

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 view additional materials from hundreds of MIT courses, visit [MITOpenCourseWare@OCW.MIT.edu](mailto:MITOpenCourseWare@OCW.MIT.edu).

**PROFESSOR:**

Welcome. One quick announcement-- if you have not yet picked up your graded exams, you can do so by seeing the TAs after the hour. OK? So today I want to continue to think about what we started last week, thinking about Fourier series. The idea is to develop a theory that lets us look at signals on the basis of frequency content, much as we looked at frequency responses as a characterization of systems, according to the way they process frequencies.

And we saw last time that there were a number of kinds of signals, for example, musical signals, where that kind of an on approach-- thinking about the signal according to the frequencies that are in it-- makes a lot of sense and can lead to insight. We also developed some formalism. We figured out how you can break a signal into components and then assemble the components to generate the signal.

And what I want to mention at the beginning of the hour today is just how to think about this operation in a more familiar way. We do this kind of a thing, breaking something into components all the time. One of the more familiar examples might be thinking about 3-space, right?

The Cartesian analysis of 3-space is based on the idea that you can think about a vector location in 3-space as having components. There's a component in the x direction, the y direction, the z direction. That's completely analogous to the way we're thinking about Fourier representations for signals.

So just like we would think about synthesizing the location of a point by adding together three pieces, and we would think about analyzing a point to figure out how big the components are in each of those directions, it's exactly the same when we think about Fourier series. We think about representing a signal as a sum of things. So the sum is precisely the same.

This one happens to have an infinite number of terms, [? eh. ?] The top one has three terms, [? eh. ?] The principles are very similar. So we think about representing a signal as a sum of components. We think about representing a point in 3-space as a sum of components, and we think about analyzing the signal or the vector in 3-space, so that we figure out what each of those components are.

And we do it in an operation where it's actually very convenient to think about the decomposition of the Fourier components using precisely the same language that we would use for thinking about vector spaces. So we would think about-- in the case of the Fourier, we think about integrating over the period sifts out a component. The analogous operation for 3-space is to think about a dot product. The way you take a vector and figure out the component in the x direction is to dot it with it, the dot product.

In the Fourier case, we think about it as being an inner product. The idea is completely analogous. So we think about having the inner product of two things-- the reference direction and the vector. So reference direction and vector-- we think about it exactly the same way, except now it's an inner product, which means that after we've multiplied, we have to integrate. That's the only difference between inner product. Inner product implies some are [? integrated ?] after you've done the multiplication.

So we do exactly the same thing, except that now we think about the inner product of a and b. That's just the integral, where we take the complex conjugate of one of the signals only because by defining it with a complex conjugate there, we set up the inner product so that the answer is zero unless we take the inner product of two things in the same direction. OK? By putting the minus sign there, if the two reference directions, that is to say the one characterized by  $k$  and  $m$ , the ones characterized by  $k$  and  $m$ , the inner product will be zero if we take the complex conjugate as long as  $k$  is not equal to  $m$ .  $k$  equals  $m$  is the only not zero component. OK, is that all clear?

So to make sure that it's clear, here's a question. How many of the following pairs of functions are orthogonal in  $T$  equals 3? Part of the goal of the exercise is to figure out what the little caveat in  $T$  equals three means. So look at your neighbor, say hello, figure out a number between 0 and 4.

[SIDE CONVERSATIONS]

OK, so how many signals, how many of the pairs are orthogonal to each other? Raise your hand with some number between 0 and 4, unless you're completely bizarre and raise five, I mean. OK, come on, come on. Higher, so I can see them. Remember if you're wrong. It's your partner's fault, it's not your fault.

OK, not quite. A lot of bad partners, no, no. Let's do the first one. Is the  $\cos$  of  $2\pi t$  orthogonal to the  $\sin$  of  $2\pi t$  over the interval capital  $T$  equals 3? Yes? No.

I haven't a clue. I don't care. No, no, no, no. You all care, no. Are they orthogonal? So what do you-- how do I formally ask the question are they are orthogonal?

OK, so it's either the last slide or the next slide. So go back to the last slide. What's it mean if they're orthogonal? Yeah?

**AUDIENCE:**

[INAUDIBLE]

**PROFESSOR:**

So how do I take the dot product? What do I do?

