[SQUEAKING]

[RUSTLING]

[CLICKING]

MICHAEL CUTHBERT: I didn't get any new topics that people wanted throughout the day, so I'm going to go through some topics that people have asked about earlier in this semester and that I think are kind of fun. And they're sort of the most fun things that, from my perspective-- maybe I have a sixth sense of fun-- that we didn't fully cover.

So one of the first things I want to talk about is a cool little two-part-- what do you call this-- a fable, a moral story, that has a little conclusion at the end. And something called gap-fill. I'm pretty sure I've talked about this a couple of times in different places. But I'm pretty sure I did not do this semester, did I? No? Good, good.

Because there's three points in the semester where this could have gone. One is anytime. We first started talking about David Huron's work. And this is one of the very cool papers. He's the second author. Paul Von Hippel, a very, very well established scholar at music perception and cognition, did this study in around fall 2000, published in fall 2000. So it's about 25 years since they did this. And it's one of these really cool things that got me into this field. So when I heard about this, I was like, this is why I want to use computers to do things.

So has anybody ever-- you all have some encounter with species counterpoint because you read the David Lewin article on an interesting global rule for counterpoint. So it's dealing with 16th century musical style and how we can recreate it.

And one of the rules that I was taught when I learned species counterpoint is that if you take a big leap and you go up, you've created a gap. And the most natural thing for your musical line to do is to go down after that. Fill in the gap. Gap-fill. Some people spell it as one word. Some people use a hyphen. Some use a space. I don't really care. Gap-fill.

Similarly, you leap down a lot, and you should go up. And so I immediately looked at a piece by Giovanni Palestrina, one of the most important composers in this period that species counterpoint is based on, around 1550, 1570 or so. This is part of one of his masses.

And here's the score, and you see it. But I've also kind of colored things in, that green notes are upward leaps that need a downward thing after it, so things. And then the blue note is a downward leap that needs an upward motion. And all around here it's all stepwise, so it can go any direction it wants.

So if gap-fill exists, and every time there's gap-fill fulfilled, every time it works, there's a little pink or peach dot. And every time there's not, there's a big yellow dot. And you can see there are no yellow dots up there. So this is what this sounds like.

[NOTES PLAYING]

If I just do that one more time so you can hear it.

[NOTES PLAYING]

I just had to add the last note. Otherwise, it'd be kind out there. So gap-fill seems to be a thing. Composers do this.

But the really cool thing that Von Hippel and Huron did was that they realized that there's two things happening when you leap up a lot. One is you've created a big leap. The other is that you're higher in your voice range.

So if I'm down here and I leap up and then I leap up, then I leap up, you guys are going to bet that eventually I'm going to come down, right? Yeah. You can't really go up forever.

And so what they were wondering is, does it have to do with the fact that eventually when you get really high, the only place to go is down, especially if you're talking about vocal music? Well, I mean, even on a piano, eventually you've got to come down.

So they looked at what the median, what the average pitch across the whole piece was and marked it as specific. And then thought that, hey, there's four kinds of leaps. And there's ones that start at or above the median and go even higher. That's what I was doing this. And I went, yeah, up there.

Then there's ones that start below the median and go above it. There's ones that start below and reach the median. And then, like I started, the ones that start below, and after they leap up, they're still pretty low. So they're still below the median.

And there might be something besides gap-fill that explains things. For instance, let's say you bought two domestic continental US flights. First, you flew from San Francisco to New York City. Now you have another flight.

Just take a guess. You've gone from-- instead of up and down, we'll say west to east. You've gone from west to a big thing on the east. And now you board, let's say, a random flight after this. It's pre-9/11. You could just board random flights. You all don't remember that world.

So which direction do you think you're likely to go? You're not in the international section. West or east?

AUDIENCE: West.

MICHAEL CUTHBERT: West, West. Why? Because there's so many more options of going west from New York than there are from-- yeah, then there are east. I guess that's a little too far for Portland. But yeah, you can get the picture.

And so if you're already above the median, you have a lot more choices going down than up. And what they predicted was that a lot of these things would happen even by chance, that if you went-- if you jumped up above from a normal note and you do a big jump up, you're probably above the median. And you're going to gap-fill just by the concept of regression to the mean.

