENCE: When two waves have the similar-- hold the same frequency, they can have a fight with each other.

PROFESSOR: Right. So basically anything has what's called a fundamental frequency. And this is the frequency of a wave that would have a whole number of wavelengths with respect to the size of the thing. Something like that.

So for instance, the cavities in your vocal cord, like the vocal tract, the various parts of your vocal tract that may or may not be blocked off, are going to have different resonant frequencies. Are you following? So it depends on the size of the thing.

So the bigger part of your vocal tract is, the lower the resonant frequency. And then smaller parts of your vocal tract are going to have higher resonant frequencies. So if you're sending air through and you're making sound, what you're basically going to get is that certain frequencies are going to resonate and be really loud. And that's what we see in this picture. Can you still see it, even though the lights?

AUDIENCE: Yes.

PROFESSOR: Yes? OK. Good. So that's what you see in this picture. So for b-- or for the e vowel--

COMPUTER: e.

PROFESSOR: We have this is the frequency of one of the big cavities in the vocal tract. And this is going to be the frequency of another one, and this is going to be another one. So this is going to be the biggest one. This is the next biggest one, et cetera, et cetera. These are called formants. I'll write it down. So does that make sense?

So basically, changing your mouth, changing the shape of your mouth and your vocal tract, is going to directly change the frequencies of the formants. This program is basically doing exactly what the cochlea does. So this program takes all of the frequencies and figures out which are the strongest and which are the weakest. It basically analyzes the different magnitudes of each of the different frequencies that are all available.

So basically, that seems to be what the speech signal is all about. So when you're producing the speech signal, you're going to be changing the shape of the vocal tract, and that's going to be changing what the formants are going to look like. And your cochlea is mapping out the formants for you. That's what it's doing. Does this makes sense?

Then, we can see what is the difference between different vowels. I'll just play the whole thing.

COMPUTER: Dee, dee, bah, dah.

PROFESSOR: So let's just look at the difference between-- let's see, this is dee and this is bah. What are the differences between the formants of the dee and bah here?

AUDIENCE: There's a very big one which has a low frequency [INAUDIBLE].

PROFESSOR: For dee?

AUDIENCE: Yes.

PROFESSOR: So the formant one-- this is formant one. It's the bottom one-- is lower in dee than in bah. Are people seeing this? This one is lower frequency than this one. Maybe if we were to show the--

AUDIENCE: [INAUDIBLE]

PROFESSOR: So the red dots are trying to track performance here. So this one is lower than this one. Do you see that? How did the formant do?

AUDIENCE: In the awe sounds, it's one or two lower [INAUDIBLE]

PROFESSOR: So formant two is very high in e sound and very low in the awe sound. So this is basically the difference between what's called a high vowel, like e, and a low vowel, like awe, is the differences in formants one and two.

If you look at the spectrum of vowels-- spectrum. If you look at-- all the vowel systems in the world, seem to have vowels that-- there's sort of a trapezoidal thing going. So e is over here, and ooh is over here. And awe and ah is over here, and like uh, oh and eh. So this is some of the vowels of English.

If you look at vowel systems, languages can have three vowels systems or five vowels systems or seven or eight or however many, but they're all going to be distributed in here. And if you look at actually the formants of these, then the big differences between them is going to be between formants one and two.

So formant one is going to be highest for the low vowels, and lowest for the high vowels. And formant two is basically going to go the other way. No. Formant two is going to be front back. Formant two is going to create this distinction. And this all happens because of the way the tongue is when you say these vowels.

So when you say awe, e, you're making different obstructions. You're basically dividing the vocal tract in different ways and creating different resonant frequencies. And that's what's going on. Does this make sense? So that's what the vowels are about.

I just want to tell you maybe about consonants. I'll play it again.

COMPUTER: Bee, dee, baw, daw.

PROFESSOR: So if you look at the wave form, this is for, what? Dee?

COMPUTER: Bee.

PROFESSOR: Bee, yeah. This is for bee. So if you divide this up, as it works out, there is silence for a while. This is the part that happens when your mouth is closed and you're saying the part that

happens before the buh. So the buh sound is a release. I don't know if you've noticed, but if you say bee, there's part of it where your mouth is closed. And then, when your mouth opens, a buh sound happens, and then the vowel.

So this is the part where your mouth is closed. And then, this is going to be the release. I don't know if you can hear anything. That's the consonant itself-- the buh. And then it's going to open up into the vowel itself.

COMPUTER: Bee, bee.

PROFESSOR: The reason it sounds like there's a consonant at the beginning is because it starts kind of abruptly. So if you were to say a vowel with no consonant at the beginning at all, it would start very quietly and get louder and then get softer again. So you can see that it gets softer again. But it starts loudly because it's preceded by this stop-- the buh sound.

COMPUTER: Bee.

PROFESSOR: So this is all the vowel, and this is all the consonant. What is mainly the difference in volume between the vowel and the consonant? Yeah?

AUDIENCE: The consonant doesn't have much volume.

PROFESSOR: The consonant is really quiet compared to the vowel. So the consonant is going to-- this is what the consonant sounds like.

COMPUTER: [SOUND SIGNAL]

PROFESSOR: Versus--

COMPUTER: Bee.

PROFESSOR: So very quiet compared to the vowel. And let's maybe look at dee.

COMPUTER: Dee, dee.

PROFESSOR: So how about this one? Can you tell this is where the vowel is happening?

COMPUTER: Dee.

PROFESSOR: And this is the consonant.