**AUDIENCE:**

[INAUDIBLE] conjugate.

**PROFESSOR:**

Conjugate 1. So I want to-- I'm thinking about  $1$  over  $T$ , the integral over  $T$ , a star of  $T$ ,  $b$  of  $t$ ,  $dt$ . Right? So the  $t$  comes in here, right, I'm integrating over a period  $t$ . So I take the two functions and I multiply them together.

So I have this function, and I have that function. I multiply them together. If you multiply two sinusoids of the same frequency but

different phase, what do you get? Another sinusoid, right? So you all know all these complicated trig relationships, right?

Here's one of them. If you multiply  $\cos$  of  $2\pi T$  times the sine of  $\pi T$  you get half the sign of double the frequency. OK? You don't need to memorize that. You just look at this picture, you look at that picture.

This one over the interval 3 has 3 periods. Right? There are 3 periods of that waveform over the period of capital T. There's an integer number 3, same here.

Here-- how many periods? Twice that. But it's exactly six. So you get a pure sinusoid.

You get an integer number of periods. You integrate over an integer number of periods, you get 0. They're orthogonal. Had I chosen the period differently, they may not have been orthogonal. It depends on the period. OK? So the inner product depends on the period, because the inner product has something to do with integrator sum. And so the range over which you sum or integrate matters.

How about  $\cos 2\pi T \cos 4\pi T$ ? Orthogonal? Yeah?

**AUDIENCE:**

Yes.

**PROFESSOR:**

And the reason is?

**AUDIENCE:**

So think if you wrapped [INAUDIBLE] together, then there's a lot of symmetry that goes on [INAUDIBLE] is going to be 0.

**PROFESSOR:**

So now we've got two different frequencies. But we still get these funny cosine relationships that have to do with sums and differences. And the sums and differences both happen to be periodic over 3, over the interval capital T equals 3, right?

So we still get the property that the average value here, which is what the interval was pulling out, the average is 0. So they're also orthogonal. How about  $\cos 2\pi T \sin \pi T$ ? OK, I've asked two

questions, and they were both yes.

So I'm getting bored at this point, so by the theory of questions in lecture, the answer is?

[LAUGHTER]

Now, wait a minute. I'm not that boring. Well, maybe. So is this periodic over a capital T equals 3? Ah, excuse me. I didn't say the right question, sorry.

Is this function have an integer number of periods in the time interval capital T equals 3? What's the period-- what's the fundamental period of this waveform?

**AUDIENCE:** 1.

**PROFESSOR:** 1. So it has 3 periods over the interval cap T equals 3. What about this one?

**AUDIENCE:** 2.

**PROFESSOR:** Period is 2. How many periods are there in the time interval capital T equals 3?

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** A period is 2. How many periods are there-- 1 and 1/2, not an integer. Bad news, right? So integer number three, not integer number.

If you were to integrate this over the period t equals 3, if I didn't multiply them if I just did that, if I just thought about that integral, I wouldn't get 0, right? There's more positives than there are negatives. And when I multiply them, the same sort of thing happens. I get two big peaks down and only one big peak up. It's because the resulting waveform no longer has an integer number of periods in the interval capital T equals 3. OK?

Last one--  $\cos 2\pi T e$  to the-- whoops. Is that what I actually said?

Good, I forgot the  $j$ . Because without the  $j$ , they would obviously not be orthogonal. Obviously, right? OK. I didn't mean to ask something quite that obvious.

So what about  $\cos 2\pi T$  and  $e^{j 2\pi T}$ ? Orthogonal. Not, I'm as clueless as I was on Part A. No, no, no, no, you're not. No, you're not. No, you're not.

So how do you think about that? You can use Euler's expression. And if there had been a  $j$  there, this would have been a correct expression. OK? It's not quite a correct expression because I forgot to put the  $j$  there.

But had there been a  $j$  there, it would have been  $\cos 2\pi T$  plus  $j \sin 2\pi T$ . And the awkward thing is that the  $\cos$  and the  $\sin$  are obviously not orthogonal with each other. A signal is not orthogonal with itself, OK?

So because part of this signal is that signal, those two signals are not orthogonal. OK? Yes? OK, so that's kind of-- so that's the idea of orthogonality. It's a very good way to think about decompositions.