And so I went and took that exact same piece from Palestrina. I guess I've been doing the same thing I did with A, B, C, D, E, F for a long time of just taking pieces and scrambling them. Remember when we were doing the similarity one, I just scrambled all the notes? So here's the exact same melody, scrambled.

[NOTES PLAYING]

And again, everything that demands by the theory of gap-fill motion in the opposite direction than where it just came, I've marked with a green or blue. And I've marked with peach when it works. And I've marked with a big yellow dot when it doesn't.

And in this scrambled-- and this was the first random run I did-- I didn't run it a whole bunch of times until I got one that worked-- it works perfectly.

What they argued was that this concept of gap-fill as something universal in music could actually just be explained by regression to the mean. And I thought, this is really, really cool.

The thing I noticed, looking at this article for the first time in the mid-- what do you call them, the aughts-- the 2000 to 2009, something-- was looking at the data that they had to do this. And so they had some songs, lieder, by Schubert. They had European folk songs. Where do you think that came from? Anybody? European folk songs and Chinese folk songs. That's always-- that's the Essen data set. So they had-- this data set's been around for a while. And they had some other things, South African and Ojibwe, Native American folk songs.

And you can see when you're going all the-- you're already high and you go even higher, then the odds are you're going to turn down when you're here, you're here, you're going to come back down. And so that's very statistically significant.

And when you start in a normal range and end in a high range, you're still probably going to go down. And they saw that a lot. This is not just down. This is other direction. But I'll assume for a second that everything is leaping up.

But then when you get right to the median, you see that you go about 50-50 in either direction. And when you start low and you go up but you're still low, more often than not, you just continue going up. There is no gap-fill. Or it's kind of reverse gap-fill, or something like that.

So I'm somebody developing music21, and I kind of look at these numbers, which are probably from 1998 or something. And now it's 2010, 2012. I'm teaching here, doing this type of stuff. And I look at their sample size. They have 86 examples of something like this and 101 of something else because it took a long time to encode music then. It still takes a long time. But we've gotten things better. There's more data sets out there.

So I thought about running it for-- I think it was originally to think of as a problem set for this class, to let's validate Von Hippel and Huron's work. But it turned into a little bit more. And it turned into a paper that my friend, Mark, who's probably going to watch this at some point. We do still need to finish publishing it. We've presented it many times, but not published. "Species or Specious?" on various rules for species counterpoint that don't seem to actually exist in the literature.

Because when we ran the same thing, you can see we have a lot more data examples. And when we looked at the Chinese folk songs, we also saw the same thing that Von Hippel and Huron did, that there's a lot more evidence for R to the mean, regression to the mean, than there is for gap-fill.

And when we looked at the European Essen data set, we still saw the same thing in all of these things. So I got to do the nice validation study that I thought we would do and I thought would make a nice assignment for us all and everything.

Until I looked at music before Bach, the music of the era that we do species counterpoint on, why we have species counterpoint exercises. And it turned out there that for these weird pieces where you're jumping up above the mean or something like that, gap-fill happens more often than regression to the mean.

And so for these low going to here, and then you would expect regression to mean, you should still go up, there did seem to be a period where people were actually doing gap-fill. And so that was pretty neat. And so it just shows that sometimes when you get a new data set, you should be re-examining existing theories.

But Von Hippel and Huron did cover their butts. At the end, they said that their conclusion-- and I agree with them completely-- so they're actually and I are actually in agreement on this-- that their conclusion applied only to the sampled vocal repertories that they were looking at, that other repertories may contain an excess of post-skip reversals, gap-fill. Given the centuries of Western composers who have been taught to write gap-fill, it seems inevitable that some of them have actually done it in their compositions.

And now we can see what errors composers were actually doing it. Though I think it's kind of interesting that the era that started teaching how to write music with gap-fill is closer to this period to Bach, Mozart, and everything, talking about how music of that past worked. So they weren't actually doing it in their own music, but they had observed something in the past. And it just shows for the 10,000,000th time, when we think we're smarter than people of the past, we're quite often wrong.

Any questions on gap-fill or this kind of work or discussion points? Doesn't have to be a question. Cool. Eh. Do it again soon, somewhat. Something in between. Yeah?