COMPUTER: [SOUND SIGNAL]

PROFESSOR: So what do we think about this?

AUDIENCE: Well, in this one, the duh sound doesn't obstruct the airflow, so it's the starting point-- or the duh sound. And the starting point of the e sound is at the same volume [INAUDIBLE]

PROFESSOR: Yeah, I mean, it looks like what's happening here is that this is the release. This is where the tongue releases so you don't have a stop anymore-- stop in the air flow-- you don't have that anymore right here. And then, there's a little bit of noise, which is probably just air passing out of the mouth. And then the vowel seems to start. But it is louder. The stop itself is louder, but certainly-- louder than the buh one, but certainly not as loud as a vowel-- or not the main part of the vowel anyway.

And you can see the difference here. So duh part looks like this. So it's got the beginnings of a formants. You're ready to say the vowel. Your mouth is shaped that way. But you don't really-- they don't get really dark until here. This is when you can really see stuff going on. So it's also quieter.

Basically, what I'm trying to say is that consonants are much quieter than vowels. So we can look at these and see if they're the same.

COMPUTER: Baw, baw, baw.

PROFESSOR: I don't know why it shakes. So this is sort of the same thing going on, versus--

COMPUTER: Daw, daw.

PROFESSOR: So again, we see that duh is a little bit louder--

COMPUTER: [SOUND SIGNAL]

PROFESSOR: --than the buh was. But it's still nowhere near as loud as the main part of the vowel.

AUDIENCE: What if you drug it to that-- can you start playing it in the middle of the vowel? Would we still tend to hear a consonant?

PROFESSOR: You would hear a consonant, but it wouldn't be a dee. Listen.

COMPUTER: Awe, awe, awe, awe, awe, awe.

AUDIENCE: [INAUDIBLE] It's almost a bee.

PROFESSOR: Yeah, you could perceive it in a few ways.

COMPUTER: Awe, awe, awe awe.

PROFESSOR: So it sounds like a bee to you? We'll talk about that. So you think that there's a consonant there because the vowel started so abruptly. First, I didn't hit play, and then I did. But that's just the only reason. So it doesn't have the characteristic swell that the vowel has.

People understand that though the larger ones are louder? People are getting this?

AUDIENCE: Higher amplitudes [INAUDIBLE]

PROFESSOR: Right. Higher amplitude makes it louder, in this top part-- the wave form. This is different. So here's a question.

If we can't tell the difference between-- or if the consonants are so quiet, then how can we hear the difference between two consonants? Yeah?

AUDIENCE: Well one thing is the shape of the volume of the vowel. If you take the e, it usually goes from-- it starts from really loud and goes to get softer [INAUDIBLE] but if you say a dee sound, then [INAUDIBLE]

PROFESSOR: Yeah. The release of the stop is quieter in the buh sound than in the duh sound, which means that the abruptness of the vowel is also different-- just plain how loud it is, is going to be more abrupt in the buh sound than in the duh sound. You can see it here. People are understanding what I'm saying? So that's one thing.

So basically, the volume of the release. What else? This is, if anyone forgot, this is--

COMPUTER: Buh, duh.

AUDIENCE: I think the starting frequencies have different formants.

PROFESSOR: Yeah, they seem to be different, right? So in this one, it looks like the formants are pretty steady throughout-- formants one and two, at least, pretty steady throughout. Whereas in this one, it looks like formant one is starting from a lower frequency and getting higher. And formant two is starting from a higher frequency and getting lower. People see this? How this is

happening?

Why might this be? Why would we see a difference in how the beginnings of the formants are depending on the preceding consonant? Yeah?

AUDIENCE: Well, in daw, the shape of your tongue is different than when you're emphasizing the awe part. [INAUDIBLE] with buh, the tongue stays [INAUDIBLE]

PROFESSOR: Basically, it must have something to do with that time-- your lips are shaped one way for the duh part of the sound, and they're shaped another way for the awe part of the sound. And the mouth wasn't designed to talk. It was designed to eat. I don't know if you guys realize this-- probably, yes. It's slow.

Phoneticians hypothesize that we're basically speaking as fast as we can. That rates of speech in different languages are more or less similar. And if they got any louder, the mechanics of the mouth wouldn't actually be able to keep up with it. Whether perception can keep up with it is maybe another question.

There's going to be some sort of lag. There's going to be some lag time between when your mouth was in the duh shape and when your mouth is in the awe shape. And similarly, some lag between the buh shape and the awe shape, and the buh shape and the e shape, and all these sorts of things.

What we're seeing here is the lag. This is the part between the way your mouth was shaped from the duh, and the way your mouth was shaped for the awe. It's kind of our ramp time. And it's very noticeable. So if you look at this, it's pretty clear. This is exactly what the cochlea is doing.

The cochlea is basically analyzing how loud the different frequencies are in real time. And so basically, what the cochlea is doing is showing you this picture, showing your brain this picture. So the fact that this is happening is sort of really noticeable. Whereas-- these are getting kind of distracting.

Whereas, the loudness of the buh sound versus the duh sound, not really as noticeable. Does that makes sense?

Let's look at bee and dee [INAUDIBLE] How about this? What do you guys see? I'll play it for you.

COMPUTER: Bee, dee, bee, dee.

PROFESSOR: Ideas?