And even though we only spent about half an hour last time, and only about 15 minutes this time, that is the whole theory of Fourier series. That doesn't mean we can't ask hard questions. There were a couple of questions. Yes, you were first.

**AUDIENCE:**

Is there a way to think about orthogonality using the Fourier [INAUDIBLE].

**PROFESSOR:**

Well, the Fourier coefficients are the result of orthogonality. I don't think you can tell-- if I just told you a bunch of Fourier coefficients, I don't know if you can tell me something about the orthogonality of the underlying signals or not.

**AUDIENCE:**

What if [INAUDIBLE].

**PROFESSOR:**

Excuse me?

**AUDIENCE:** [INAUDIBLE] [? the period ?] and the Fourier [INAUDIBLE].

**PROFESSOR:** Let's see, so I'm not completely sure I know what you're asking. Certainly if you tell me that the Fourier's coefficients are blah, blah, blah, 3, 2 7, and 16. And if you tell me that you're working with a simple Fourier series periodic in 3, then you've told me everything. And so there's a way for me to backtrack that it was orthogonal. I am not sure if I'm connecting with you, so if I'm not, ask me after lecture to make sure that--

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** Sure.

**AUDIENCE:** I think he's saying if you have two signals [INAUDIBLE] coefficients [INAUDIBLE] two signals, can I tell if those two signals are orthogonal [INAUDIBLE] the coefficients are orthogonal [INAUDIBLE].

**PROFESSOR:** If they have components in common, they couldn't possibly be orthogonal. So I would answer yes to that question. So if that's what you were-- so I think that's probably right. Does that sound right? Yeah, OK. Yes?

**AUDIENCE:** I'm awfully confused by the complex conjugate, the [INAUDIBLE].

**PROFESSOR:** Yes, yes, yes.

**AUDIENCE:** So does that mean we're taking the complex conjugate [? of a ?] and we're [? applying it ?] to b?

**PROFESSOR:** We're taking the complex conjugate of the entire function. At every point in time, we take the complex conjugate of it. And it's especially useful to think about if you're doing something of the form-- if a of t were  $e$  to the  $j 2 \pi m t$  and if b of t were  $e$  to the  $j 2 \pi l t$ .

The only thing we're trying to do-- but this comes up quite frequently-- the only thing we're trying to do is when you conjugate one of these, you rig it so that when you add the exponents, the result goes to 0 by

putting the minus up there. That's all.

**AUDIENCE:**

It doesn't seem like we had to do any of that for the example we just worked on. It seems like there were just like [? signals ?] [INAUDIBLE].

**PROFESSOR:**

Oh, interesting. That's a very good point. That's interesting. So I didn't intend to throw you a ringer.

These were signals, all of these except that one, are real functions of time. That's why the complex conjugate didn't come up. So I apologize. I wasn't trying to make it seem tricky.

OK, so it's because this function of time is everywhere real that we didn't need to rehearse this. We did have to do it in that one. OK? OK, so the point is that we've already covered, even though we've only done a little bit of work in lecture, we've already covered all of the theory.

What remains though is to do some practice. And also what remains is to understand how this is useful. So it's not just music. The example that I want to talk about today is speech. The same sort of thing that we could do with music last time, we can do with speech. And here are some utterances.

[AUDIO PLAYBACK]

- Bat, bait, bet, beet, bit, bite, bought, boat,  
but, boot.

[END PLAYBACK]

**PROFESSOR:**

All right, it was just intended to be a bunch of sounds that we can analyze with Fourier analysis to get some insight into how to think about, in particular, speech recognition and speech synthesis. So we can take those utterances, and all I did was write a little Python program to do the decomposition that I showed on the previous slides, so that I could break these time waveforms. Here I'm illustrating one, two, three, four, five, six periods. So I took one period of that sound



and ran it through that kind of an integral to break it into Fourier components, which are showed here.

And what I want you to see is just like you could have recognized a pattern here, and you might try to recognize which vowel is which by the signature in time. An alternative, and far more useful way of thinking about it, is to try to recognize the pattern in frequency. So there are characteristic differences in the sounds, and we'll look at the basis for why there are. There are characteristic differences in the sound that can help us to identify automatically, by a machine, what was being said.