AUDIENCE: Can you talk a little bit through the numbers? What exactly does-- how do you measure significance exactly?

MICHAEL CUTHBERT: Yeah, I can't remember. This is why I need a little bit of time to come back to this, how we did this. I think it was the scholar's t-test. Do you know what that is? What the significance of it? Basically, how far is this difference between these two things? How far is it compared to random chance?

So one of the things that-- it doesn't work. This isn't a statistics class, and I am not a professional statistician. But one of the ways that you can often test something is if your null hypothesis, if your hypothesis going into experiment, is that gap-fill doesn't exist, that there shouldn't be a difference between gap-fill and a random motion or motion toward the median to regression-- I should say regression to median, I suppose, than to mean, but we don't know that.

And so one of the things is that significance goes-- the amount of difference between two hypotheses necessary to say that something is highly significant goes up. The amount of difference in percentage goes down a lot as the number of observations goes way, way up.

And one of the places I saw this on was on the article I wrote with Sophia's son on the lyrics emphasis. Are their high notes on anticipation or surprise? When we're talking about, oh, anticipation is higher than disgust, we're talking about an average. I think those two were about one semitone, the difference between F and F-sharp.

This is not a huge difference. But when you get 80,000-- or I suppose that one was about 10,000 observations-- one semitone starts becoming pretty significant. So that's one of the ways that sometimes if you're just looking at 10 pieces, you might see that well, six of them do this and four of them do the other thing. That's probably not going to be a significant difference, because if you flip a coin 10 times, you might get six heads and four tails or something. Yeah?

AUDIENCE: The numbers for gap-fill and regression to, I guess, the median, those are instances where this happens in the pieces of that specific?

MICHAEL CUTHBERT: Yes.

AUDIENCE: Sometimes can it do both, I guess? If it starts above or at the median, it will technically be both if it just goes down, right?

MICHAEL CUTHBERT: Yes.

AUDIENCE: So you can have-- but some of them are one or the other?

MICHAEL CUTHBERT: Right. And so the ones that cross over, I think it's only the ones that reach to the median or that don't reach it that end up being in the observation set.

AUDIENCE: OK.

MICHAEL CUTHBERT: Because all the rest should always be the-- should always be the same. They should be the same. That is to say, if you start-- if you start middle or start middle high and go really high and then you go even higher, well, then you've broken gap-fill and broken regression to the mean. And if you go down, then you fulfilled gap-fill and regression to the mean, so you haven't learned anything about your hypothesis.

So I think these were only various-- these particular cases where the two hypotheses would result in different directions predicted. Yeah, great question. Great question. Now we're getting on. I just need to give more time. Any other things?

This isn't a statistics class, so I'm not going to go into things like this. But one of the mistakes that people who dabble in statistics but whose main work is on medieval music, music theory, programming, all these other things we often get wrong is something called a multiple comparison test.

And that is to say, the more-- if I looked at 7,000 repertories, and I'm only reporting the top 10 or the 10 best or something like that, then my significance level needs to be much, much, much higher. And a general rule is if-- who's heard of p-value 0.05 or 5% as a standard thing?

So if you look at five different things and five different hypotheses, start with a p-value instead of 0.05. Divide that by 5 is 0.01. And so this was a criticism on some earlier work that I did where I was looking at-- what is it-- 10 different emotions or sentiments that can be compared with nine other ones. So I'm doing 90 comparisons for each attribute.

So I should be whatever 0.05 divided by 90 is, somewhere around 0.001. We just looked at the number of zeros compared to the number of hypotheses. Yeah, so look out for things where people are reporting p 0.046 as significant because it's lower than 0.05 but not very much.

But on the other hand, when you get something like e to the negative 45, you have 45 zeros before some digit that I don't even care what the digit is. There's 45 zeros, and that's what's important. That ends up being pretty good.

I like that the European folk song we can actually do even better than-- they were giving two stars for their work because they only had this many observations. But now we can give them a lot more observations and say, yeah, this is absolutely right. Once you get in 64-bit numbers, once you get more than 308 zeros, it just gets rounded to 0, which ends up being a lot of 0 division errors. Cool.