AUDIENCE: It will take the lag time [INAUDIBLE] because the dee shape is closer to-- the duh shape is closer to the [INAUDIBLE]

PROFESSOR: So it seems like we don't see that drastic lag time that we saw with the awe vowel, so exactly right. So it must be something like the shape of your mouth for a duh is closer to the shape of your mouth for e, dee. I mean, you can kind of tell-- dee-- we're thinking mostly of what your tongue is doing here, because that's the part of your mouth that's moving the most in this whole process.

And b, we also don't see some sort of exciting formant transition. This is what is called formant transition. We're talking about the beginnings and ends of the formants. So we do see something over here in formants three and four. They seem to dip. I'm sorry-- they seem to rise at the beginning, for bee more than with dee.

But formants one-- formant three is certainly important. But it seems that the lower numbers of formants, the lower formants tends to be more important in our perception than the higher formants. I guess because they represent larger parts of the vocal tract. Yeah?

AUDIENCE: Does each formant [INAUDIBLE]

PROFESSOR: Sort of. So basically for formants one and two seem to be more or less that-- there's typically one obstruction in the vocal tract going on at any one time, like one big one. And formants one and two seem to be related to each other, in that formant one is always the bigger half and formant two is always the smaller half.

And if you actually look at the vowels and you think about which formant corresponds to which half, you'll see that they switch, as one half gets bigger and one half gets-- it crosses the midpoint. But that's beside the point.

And then after that, are other resonant frequencies. The formants are going to take into account basically all, like-- they keep going. I've cut this off at 5,000 Hertz, but they keep going. There are many, many formants up there and they're resonating with all of the different parts of your head, all of your various sinuses, and all this sort of thing, and to various

degrees. So they get quieter and quieter as they go up.

So does that answer your question? Yeah. So what were we saying before?

We don't have this drastic difference in formants one and two for the difference between bee and dee. So maybe we're not going to really-- what I'm trying to get at is, what is-- if you look at this, and you ask yourself, how does a hearer perceive the difference between a bee-- the difference between baw and daw? You might, without thinking, or without looking at this picture, you might think that, well, the difference between baw and daw is the beginning buh and duh sounds.

But as you can see from this picture, the actual buh and duh sounds are extremely quiet with respect to the vowel. But if you look at the formants of the vowels, you can clearly tell that they are different. So one hypothesis might be that when you're perceiving sound it actually has more to do with the formant transitions of the vowel than it does-- with the difference in release, the volumes of the stops, or whatever.

And then, if you look at bee and dee, there's not quite this obvious difference in the formant transitions from the buh and the duh sound to the e. So maybe when perceiving bee and dee it might have more to do with the loudness of the stop release or something like that. Does that make sense? Go like this if you understand what's going. Go like this if you don't. I didn't see any perplexed faces, I guess.

So let's look at a different sound.

I'll just play this.

COMPUTER: Gun, gum.

PROFESSOR: Can you understand that? Is it a little loud? What did it say?

AUDIENCE: Gun, gum.

PROFESSOR: Gun, gum. Right. Gun, gum.

COMPUTER: Gun, gum.

PROFESSOR: Let's look at the guh part for now, try and review what we learned.

COMPUTER: Guh, guh.

PROFESSOR: That's a little bit zoomed in. So what's going on with the-- this is gun, I guess.

COMPUTER: Gun, gun.

PROFESSOR: What's going on with gun here? Can anyone summarize this situation? Here, I'll give you a hint. I think the guh is probably this much of it.

COMPUTER: [SOUND SIGNAL]

PROFESSOR: And then--

COMPUTER: Gun, gun, gun.

PROFESSOR: So what do you think is going on?

AUDIENCE: [INAUDIBLE] awe, except the middle part of it, [INAUDIBLE] an un, so that's part of it.

PROFESSOR: Yeah, the n. Probably this--

COMPUTER: un, un, un.

PROFESSOR: Can you hear that? That's an un.

COMPUTER: Un, un.

PROFESSOR: So what is happening with the formant transitions for just the guh part? Just the beginning.

AUDIENCE: I think one [INAUDIBLE]

PROFESSOR: Right. Oh, yeah. These dots at the beginning are trying to find formants in the guh, and may not be successful. So sorry, I should have mentioned probably, you can pretty much ignore anything before about this line. Any of this stuff is probably-- here, we'll just turn this off.

So the formant one has this distinctive rise at the beginning. How about formant two?

AUDIENCE: [INAUDIBLE]

PROFESSOR: Yeah. What happens at the beginning? This is formant two, right?

AUDIENCE: It's decreasing. It's going down.

PROFESSOR: Right. So it seems to start high and dip a little bit. And formant one, starts low and rises a little bit, just in the beginning. Let's just, for kicks, see if it's the same over here. This is also, if you remember, also starts with a guh, so probably should be similar. What do we see? Similar? Go like this if you think it's similar. Yeah.

So a little bit of a rise here, a little bit of a dip here. And formant three seems to be having also a rise. So that's kind of interesting. So that seems to be what's going on. Now let's look at the other transition, the one that's going into the n or the m.

What's happening here? Can you you guys see this? So this dark part is the vowel. And then, you can see it corresponds with loudness. It's very loud. And then this quieter part is the n sound. So the line is probably somewhere around here.

COMPUTER: Guh, n, n, n.

PROFESSOR: So what is going on with the formants here. We can turn that back on.

AUDIENCE: They get a lot quieter.

PROFESSOR: They get quieter for sure.

AUDIENCE: Looks like formant one drops off almost to zero.