And so what we want to do is learn to think about a pattern that characterizes ah, ee, oo in the frequency domain, as opposed to the time domain.

[AUDIO PLAYBACK]

- Bat beet, boot.

[END PLAYBACK]

**PROFESSOR:**

So there's something different about those sounds that manifests a difference in this Fourier signature. So that's one of the useful applications of this. And we'd like to understand that better. There's a really good physical reason why that happens.

And it has to do with the way we produce speech. So you can think about speech as being generated by some source. Ultimately, the source of my speech is somewhere down here, which always amuses me when I see the cut off heads talking, like at Halloween and like on some cartoon shows. Because you can't do that, right? Because the source has to do with down here someplace, right?

My lungs push in. That pushes air through something and starts making noise somehow. I'm going to focus today on things that we call voiced-- in a voiced sound. Ah, in a voiced sound, it's caused by vibrations of the vocal chords. So if you were to stick a camera down

someone's throat, this is the sort of thing that you would see.

It's an enormously complex structure whose mechanics are extremely difficult to understand. Because what happens is when you want to make a high sound, you tense the structure. You pull on some muscles that pull the cords. The cords are normally rattling pretty fast.

And what you do is, you pull on a muscle that tenses them to make it go higher. But your intuition should say, now wait a minute-- you're making it longer to make it higher? And your intuition would be right. Normally, long organ pipes are higher or lower in frequency?

**AUDIENCE:**

Lower.

**PROFESSOR:**

Lower. So you have to do a lot of mental calculations in order to move these muscles correctly. So that the resulting frequency of the vibration comes out right. It's not obvious, because two things happen as you tense the muscle. The folds-- the vocal cords get longer which you would think would make the frequency lower, but they get tighter, which, of course, goes the other direction, right?

So it's a very complicated thing. And in fact, it's something that goes bad with professional speakers. But even more, professional singers often have a lot of trouble with the enormous stress that happens on this structure with repeated use and repeated overuse. Anyway, this takes a real beating. But that's ultimately the source of speech.

But if that were all you had, it wouldn't sound much like speech. A lot of the interesting stuff comes from these cavities that you intentionally manipulate as you're speaking to make the different characteristic sounds. So the idea then is that you have a source that contains information like frequency. What's the pitch of the utterance?

But you have this other thing, which is acting like a filter. If you think about the whole thing as a system, we have a block which represents a filter, which is the thing that has a frequency response. The frequency response depends on how I've put my tongue in my mouth and how

I've opened my lips and stuff like that. We'll see that in a minute. But it also depends on how the vocal folds-- it has an input, which is the vocal folds.

So the idea then is the same kind of a source filter idea that we motivated last time by way of the RC filter example. If you put a resistor and a capacitor together with a source, a convenient way to think about that is as a low pass filter. We think about it having a frequency response. So the system, just the RC part, has a frequency response which we can characterize by a Bode diagram.

So we can think about-- we did this last time-- so we can think about the low frequencies go through without attenuation. The gain is 1, and the phase is 0. So basically low frequencies go through the filter without any change. High frequencies are attenuated. The higher the frequency, the more the attenuation, and phase shifted by lagging  $\pi$  over 2. So that's a way of thinking about the RC circuit as a low pass filter.

And it gives us insight in the kinds of signals that go through and don't go through. So that, if we think about a signal like a square wave having a Fourier series decomposition, it only has odd components and the odd components fall with  $k$ . The magnitude of the component is inverse with  $k$ . So we get components that, if I plot on a log scale, the reciprocal relationship of the weight of the components, the magnitude of the components, makes it a straight line with a slope of minus 1.

And now we can think about putting this signal into the RC filter and thinking about what the output should look like. If the frequency of the square wave, if the fundamental frequency,  $2\pi$  over the period, if  $2\pi$  over capital T, if  $2\pi$  over capital T is some frequency that's low compared to the corner frequency of the low pass filter, basically the output of the filter, which is showed in green, overlaps the input, which is showed in red. You can't tell the difference because all the components have the same magnitude and phase as the input.