Now I'm going to move on to the second topic, which is a little bit briefly. And it's something I feel-- I rightly should feel apologetic for. We're moving into an era where people here in New England, in Boston, at MIT, in the United States, are recognizing, finally, that Western Europe is not the only place that does music or culture. I know. Big, big, big shock.

So why has this class been primarily about Western music? In fact, we can start looking at other music as part of a music theory. So there was just earlier this semester-- I think I emailed it to somebody. It might have been Vincent, even, who just came in, maybe Sruthi.

A new book that just came out. I think it's like 600 pages, all on computational approaches to Indian art music-- that's to say music that's in a high tradition that is not folk music or modern popular music. And if you think about-- or Bollywood, or something like that, if you think about it, you can write 600 pages on Indian art music. You could probably write 600 pages on computational approaches to any of those things, too. And I'm really looking forward to seeing what many of my colleagues down in South Asia are doing.

But one of the things that comes up here-- oh, no, I didn't highlight this-- is that most of my training, my personal blinders, I'll say bias-- although maybe it is also that-- has been that I was always interested in musical scores. And now there's large parts of music theory that are interested in things that are not score-based.

And most of the rest of the world's music is not primarily score-based, just as, shockingly, most Western music today is not primarily score-based. It is really, really great when you go see your favorite heavy metal band, and they're all page-turning for each other. So no, that doesn't happen. We're much more of an oral culture today.

So it says basically that there are no standardized tools yet such as humdrum or music 21 for Indian art music scores that's still going to need collections of scores, whatever it is, if we're going to be using these tools.

And I want us to always be trying to break out of the chicken and egg problem that I've done this semester with-- we'll do another Bach chorale. I'm really-- my apology to you all, everybody, is that I wanted to de-emphasize more of those things. And music21 version 9 has a bunch of other substitute scores that are designed for us to be using. And it just has continued to be too buggy to recommend for us all to use in this class-- not the scores, the surrounding code.

So we don't want the chicken and egg problem that, well, we don't study this music because there aren't enough scores in there. Well, why don't we encode a whole bunch of scores? Well, if we did that, there's no software to analyze it. Right? So these are some of the problems that are there.

So they're absolutely right that these tools, the authors of this book, editors, that these tools don't yet exist. But we wanted to, those of us who were building music21 first and many of us in scores. We just never got there.

This is one of the slides-- I know I haven't changed my font in a long time-- from the original funding presentation to get funding to take music21 from a little hobby Europe project and to make it a bigger one. And we talked about these are probably things-- doesn't exactly work like this, but these were the types of things music21 was going to be able to do-- generic intervals, specific, chromatic, isStep. You all have worked with this by now.

But then I thought, well, you should be able to apply a context to any object and transform how that object works. So you could apply 13th century France cons-- con-- oh, I didn't even make this consistent. This is all pseudocode.

In 13th century France, this major third would not be called a consonance, only perfect intervals. That's why they're called perfect or consonances. In fact, this lower note wouldn't even exist. There is no A-flat back then, so we could do that. And then we could apply other things like 20th century harmony, applying octatonic scale and count the number of notes in this. And we get three instead of two, and that one we can do.

But I did think about, oh, wouldn't it be great to be able to take this interval, which we started with calling it a major third, and convert it into the Indian swara and get 7 shruti as the interval? That these types of things should be able to be done.

And a lot of the basic design decisions, some of the things that are overcomplicated in what you're trying to do, were in order to enable more of these things. Or that we can apply Okinawan pentatonic scale and say that A flat to C is a step. It is not a skip. It's a leap because there is no note between those.

Or one of the last things in there, that one of the things music21 will be able to answer-- I suppose I never put a timeline on that, so maybe I'm still telling the truth-- is that West African cultures often hear the same rhythmic pattern as having different strong beats, different parts of the bell pattern. what can we learn from analyzing more music to try to figure out what clues give a different sense of strong beat among different listeners? So these are all ideas that computational music theory should be able to do.

So we wanted to get there originally. A whole field has wanted to get there to be able to do some of these things. But we're not there yet to really have cross-cultural symbolic music analysis.