PROFESSOR: Yeah, formant one disappears. And then, formant two is pretty much steady. Let's check out--

AUDIENCE: [INAUDIBLE] formant

PROFESSOR: Formant one is going to be the bottom formant. You see this? So they count from the bottom up. So this is the first formant, second formant, third formant. The formants are these big dark smudges we see. And also, the red dots are trying to track them. So that's the program trying to find formants for you, so you can measure them or whatever.

Let's check out the other one. So that was gun. How about this? What's happening?

AUDIENCE: Same thing. It's dropping off [INAUDIBLE]

PROFESSOR: Right. So it's getting quieter.

AUDIENCE: Also, formant two jumps up a bit before it comes steady.

PROFESSOR: Right. So formant two seems to, rather than have this straight across going from the uh sound

to the n sound, or going from the uh sound to the n sound in gun, we saw that it was more or less-- formant two was more or less steady all the way across.

But this one, it's more-- it dips and then rises. Well, the dip is probably just a vowel. But it seems to rise up abruptly for the n. So if we listen--

COMPUTER: M, m, m.

PROFESSOR: You can hear that it's an m sound. You can hear that it's mm and not nn. Yeah? You agree? And it seems like the main difference between mm and nn is where the second formant is. Assuming that the rest of the world is the same, which I I think we can pretty much safely assume that. Where did my mouse go? Down here? Oh, there it is.

We can even see it here. So pretty steady transition here, whereas sharp rise here. Make sense?

We could hypothesize that the difference between when you're perceiving gum and gun is whether the-- is going to be the formant transitions. Or we could hypothesize that it's the actual sound of the mm and nn. You could hear the difference, right? When I didn't play the vowel, you could hear the difference between mm and nn, right? Let's just try again.

COMPUTER: Nn, nn, mm, mm, mm.

AUDIENCE: [INAUDIBLE] play the beginning. I think if you play it from the beginning of the vowel [INAUDIBLE].

PROFESSOR: But I think even if you don't play-- if you don't get the formant transition, you can hear the difference.

COMPUTER: Mm, nn.

AUDIENCE: Unless this a placebo. Why don't we try it by [INAUDIBLE] Everybody close our eyes [INAUDIBLE] and then raise your right hand if it's nn, and left hand it's mm.

PROFESSOR: Good idea. Everyone close your eyes. We'll see if we can hear this.

AUDIENCE: [INAUDIBLE]

PROFESSOR: I'll just say who thinks it's mm? Who thinks it's nn? I'm going to play it.

COMPUTER: Nn, nn, nn, nn.

PROFESSOR: Who thinks it's m? Who thinks it's n?

PROFESSOR: That was like four and five, so I guess we can't hear it. OK-- can't hear it. But it is maybe different. It was n. It was an n sound. Congrats to the four people who thought it was n-- half of the class. So maybe not-- maybe we can't really hear it.

So if we can't really hear the difference between mm and nn, how can we tell the difference between gum and gun?

AUDIENCE: [INAUDIBLE] Context.

AUDIENCE: But she just said both and it wasn't in any [INAUDIBLE]

PROFESSOR: Right. I think I can say it out of nowhere and you can probably tell, even if you can't see my face. Gum, gun, gun, gum, gun, gun. Yeah? Maybe you can--

AUDIENCE: [INAUDIBLE]

PROFESSOR: Yeah. Or you could-- sometimes, like-- So if it's not the sound of the m or n sound, then how can we tell the difference? I just played it for you just without the transitions.

COMPUTER: Mm, nn.

PROFESSOR: And you couldn't really tell the difference. But what if I played you the whole thing?

COMPUTER: Gun.

PROFESSOR: Yeah?

AUDIENCE: Is it the actual change from the vowel to the consonant?

PROFESSOR: Yeah. So it's probably the change from the vowel to the consonant, because this is pretty much steady, and this is different. So it's probably like our baw, daw example where the formant transitions were different, really different. And in comparison, the actual stop release was not such a drastic difference.

So we might hypothesize that it's the formant transitions that we can tell. So if this is our hypothesis, how would we test it? What do you think?

AUDIENCE: You mean by different [INAUDIBLE]

PROFESSOR: Yeah.

AUDIENCE: We can start off how we just did with our eyes closed and half of us [INAUDIBLE] if you started it from an earlier point where the transitions [INAUDIBLE] and see how many people could guess what it was.

PROFESSOR: Yeah, exactly. So maybe we could play smaller parts of the speech signal and see what parts of speech signal you actually have to hear in order to hear the difference. I will point out, by the way, that when I first brought this up, all I did was play it and I asked you what it said-- everybody knew, even though it was completely out of context.

So maybe if we chop up the speech signal in strategic ways, and then play it, we can maybe hear-- statistically figure out what people are actually hearing. So why don't we do something like that?

So we could, for example, if we just play the guh part.

COMPUTER: Guh, guh, guh.

AUDIENCE: [INAUDIBLE]

COMPUTER: Gum, gum, gum.

PROFESSOR: Yeah just through the transition.

COMPUTER: Gum, gum, gum. Gun, gun.

PROFESSOR: Can you hear the difference? So it seems to me more this very beginning part of the nasal or the end part of the vowel is doing more for us. Maybe we could try just the end of it.

COMPUTER: Un, un. Um, um, um.

PROFESSOR: Could you hear the difference then?

AUDIENCE: Yes.

PROFESSOR: Yeah, much clearer, right? Much clearer. You agree? Go like this if you agree. Go like this if you don't agree. Everyone agrees. That's very nice.