But if you change the frequency of the square wave so that the fundamental is higher, some of the higher frequencies are attenuated and phase shifted. The shape of the waveform, showed in green, starts to deviate. If you go to still higher frequencies, the deviation's even greater. And if you go to high enough frequencies, they're all in the region where the magnitude is being attenuated by whatever frequency. So my dependence of  $1$  over  $k$  becomes  $1$  over  $k$  squared, and it goes from being a square wave to a triangle wave.

So that's a way of thinking about the signal transformation in terms of a filter. We did that last time. What's going on with speech is exactly the same thing. What we want to do is think about-- the glottis makes some kind of a sound that goes into a filter.

The filter is this thing that is controlled by my tongue's position and my jaw position and my lip position and stuff like that. And what comes out is speech. To demonstrate that, here's a film that was made by Ken Stevens. Ken Stevens was a professor in this department. He just recently retired.

This was done when he was a graduate student. It's very hard to see because the contrast is not great. But you have to take into consideration this was made with X-rays. OK, we probably wouldn't do this today. It was a relatively large exposure to x-rays, which we sort of frown on these days.

Just so you're not too worried, Ken Stevens, when he retired, had the longest teaching career in our history. He was a lecturer who actively lectured for 50 years. So he seemed to have done OK. He survived. So you don't need to worry about what happened to him.

But we probably wouldn't repeat this. It's a little hard to see. The bone is easy, right, because x-rays don't go through bones very well. What you can just barely see is his lips.

And it's important to watch the lips, too. It's also important that his chin

is on a chin rest to simplify analysis. The idea of this was to get quantitative measurements to fit the source filter idea. OK, so now I'm going to play a film, a recording of him.

[VIDEO PLAYBACK]

- Test. Test. Test. [? The tongue. ?] [? The tongue. ?] The [INAUDIBLE]. The [INAUDIBLE]. [? The neck. ?] [INAUDIBLE]. [INAUDIBLE]. [INAUDIBLE]. [INAUDIBLE]. [INAUDIBLE]. [? Fox. ?] [? Clock. ?] The [INAUDIBLE]. The [INAUDIBLE]. [INAUDIBLE]. [? Took. ?] [? Two. ?] [INAUDIBLE]. [? Tot. ?] [? Tech. ?] [INAUDIBLE]. [INAUDIBLE]. [INAUDIBLE]. [INAUDIBLE]. [INAUDIBLE]. [INAUDIBLE].

Why did [INAUDIBLE] set the [INAUDIBLE] on top of his desk? I have put [? blood under ?] [? two clean ?] [? yellow shoes. ?]

[END VIDEO PLAYBACK]

[LAUGHTER]

**PROFESSOR:**

OK, so what you were supposed to see is that the thing that we associate with speech is only a small part of it. His lips were obviously moving. That's what we see.

But if you were paying attention, his tongue was going up and down not a little bit, but a lot. So the gap between his tongue and the roof of his mouth was going from 0 to about that far. The velum back here was opening very broadly on occasion. So there was a significant variation in the shape of the structure through which the glottis wave form was passing. And that's the basis of the filtering that gives rise to the different speech sounds.

So to convince you of that, here I have a carefully machined item. Let's see. I don't want this one. I want this one.

So this is a Japanese oo. OK, now I don't know Japanese, so I have to just sort of trust the guy who made this that it actually sounds like a Japanese oo. The second one I'll show is a Japanese ee, which actually sounds more like an ee to me.

But anyway, this model was made from measurements of the type that I just showed with Ken. So the idea was to estimate the size of those cavities through which the air was passing, and then make, by machining in Plexiglas, a structure that has that shape. So this was an early test of whether the source filter idea works.

So if that is the explanation for how speech is generated, then I ought to be able to take a boring sound of the type that's generated by the glottis--

[BUZZING SOUND]

And put it through this, and it should sound more like a vowel. OK? Got it? Know what I'm talking about? So this is a Japanese oo.

[BUZZING SOUND]

[COMBINES BUZZING SOUND WITH 'OO' SOUND]

I don't know if anybody knows Japanese. I don't know if that sounds like an oo. Does anybody know Japanese, and does that sound like an oo or not? OK, I'll pass.

[BUZZING SOUND]

This is an ee. OK, now notice that the ee looks very different. Right? The question is whether that's a big enough difference to make the difference between an oo and an ee. I'm pressing the same button, nothing up my sleeve, nothing-- OK, so same button.