But I want to keep trying. I'm going to be spending-- this summer coming up for me, I'm going to be trying-- this has really inspired me to make sure that at least one more of these things are possible. Cool.

So that's one more, the second of three things I wanted to talk about today. Discussion? Ideas? Anybody know of anything very cool happening in this world? Oh, I meant to talk about.

There is a project, Tarsos, that is specifically a music theory toolkit for looking at music throughout the world. It's a little bit older. It's one of these things that you need to find your Java virtual machine to work properly on.

But it's a very cool set of tools for analyzing especially scales and scale usage throughout the world that doesn't rely on saying oh, well, that note's kind of like an A, and that note's a C-sharp but 20 sets high. No, no, those are notes. These are not alterations on some platonic, correct note that just happens to be the Western scale.

So I think any conceptual system for non-Western music needs to start with anything that is fundamental in that system is a fundamental object, that maybe we can translate them to other systems, objects like Western music. But first, the manipulation should happen in the space that people are actually conceiving. I know. It's very bobby, lecturey today for me, but I don't want you to have to do too much more work while you're doing many other things.

So I'm going to talk about the last really cool area that I think there's still so much for music theory to be doing in order to get us more scores. And that is improving optical music recognition. So optical music recognition, the process of changing a musical score into something that a computer can understand, a symbolic score.

And this is done-- why is my name bigger on this? I think we reset the-- that did not supposed to happen. I think there was something-- I think it was in an autoshrink mode. Sorry about that. Sorry, Maura. She's doing amazing things out there at Duolingo.

One of the opportunities is-- and this is from a presentation-- I'm reusing slides-- I don't update them every year. --from 2014 that back when IMSLP had about 300,000 scores. I think it's over double that now. And so we'd love to be able to convert all of them into music.

So I always might as well start with "Eine kleine Nachtmusik" by Wolfgang Amadeus Mozart, late 19th century. And so if we want to convert it over, this is use from the state of the art in 2014. It's gotten better, but not a huge amount better.

So this is the original score of one part. And this is what the optical music recognition score came up with. So you can see-- oops, we've lost some notes here. Somehow we have another rest here and a whole bunch of dots at the top.

And on the other hand, a lot of it looks kind of, kind of right. But how do these errors add up? Especially rhythmic errors and misidentified part names, clefts, can lead to big differences in perceptions.

So I'm just going to remind us for the first couple of bars of what "Eine kleine Nachtmusik" should sound like. Actually, it shouldn't sound like this. This is an unretouched finale performance from musicXML.

["EINE KLEINE NACHTMUSIK" PLAYING]

So did y'all recognize it? Good, so I don't need to play more. Let's listen to a little bit of what happens when you play back the state-of-the-art optical music recognition.

["EINE KLEINE NACHTMUSIK" PLAYING]

I have no idea how it got what instruments should be playing. Anyhow.

So one of the things I've been talking about, especially since our sound to score lecture-- we didn't really do an image to score lecture-- is how music theory knowledge can work with inaccuracies in OMR.

Fortunately, some other people have been thinking about it. And this is one-- I should be citing my sources-- I can't remember exactly where it came from-- one pipeline for how scanned music can end up as a music encoded data file-- staff line identification, object location, and so on.

But just like we saw in the sound to score things, musical semantics and musical knowledge can be used to go back and forth in a two-way pipeline. And that was the idea that almost everybody knew. But very few musical semantics have been brought in to what we're doing.

So here's some things, areas that music theory can fix. You might have a pitch out of context and you want to figure out what we have misread. Probably that was intended to be a sharp. And now it's been read as a natural.

Instrument or clef misidentification. This is something that you don't need very much programming to figure out something's wrong. The correction isn't necessarily the easiest thing. Should it be change that to a tuba, or change the clef to a treble clef? Not sure, but we can definitely point that out if we have music theory knowledge to point something in.

And then metrical inconsistencies is what we focused on because metrical inconsistencies have a really, really nice thing that pitch inconsistencies don't. There's no answer key to whether you do it right or wrong, but there is a checksum to see if you've probably got it right or probably got it wrong. And that is the time signature. When everything that you've recognized does not add up to the time signature-- maybe it's a pickup bar-- maybe it's some other weird thing-- but most of the time, you've gotten something wrong.