Or let's say that we were to zoom in so we can do some precision work. Oh, this is gum, I think. If we were to just stick the m in here-- I included the transition here sort of. You see, it might sound weird. Let's see what it sounds like.

COMPUTER: Gun, gun, gun.

PROFESSOR: It sounded like gun, I think. Let's try-- that's not what I want to do. Can I undo again? No. Sorry. Where am I? [INAUDIBLE] at the end. Is it back? Oh, no, it looks scary. Hold on. let me fix this.

Let's try and grab this end bit. We should copy. That sounds good. We'll stick it here, make sure that we get-- we're moving over the transition as well.

COMPUTER: Gum, gun, gun, gum.

PROFESSOR: What did I do?

AUDIENCE: You changed it from gun to gum.

PROFESSOR: Yeah. So--

COMPUTER: Gun, gum, gum.

PROFESSOR: So it seems that if I make very careful to include the end bit of a vowel, then I can successfully change the word. So to review, what I did here was I took the m, including, making sure to include the formant transition. So I was including the end of the vowel, and I successfully changed the word to gum. So let's try again, except this time, I'll not include the end bit of the vowel. That's actually what I did the first time. Now it's here.

COMPUTER: Gun, gun, gun.

PROFESSOR: Now what is it?

AUDIENCE: [INAUDIBLE]

PROFESSOR: Now it's gun, right? So this is very interesting. So what this means is that it must be something about the formant transitions that we're hearing, because-- that is the most salient auditory cue. Because if I included the formant transitions, I could successfully take the n and paste it onto and change the word. But if I didn't include the formant transitions, I couldn't. I didn't change the word. People following what's happening here?

So this experiment suggests that it must be something about the formant transitions that's making us hear the difference. And there may be other cues. Certainly there are other cues, like m and then n itself look different. Let me just undo this. A little bit.

But for one thing, the m and the n parts are significantly quieter than the vowels. See, here is the vowel-- much louder. Here's the nasal-- not loud. And for another thing, they're not that different. It must be-- so it makes sense that it's the formant transitions that we listen to.

AUDIENCE: Do you call it [INAUDIBLE]

PROFESSOR: Yeah, sorry, the nasal consonant. Right. The consonants are much quieter than the vowels. So it makes sense that what we're really listening to is the vowel. Let me play something else for you.

COMPUTER: Lose, lose.

PROFESSOR: What's the word?

AUDIENCE: Lose.

PROFESSOR: Lose, like lose a game, right? Something like that?

AUDIENCE: [INAUDIBLE]

PROFESSOR: This guy's British.

AUDIENCE: Oh, yeah. That makes sense.

COMPUTER: Lose.

PROFESSOR: He's very British.

AUDIENCE: Lose.

PROFESSOR: Lose-- yeah. It's not quite the vowel we use, right? It's not quite the American vowel. There's something about the vowel that's a little bit different, and that's what makes it sound British. So let me break it down for you. Here is the l part, the luh part.

COMPUTER: Luh, luh.

PROFESSOR: Well, maybe I got a little bit of oo.

COMPUTER: Loo, loo.

PROFESSOR: And then we have the ooh. See this.

COMPUTER: Ooh, ooh.

PROFESSOR: And then here is the zuh.

COMPUTER: [SOUND SIGNAL]

PROFESSOR: So let me ask you, what is the difference between between the word lose and the word loose?
Yeah?

AUDIENCE: The [INAUDIBLE] consonant in his voice.

PROFESSOR: In which one?

AUDIENCE: In lose.

PROFESSOR: In lose, right. So in lose, it's a Z sound. Lose-- I'm going to write them down. Make sure I-- So this a Z sound. And this is an S. People understand? You get this? This is a Z-- lose. This is an S-- loose. Yeah?

So this is lose, so it should have a Z sound at the end here.

COMPUTER: [SOUND SIGNAL]

PROFESSOR: What does it sound like? It sounds like an s, right?

COMPUTER: [SOUND SIGNAL]

PROFESSOR: So in theory, this is voice, but in actuality kind of not. If we look at-- so the characteristic of voicing is this striations, these up and down stripes, in the bottom here. If we look, they peter out really fast. So for the zuh part of this, it's only really a Z for this much time, and then it is all s.

Furthermore, the furcation noise-- that's the fuzziness that happens when you say a Z or an S-- is really loud, compared to what's going on here. This is not very loud, and this is kind of a big deal. Does this makes sense? So I ask you, if it's not really a zuh that's happening, if it's

really more like an S-- This is s sound-- then how can we tell that this is lose and not loose?

AUDIENCE: [INAUDIBLE]

PROFESSOR: Sure, maybe it's just the beginning voicing. Yeah?

AUDIENCE: Does it have to do with the volume-- like what happens after?

PROFESSOR: After what?

AUDIENCE: After the transition from the ooh to the suh.

PROFESSOR: Yeah, maybe it's something about the furcation noise. What it probably is, and I don't really expect you guys to have known this, but the actual time that it takes to say the vowel, the actual length the vowel is going to be much longer in lose than in loose. If and you actually-- I don't have loose. I really should.

If you actually look at lose versus loose, lose is going to be-- are you OK?

AUDIENCE: What are you doing?

AUDIENCE: [INAUDIBLE]

PROFESSOR: Oh.

AUDIENCE: Never mind. [INAUDIBLE]

That was really strange.

PROFESSOR: So the vowel in lose is actually like-- it takes twice as long to say. You actually spend more time saying it. It's a long vowel. Whereas in loose, it's not. And I'm not going to do this, because it's going to take a really long time. But if you actually take half of the length of this vowel away, and you just cut, use edit, cut, control whatever, X or whatever, then it will actually start sounding like loose.