[COMBINES BUZZING SOUND WITH 'EE' SOUND]

OK, so what you're supposed to be convinced of is there is enough information in the shape of the vocal structures to account for the difference in the sounds. Now of course, we don't really care about the acoustics if we're trying to, for example, synthesize or analyze speech. We don't particularly care about that.

We do like to know that there is a theory that underlies it, right? And there's a very sound physical basis for why we should think about the source filter idea. When I say source filter, source filter-- so everybody calls it the source filter model of speech. So is there any good physical reason for why that should be true?

Of course, what we care about is the frequency response. So here what's showed is measurements of frequency responses taken from speakers. So now we don't do the x-ray thing. All we do is record somebody saying heed, had, hood, haw'd, who'd.

And we look at men, women, and children, and we characterize how their frequency responses change when they make those different sounds. So what's showed here is that you get a relatively good fit by thinking about the frequency response having three formants. The formants are the peak frequencies.

There's a theory, which I won't go into, for how you take this shape and turn it into a formant frequency. And given just the formant frequencies, or given the frequency response measured at uniform spacing across frequencies, there is a theory for how you can generate the smooth line, which is really-- this is an 11th order [? fit, ?] which means that there are 11 poles and no zeros. So what you do to get this shape then is take the locations and amplitudes of the formant frequencies and do a fit using poles.

And so here's a table showing measured formant frequencies, F1, F2, and F3, for whatever, six different sounds for three different categories of speakers. OK? And that's kind of a complete analysis then in terms of the source filter idea. So this figure summarizes the idea.

We think about source filters. So the source is the glottis. The filter is the formants created by the throat. And speech is the thing that comes out of the source filter. The source is some periodic waveform caused by the banging together of the vocal folds. The filter is the frequency response of the throat.

And the result, then, is just passing this glottis wave form-- so this is a measured-- by sticking a microphone in somebody's throat, this is a measurement of what the glottis acoustics looks like. This is a Fourier decomposition of that periodic waveform. Then this is the frequency response of that thing. And this is the Fourier coefficients of the output for different sounds.

So here is the frequency. So the same glottis signal underlies and ee sound and ah sound and generates two different spectra. We call that combination of magnitudes and angles. We call that the Fourier spectrum. So you get two different spectra, depending on the filter shape. OK?

And that's the basis-- this theory, this source filter idea-- is the basis of the current technology for speech recognition and speech production. So I actually cheated. Those sounds that I played earlier-- bit, bat, bought, beat, all those things-- those were actually synthetic speech.

OK, all I did is I ran a speech synthesizer, and I said, synthesize bit. So that was really a synthesized thing. That was not a real person.

And so the synthesizer used this theory in order to generate this synthetic speech. We also use this theory in order to recognize speech. And you'll do a homework problem in Homework 10, I think it is, in which you'll build the primitive front end of a speech recognizer using this theory. And I'll give you a couple of utterances of different vowels and you'll have to classify which vowel is being said, according to some automatic speech recognizer based on this theory.

The theory is also just fun, because a theory lets us figure out



anomalies. So when somebody has a speech impediment, for example, when I did, when I was a little kid I did. And I was sent to speech school. Now they do a much better job because they do analysis to figure out what I'm doing wrong, using this sort of source filter idea.

We can also use the source filter idea to understand paradoxes. So for example, I've told you before I work on hearing aids. I tried to make hearing aids hear. And so people with hearing deficiencies like mine, where I have sort of progressive age-related because I'm old, right, that's what happens.

I have age-related hearing loss, which means that I'm losing high frequencies. I'm less sensitive to high frequencies. People like me, which is the vast majority of people my age-- it's easier to understand male speech than female speech. Why? Higher frequencies.

So those higher frequencies shift some of the important stuff that I should be listening to into frequencies I don't hear anymore. Right? So that's a way of using this theory to try to understand what's wrong with me. But there's also things-- it's not just me. Normal people have trouble distinguishing female speech, especially in taxing environments, and one of those is singing.

So if you consider altos and sopranos, sopranos are like, the worst, right? Because they are not only female, but they're at the high end the females. And there are those who complain about not being able to understand female singers. OK, so here's a demo that will help us to understand whether that's a valid kind of a criticism or not.