So we want to talk about the rhythm replacement. And what we're doing all begins with the assumption that context will give you an idea of the correct rhythm. So here's that measure that was screwed up. And here, it becomes very obvious that we have a dot that's been added, given this context.

But now we're going to come up with a different conclusion. Same middle measure. But now what's wrong is almost-- is very, very likely that we've lost a flag on the first eighth note. So we have the assumption that the context will give an idea about the correct rhythm. And that requires a certain amount of rhythmic repetition and metrical stability throughout a piece or across parts.

There's some 20th century music-- John Cage, Pierre Boulez, Ruth Crawford Seeger, Gregorian chant-- that can't do it. But most of the music doesn't behave like that. Most music works like "Eine kleine Nachtmusik."

So we're just stripping out all of the melody and just giving the computer a score that looks kind of like this. There'd be a separate problem if there's a note deleted to try to figure out what pitch to put in. But we're trying to figure out, well, at least what rhythm is. So this is "Eine kleine Nachtmusik," the opening with just the rhythm.

So the first thing we looked at was distance. The probability measure offers a solution based on its location in the score, something called distance probability. So distance can be part-wise or vertical or horizontal. What do I mean by vertical probability measure?

So we're just going to look at this third measure at the top of "Eine kleine Nachtmusik" where there's an added-- well, is something doesn't add up here. The computer doesn't yet know what you all know, that there's a dot that's been added, dotted eighth note. But it could be anything. So-- yeah, there it is.

So one of the things to do is to look at the rhythms in all the other parts at the same time. What we can do is build a histogram for the violin. How often across the whole piece-- now again, this is being built based on the incorrect, possibly incorrect, OMR. So you're learning from error-prone scores to build other error-prone scores, perhaps.

But you hope that this is just a little bit close to reality. So the violin, violin I, mostly has the same rhythm or has the closest to the same rhythm as violin II and the cello. But for the most part, it's doing its own thing. When we look compared to the viola, you can see, hey, when the viola is wrong, 80% of the time, if that measure is like the other measures, the cello might have the same rhythm.

So that gives the probability based on horizontal or vertical. And then the other thing you can do is look left and look right to see if that same rhythm is somewhere over there. In this case, it's not because it should have been. But there's another mistake in the first measure.

And we can look throughout the whole score, not just-- what's cool about the horizontal thing is that scores tend to be wider than they are tall. In other words, more measures than there are parts, so you have more things to do.

And we can look-- I'll skip over the violin-- I think it's more interesting to look at the cello-- that obviously each measure is always exactly identical to itself. So you have a probability of 1 it's going to be. But then you see in a lot of the classical things, the most often similar measure is two measures on either side, not one, because of this four-bar structure.

And the other thing that I learned from this is anybody know sonata form? Exposition, development, recapitulation. At some point later, the music comes back. And so you can see this temporary spike in that if you can't get it right, look 97 measures later, and you might also get it right.

So the computer can learn just from the scores what all the probabilities for this particular piece are and then hopefully apply them to put it in. So now we know what measures-- what the probability of any measure having the same rhythm would be ahead of time.

But then we need to look at a second probability. What's the probability that that measure, that each other measure, might have the solution based on how often certain errors happen? And here's a chart that I think we need to keep looking at and keep judging. But it was the only one that I knew of at the time where somebody looked at what the probability that any given musical object would be recognized properly.

So 97%, 99% some of them. I don't believe anything that says 100.00. That means not enough data, right, but a lot of things. So we can look at various probabilities. Yeah, natural-- that's quite, quite, quite low.

So then this is where there should have been some gigantic formula with regressions. But it turns out all you need to do is find the probability of each measure being the same. Find the probability that how many changes you would have to make. And multiply those two probabilities and take the max value. And yeah, it is that simple.

So the workflow. Normally take a PDF. Turn it into musicXML. We kind of stuck ourselves into in the middle. Run it through one more time. I would love to be able to send it all the way back into the OMR pipeline and say, no, reject that. Now rerun the recognition with this being a possibility, but didn't get to that.