The reason I'm not going to do for you is that to preserve this beautiful arc here, you actually have to take away every other period sort of thing, which is painstaking work.

This is kind of another example of-- what we are trying to signal is a difference in the consonant. We're trying to signal the difference between a ss and a zz. And that's supposed to be a voicing difference. But in English, we don't-- it's not that the voicing is what's different.

The voicing is hardly different. It's just this amount of different. Hardly different at all. But the vowel length is really very different.

So this is just another example of where we're listening to the much louder vowel to get our signal, to get relevant cues, than to the actual consonant. Yeah?

AUDIENCE: Doesn't it have to be something like a consonant too, because if you say two words that have the exact same length, vowel, you can still the difference between an S and a Z.

PROFESSOR: I mean, maybe. Although-- what?

AUDIENCE: [INAUDIBLE] hearing sounds [INAUDIBLE] same vowel length.

PROFESSOR: I mean you can say a Z with full voicing. You can go zzz, if you want to. It's possible to make that distinction. But, like I said, if I were to take this signal and make the duration of the vowel shorter, you would hear a loose. If I were to make this half the length, then you would actually hear loose. So that suggests that the vowel length is some really relevant signal. Does that makes sense? That's very interesting.

Here's one. Hello? You guys don't need to see that, I realize.

So before I play this whole thing for you, let me just play the n bit, and you guys can tell me what you hear.

COMPUTER: A, a, a, a,

PROFESSOR: What does it sound like?

AUDIENCE: A.

PROFESSOR: A, sure. Yeah.

AUDIENCE: The consonant a.

PROFESSOR: Yeah, an a that comes out of nowhere, or something. It's not a or I don't know-- hard to do. Yeah, some sort of consonant maybe.

COMPUTER: A, a, a, a, a.

PROFESSOR: So it sounded like a.

AUDIENCE: Yes.

COMPUTER: A, a, a, a.

PROFESSOR: Is it sounding different at all to you guys?

AUDIENCE: [INAUDIBLE]

PROFESSOR: More like bay.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Yeah, right. Sounds like-- kind of like bay.

AUDIENCE: Sounds like grey.

PROFESSOR: Grey?

AUDIENCE: Sounds like gay to me.

PROFESSOR: Gay, bay, something like that. Yeah. It sounds like there's a consonant there-- some consonant at the beginning. How about here?

COMPUTER: Bay, bay, bay.

AUDIENCE: Sounds like [INAUDIBLE]

PROFESSOR: What?

AUDIENCE: Bay.

PROFESSOR: Bay, still-- is that what you're hearing?

AUDIENCE: [INAUDIBLE] I heard a B at the end.

PROFESSOR: At the end?

COMPUTER: Bay, bay, bay.

AUDIENCE: It's not one vowel. It's a diphthong.

PROFESSOR: Yes, it's a diphthong. We're focusing on the beginning here.

COMPUTER: Bay, bay, bay, bay.

PROFESSOR: What do you guys hear? Still the same?

AUDIENCE: Day.

PROFESSOR: Day? hearing day. Some people are hearing bay.

AUDIENCE: [INAUDIBLE]

COMPUTER: Bay, bay, day, day, day.

PROFESSOR: Still bay, day--

AUDIENCE: It switches off between bay and day.

COMPUTER: Day, day, day, day.

PROFESSOR: How about now?

AUDIENCE: Date.

PROFESSOR: Day, they, date.

COMPUTER: Date, date, date.

PROFESSOR: This was actually taken out of a sentence. So it makes sense to you here also a consonant at the end, because the vowel stops abruptly, because it was just sort of cut and pasted from the middle of a sentence.

COMPUTER: Day, day.

PROFESSOR: People are hearing mostly day here?

AUDIENCE: Yeah.

PROFESSOR: Day? Let's keep going back.

COMPUTER: Day, day, day, day.

PROFESSOR: Still day?

AUDIENCE: [INAUDIBLE]

COMPUTER: Day, day, day, day.

PROFESSOR: Still day? I'll play you the whole thing.

COMPUTER: Say.

AUDIENCE: Say.

AUDIENCE: [INAUDIBLE] the last time right before that, and I was like, oh.

COMPUTER: Say, say.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Say-- anyone surprised?

AUDIENCE: [INAUDIBLE]

[INTERPOSING VOICES]

PROFESSOR: Huh? Oh, yeah. Right. The name of the file is set to yes. What's going on here?

AUDIENCE: I heard it right before.

PROFESSOR: Yeah.

COMPUTER: Say, say, say.

PROFESSOR: You can see from this demonstration how we're interpreting the signal. So back when we were over here, you guys were saying it sounds like a or bay or something. It sounds like a vowel and it comes out of nowhere, so it must have a consonant at the beginning. So your brain sort of fills that in, like oh, there must have been a consonant there, because it's a vowel coming out of nowhere. If you were just to say a, then you would probably have a sort of characteristic rise that we've seen before, sort of swell into it.

And then as I got farther-- what happened? My computer is freaking out.

So as I got farther back, you guys started hearing day. Why do you think you would have heard day? As I got closer and closer to here, you guys started hearing day rather than bay,

right? So why was that?

AUDIENCE: Maybe because bay was more abrupt-- stop [INAUDIBLE] it's so abrupt, and so we assumed it was that until we started hearing the ss sound.