So what I've got is a professional singer singing, la, la, la, la-- on a scale. So from low frequency to high frequency, then a different sound, a different sound, a different sound, a different sound. So the first thing that I want to do-- I want you to listen to those different sounds as she goes across the scale. Then I'm going to play just the low ones and just the high ones, just the low frequency ones and just the high

frequency ones. So first, the different scales-- la, la, lore, loo, lee, OK?

[AUDIO PLAYBACK]

- La, la, la, la, la, la, la, la, la, la, la, la, la,  
la, la, la. La. Lore, lore, lore, lore, lore,  
lore, lore, lore, lore, lore, lore, lore, lore,  
lore, lore, lore, lore. Lore. Loo, loo, loo,  
loo, loo, loo, loo, loo, loo, loo, loo, loo, loo,  
loo, loo, loo. Loo. Ler, ler, ler, ler, ler, ler,  
ler, ler, ler, ler, ler, ler, ler, ler, ler, ler. Ler.  
Lee, lee, lee, lee, lee, lee, lee, lee, lee, lee,  
lee, lee, lee, lee, lee, lee. Lee.

[END PLAYBACK]

**PROFESSOR:**

OK, so now what I've done is I've sliced out the lowest frequency, the very first of the scale from each of the sounds and pasted them together to get the low frequency run. And then I took out the high ones and pasted them together. OK? Exactly the same sounds, just played in a different order. So first the low frequency ones.

**SINGER:**

La. Lore. Loo. Ler. Lee. And the high frequency ones. La. Lore. Loo. Ler. Lee.

[LAUGHTER]

**PROFESSOR:**

It's not her fault. She's doing everything right. And you can see that.

Here is, again, a Python program analyzing those same segments. So what's shown here is the ee, the filter derived from the ee, by thinking about the lee, lee, lee, lee, lee, lee-- by looking at that sequence, and averaging across the frequencies. So here's the filter.

Here's the filtered glottis spectrum for a low frequency and intermediate frequency and high frequency. What's the difference between the low, middle, and high? What's characteristically different at low and high?

**AUDIENCE:**

[INAUDIBLE] frequency, like high amplitude.

**PROFESSOR:**

So if you look at the low frequency, the low pitch, there are more frequency components in a given range. So if I say analyze the frequencies between 0 and 1,000 hertz, 1,000 cycles per second, there are more lines when you have a low frequency. And so you get the density of the lines is greater for the low frequency utterance than it is for the high frequency utterance.

The low frequency utterance is spaced close enough that you can clearly figure out this pattern from that spacing, because there are multiple lines per peak. The problem is that the speech waveforms have very sharp resonances. The peaks are narrow.

So that as you go to a higher frequency, now it's very hard to see. So where there was two lines characterizing this guy, now there's one. And at the highest frequency, there's nothing there.

Similarly with these peaks, again, several lines representing each peak. One line representing-- nothing representing this peak, nothing representing that peak. There is nothing about ee in that signal.

And if you do the same analysis for ah, you get the same result. There's nothing about ee, and there's nothing about ah. There's just nothing there. There's no way anybody is going to tell those two sounds apart.

So if the singer put her voice, her vocal tract, in precisely the right location, there would be no difference between those sounds, OK, regardless of what the director said. OK, so that's the problem. So that's a way of using the Fourier analysis to gain some insight into some anomalous situations. Yeah?

**AUDIENCE:**

Does this have more to do with the rate at which you're sampling?

**PROFESSOR:**

No. It has to do only with the frequency content of the glottis waveform. You can think about it as sampling. And that's a good insight, because

the Fourier series only has components at integer multiples of a base frequency. So that means we're sampling in frequency, not in time.

So we have this potentially continuous frequency response, which is characterizing this. That is continuous. I could excite this at any frequency that I want to. But the glottis waveform of the singer is only sampling that at particular frequencies-- C, C prime, C double prime, B, B prime, B double prime, right? So there's only certain frequencies at which the singer is sampling this.

So there is a way of thinking about it as sampling. But it's not sampling due to my A/D converter or anything like that. It's sampling in frequency. So the point is that this kind of a source filter idea, and more generally, the filter idea, is such a powerful representation that next time we'll think about how to do the same sort of thing for non-periodic [? stimuli. ?] See you then.