Then we had a whole bunch of scores where we had-- not as many as I would have liked, but we had scanned optical music recognition, musicXML. And we had the correctly encoded one. And we can look at the number of differences before, run through the model, and then look at the new number of differences. All of this code is in music21, in case you want to use it to improve things. Everything works. And the music we looked at-- again, too many canonical composers.

They average about 19% decrease in error, which made us very happy in a field where usually you're trying to improve by 1% at a time or something. And Maura Church earned the best grad student paper award, which was great because she wasn't technically qualified because she was still an undergrad.

Yeah, so it worked pretty well here. And some of the pieces got as high as 36% error reduction. So what happens? You have all original score. We've marked what's the errors. We correct that, and then post-correction. It's correct if we got the same rhythm, not necessarily the right pitch because we didn't work any time on that.

So these are all the places to be looking at. Here's what the optical music recognition software got wrong. And the correction. So it got, in this case, three out of five corrected. You can see them all there.

But what really came out of this work for me that seemed very important was that we are evaluating the quality of our work a little bit wrong. So there's that we need to be looking at optical music recognition that allows for musical semantics to go back and forth into things. That was a possible implementation for how you can make it faster to do things.

So there's proposals for judging how good optical music recognition software is that you feed in a bunch of scores and you see, how well does it do? But some of the standard scores being used are ones that if we had to run our software through would run terribly on.

So this is one of the ones that you should be able to recognize. But this particular score has no motivic repetition. I've never even played it. Let me play it. Just hear how unusual that score is.

[NOTES PLAYING]

So it doesn't really behave like a normal score. And there's even things like if my optical music recognition saw that C-sharp when it's just recognized this key signature, I would say you probably got it wrong. That's probably a C-natural. Or if you've got your key signature wrong because there's no reason why you would write a sharp there immediately.

And there are other things that are in the testbeds that I still think we need to do, like being able to recognize properly this measure. And I get that it's trying to test, can you recognize a whole bunch of rests? But if I saw this from a regular piece of music, I would say those are probably notes in here, right? Because there's no reason to not just write every measure like this.

And then other things that you get really smart and you could say, OK, well, that's a pedal mark, and that's a clef mark. I don't know a single alto clef instrument that uses one staff and has pedals. So I would send it back, and say, you got to change one of these things in your recognition. So I think in our quest to try to give simple tests that can be run very quickly, we're ending up optimizing our models for things that don't resemble the music that most of us learn.

Yes, so there's still a ton of work to be done on music theory and semantics in optical music recognition choices. And this is something that's getting a little bit harder now with the rise of most people who are trying to work in this field now obviously want to use machine learning, artificial intelligence, and deep learning models, which will be great once we get more ground truths encoded and we can do things right now.

That's the hardest part of having more things lined up so that the computer can learn. But it's going to be harder and harder as long as we're just using black boxes to be able to get this into there.

You all have actually met the world's expert in doing this. His name is Gus Xia. He was here when we moved to another place. So there are people who are trying to figure out a way to add-- not just in optical music recognition but in any of artificial intelligence working with scores-- how that we can still use these big deep learning models but incorporate music theory into that. So that's the big challenge for our time.

Questions on this? Yeah?

AUDIENCE: What do you think the state of using deep learning, end-to-end deep learning, for optical music recognition is?

MICHAEL CUTHBERT: Yeah, well, so what do I think the state of the art of using end-to-end deep learning for optical music recognition? I think it is a place with great potential. It is not nearly as good yet as the things you can buy-- SmartScore, SharpEye, the things that are there. But it has a lot of potential.

And one of the things most of the work that I've seen on it is missing are the skills that you guys have here. I saw-- OK, I'm going to forget, and you're going to tell me you put this back up in a second. I just need a couple extra lines.

I actually saw a model where it was using generated scores, which are a little bit different because they're always the same. So you take an output from Finale or MuseScore or something. You create an image file and then you know exactly where every note is because you wrote it and everything. So you can generate a million ground truths and stuff like that. And so it was running through a whole bunch of scores and things.

And this particular one, it had a particular reinforcement learning problem that it successfully identified where all the notes were and where all the staff lines were. It was getting incredibly, incredibly good, but could not break around 80% on then recognizing what the note was.