PROFESSOR: Right. So you guys started hearing day-- I don't know when my computer is freaking out-- you guys started hearing day around here, Right so maybe it was a little less abrupt. Anything else?

AUDIENCE: [INAUDIBLE]

PROFESSOR: Yeah, you didn't hear the furcation noise, is what this is called, of the-- because S is a fricative, and that's a furcation noise. So you didn't hear that happening. So you weren't perceiving S. But the question is, since you weren't perceiving S, why was it day that you were perceiving, and not something else, like bay. Because you were perceiving bay a lot in here.

So you suggested that it's a little less abrupt to go into there, maybe because of the release for day being louder. What do you think?

AUDIENCE: Maybe the shape of [INAUDIBLE] the formant changes the closer you're [INAUDIBLE]

PROFESSOR: Yeah. So if you think about it, the way your mouth is for an S and a d, there's a closure in a very similar place.

AUDIENCE: Didn't we notice before that with the d that there was less of a formant change.

PROFESSOR: Yeah for awe, right? So we have yet to see that for the a vowel, diphthong. But if you think about it, the way your mouth is for an S and duh, both have a closure right here. Try it-- ss, duh, duh, duh-- very close, like behind your top teeth, behind your teeth, that's where the closure is happening.

Basically, what that means is that the different sort of cavities in your vocal tract are going to be very similar. That's too bad.

The cavities in your mouth are going to be really similar for day and say. So the formant transitions for day and say are also going to be really similar. So the way to figure out the difference between day and say is going to be something like is it a fricative or is it a stop?

So when we were starting just at the end of the furcation or the beginning of a vowel,

somewhere in there, we were thinking it's a stop, because we didn't hear all that furcation noise. We heard the vowel coming very loudly out of nowhere. But when we moved it back to hear the whole furcation noise, then we could hear that it was say. People are getting this? Understanding this? Yeah?

Do you guys like this? I'm trying out a new background. I just found this yesterday. I'm not sure whether I like it.

AUDIENCE: [INAUDIBLE]

PROFESSOR: I know. I kind of wish that they were leaves all the way down, and not stars.

AUDIENCE: Did you make it?

PROFESSOR: I didn't make it.

[INTERPOSING VOICES]

AUDIENCE: It's like maple leaves.

PROFESSOR: I mean, this one looks like a star, right?

AUDIENCE: Oh, I see it.

[INTERPOSING VOICES]

COMPUTER: Say, say.

PROFESSOR: What was I going to say? Oh, yeah. Incidentally, a lot of these studies are done with women-- sorry-- a lot of these studies are done with men.

[INTERPOSING VOICES]

This is a sort of exceptional, in that this is a woman. It's sort of exceptional, because the thing is that, on average, men and women speak at different pitches. And yes, really. So those interact with the formants. So if you think of this, the formants have something to do with the resonances in your vocal tract. That's what they have to do with. And the pitch of your voice has to do with how quickly your vocal chords are vibrating.

So if you think about it, these are completely independent things. Do you understand this? The

formants have to do with the shape of your mouth, and the pitch has to do with the speed of air and all that sort of thing. And these are independent.

But what you can see is that, depending on the frequency-- So if you're producing sound at a certain pitch, if you're producing a pitch of a vowel, then it has harmonics. So if you produce something at like 400 Hertz, then it's going to be loudest at 400, and 800, and 1200, and 1600, and so on. And then, in between--

AUDIENCE: Don't you multiply by two each time?

PROFESSOR: 1200 comes in there, I believe. And in between them, you're going to have like-- at 400-- sorry, 400, 800-- so at 600, it's going to be less quiet. So what ends up happening is that like regardless you're not going to see much going on at 600.

Basically, these little dots are indicative of where those gaps are. If you see the white dots, that's where you can't really hear it because of the pitch that her voice is at. And for men, because it's a lower pitch, they're going to be differently spaced. So if we were to look at Edward again, which we'll do in a second, I guess, then we could see that he didn't interfere with the formant here.

Sometimes, you have some issues actually finding where is formant two, or where it's not, because you have this pitch thing going on. So that's why a lot of this stuff is done with men's voices, because it tends not to interfere with the lower formants when you're just looking at the pictures. That isn't really relevant.

I guess that is more or less what I had to say. Do you guys have questions? No questions? What?

AUDIENCE: I see the stars.

[INTERPOSING VOICES]

PROFESSOR: Do you guys want to see some cool Burmese stuff? Burmese?

AUDIENCE: Yeah.

PROFESSOR: OK.

[INTERPOSING VOICES]

AUDIENCE: Oh, Burmese.

[INTERPOSING VOICES]

PROFESSOR: Yes, with the Buddhists.

[INTERPOSING VOICES]

PROFESSOR: This is Burmese.

[INTERPOSING VOICES]

PROFESSOR: That was a guy, yeah.

AUDIENCE: That was a guy?

PROFESSOR: Yes.

COMPUTER: Ma, ma.

PROFESSOR: This is the same guy.

AUDIENCE: Oh, it's a guy.

COMPUTER: Ma, ma.

PROFESSOR: This guy-- so Burmese has tones. It's a tonal language. Do you guys know what that means?

AUDIENCE: Is it like Chinese.

AUDIENCE: Japanese, Chinese.

PROFESSOR: It's like Chinese. Yeah.

AUDIENCE: Japanese [INAUDIBLE]

PROFESSOR: Japanese isn't really tonal.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Yeah. So the difference-- So what makes it a tonal language is that something about the pitch

of the vowel or the shape-- whether it's going up or down, makes a distinction. So like in Chinese, there's going to be difference between Ma and ma, something like that. And those are different words.

AUDIENCE: It's like how you express anger. because like in English, you use your tone to express differently.

You just yell the word loudly.

[INTERPOSING VOICES]

PROFESSOR: So this is a good question. And the answer to your-- is the word you should Google is prosody. Prosody is the study of the pitches of sentences and entire utterances, and how those pitches change, and how that relates to what you're saying. So for example, in English, we can use prosody to make the difference between a declarative sentence and a question, like you ate earlier today. Versus, you ate earlier today? Or something like that.

So it's not the words themselves, but the porosity that makes a difference. And porosity in Chinese is different from porosity in English is the short answer. Burmese is tonal. Burmese also has this weird thing called voiceless nasals. So let's just look at one.

If you think of a nasal like mm in English, pretty much whenever you say it, it has voicing going on. Right there-- their vocal chords are vibrating during the mm. Mm-- it's happening. You can feel it-- mm. And in Burmese, this is true as well. But they have something called the voiceless nasal. So this is supposed to be the same word with the same tone, and the difference is that this one-- or sorry, these are supposed to be different words, but they're supposed to have the same tone and the same vowel-- and the difference is that this has a voiceless m versus a voiced m here. So let's see if you can hear the difference.

COMPUTER: Ma. Muh.

PROFESSOR: Anyone?

AUDIENCE: Yeah, slightly.

PROFESSOR: The vowel is shorter.

COMPUTER: Muh, ma.

PROFESSOR: I mean, in truth, they're both voiced. We can see this beautiful voicing down here, and down here. And maybe the length is different. I don't know. Let's see. The length of this one is like 57 milliseconds, and this one is like 42 milliseconds. So it's maybe a little bit shorter. Yeah?

AUDIENCE: I think it might be easier to pronounce voiceless nasal [INAUDIBLE] shorter. It might be easier, because--

PROFESSOR: So phoneticians--

[INTERPOSING VOICES]

PROFESSOR: Yeah, you have to voice it to hear it basically. So phoneticians and Burmese speakers have noticed that in the voiceless nasals, what happens is that before the nasal happens, there's some air flowing out of your mouth-- or sorry, flowing out of your nose.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Ma, right? So ma. So there's air flowing out of the nose before the nasal. But I found this kind of unlikely. It's really unlikely that the difference that you're hearing is actually the air flowing out of your nose-- not very loud, right? Really quiet.

So I actually did a study where I took-- I basically did a copy and paste study. So it took everything before the vowel for two words that were supposed to be the same, except for the voicing of the nasal, and I pasted them in the wrong way, and then I also pasted them in the right way with other instances of the same word, and they always judged the meaning of the word to be the one that associated with the correct vowel.

So if I took this off here, and put on the voice one, they would still think it was unvoiced. And if I took off the voiceless one-- or if I put a voiceless onto here instead of this, they would still think that it was voiced. So I never figured out exactly what the difference was, but it must be something to do with the vowel that's making the difference. Does that make sense? This is another one of those stereotypical things.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Totally out of context. So I just played them the word. I randomly mixed them up using-- I think, actually, I drew numbers out of a hat-- named them things that had nothing-- like just letters or something-- had nothing to do with what they meant, played them for them, had lots

of controls, and just said, what does this mean? Does it mean to marinate or a celestial body, or whatever, because this is like the differences that these have.

And they give me really significant results. So it was really cool. And it has everything to do with the vowel. I hypothesized that it might have something to do with breathiness of the vowel. But that's also intertwined with the Burmese tone system, so it's definitely not conclusive.

So that's an interesting thing that Burmese has. It also has tones, so I wanted to play those for you.

AUDIENCE: Do you speak Burmese?

PROFESSOR: I do not.

AUDIENCE: So wasn't it really [INAUDIBLE]

PROFESSOR: Hmm?

AUDIENCE: Wasn't it like really difficult [INAUDIBLE] experiment [INAUDIBLE]

PROFESSOR: I was asking Burmese speakers.

AUDIENCE: Oh.

PROFESSOR: I wasn't doing it with myself, obviously. So here's a voiced-- So here is one tone.

COMPUTER: Ma.

PROFESSOR: Here's another tone.

COMPUTER: Mah, mah.

PROFESSOR: Can you hear the difference?

AUDIENCE: Yeah.

PROFESSOR: So the first time is--

COMPUTER: Ma, ma. Mah.

PROFESSOR: So the first tone is falling and shorter-- falling, shorter, creaky, while this is kind of breathy. It's

not really a tone system, because there's actually not just pitches involved, but also different stuff going on. So it's kind of a crazy tone system Burmese has. And it has two more tones in addition to this. And it also has-- some syllables have no tone, somehow. But they are special syllables. So it's pretty complicated.

So for example-- maybe I'll play you one more thing since we have two minutes or zero minutes.

COMPUTER: Ma, ma, mah.

PROFESSOR: Those are different. And I think this doesn't always go--

COMPUTER: Ma, ma.

PROFESSOR: --as reliably. And this one sometimes falls. It's kind of complicated. So these are--

AUDIENCE: What does ma mean?

PROFESSOR: I don't know. I don't know. Once upon a time I knew.

COMPUTER: Mah, mah.

PROFESSOR: So this is supposed to be a falling tone, but it doesn't fall very far, as you see. Maybe it isn't a falling tone. I don't know. So that was kind of cool. That's basically it. Class is over.