Why? Because they had trained it on this is A, this is D, this is G, this is F, this is B-flat. This is B-flat. This is E. Actually, we'll make it even better and put that there. That's an E, F-sharp. So these are all being recognized.

AUDIENCE: So the fourth one in treble would also be F-sharp, right?

MICHAEL CUTHBERT: Yeah, that's right. Thank you. Yep. Thank you very much, Adam. But now imagine you're the poor computer that's having to predict, OK, A, I got it right. D, I got it right. Oh, sorry. I didn't mean that. G-- nor that. G-- I got it right.

I think that's F because it's only looking at-- sorry, camera and people on this side-- because it's only looking at this amount of data, whereas it needs some way to know that's sharp. Yeah. And this one it could probably get right is B-flat.

And then, again, it was saying I think that's G-flat. I think that's C. And I think that's D. Because the training data had prioritized treble clef. And once it's here, it has no way of getting back to here. What should the computer be trying to predict on this basis? Anyone take a guess? Based on the amount of information you're trying to get from the computer, what would be essentially an incredibly good prediction model? Yeah?

AUDIENCE: Maybe you just have to predict which line or space?

MICHAEL CUTHBERT: Yeah, what's-- yeah. I see a quarter note on space three. I see a 16th note on line 5. And then when you put all that together, then you have a separate detector trying to detect all the clefs. Or it can be the same one. These things are pretty smart. But yeah, try to do that. And then only afterwards connect those two data to other things.

So that was really nice that-- that was about, I think, about four years ago that this was shown to me as the state of the art. I think that they've gotten better since then. But there's still-- if this kind of mistakes were still being made four years ago, I imagine that there's still many stumbling blocks that you can put in for.

And I think a team from here now with the knowledge that we have could, in a year-long project, produce a real, state-of-the-art tool. Yeah, Misha?

AUDIENCE: With some of the best large language models, they use attention-based models. So in this case, this actually might fix this problem because it would really prioritize the stuff at the beginning of the measure and ignore the A-D-G when evaluating F-sharp.

MICHAEL CUTHBERT: Yeah. Yeah, some of you who are also taking Eran Egozy's Interactive Music Systems had a lecture from one of the experts, Anna Huang, who is an expert in the attention models and the things where you have a memory and thinking back about what's important earlier in the data.

And so I think that her work-- and that's over at Google, Project Magenta-- I think that that could have a lot of impact on the field. And there's other people. Naming a few people doesn't mean that these are the only people doing great things. These are just people who I seem to know.

But none of these people are working, particularly on the optical music recognition problem. So again, we need to someday have 21m.384, where we have a whole semester to just take our projects and do something bigger. But we don't have something like that yet because that would be a great group project. We're going to do all this.

Any other questions? Yeah? Paul?

AUDIENCE: Did you know of any?

MICHAEL CUTHBERT: Are there any classes that I know? No. Right now-- and Paul, are you a senior? Are you? You're a junior. There may be by the time you leave. We are looking at creating more classes on artificial intelligence in music and these kinds of problems. Yeah.

So there's probably-- but definitely-- well, I suppose if you're a junior, yeah, we're already too late. We're in the second semester. It's just been so cold. I still don't feel like it's spring, end of semester there. But yeah, there should be more. And there are other schools that offer some of these things and various workshops that you could go to on some of these techniques.

But we're all still learning. We're all still trying to create the musical equivalent-- for those who have done machine learning know the handwritten 2 2 2 2 data set, the classified thing which everybody does as their first artificial intelligence thing to here's all the different ways 2's been written. Now here's a new number. What is it?

We still need that data set for optical music recognition. Here are 10,000 quarter notes on space 3 from all different things. And now, can your system do that? Cool. Anything else?

Well, I'm going to give you time to either work on your projects or stay here or take advantage of the fact that finally I didn't make it, use up every inch of my time. So just this is the end for my teaching in this class.

And next week, I'm excited because I get to learn a lot new. Now, I've been learning from you all, but I'll get to learn a lot more. So I'm looking forward to that. I'll have office hours as soon as I'm done here. And we'll look forward to projects. So thanks.

[APPLAUSE]

Oh! Oh, that's nice, that's nice. Thank you.