[SQUEAKING]

[RUSTLING]

[CLICKING]

MICHAEL SCOTT ASATO CUTHBERT: Hello, computational music theorists. Here's a short video on what some of the problems in MIR, or Music Information Retrieval, are and how they can be connected to music theory and musicology, which is the main topic of this class, 21m.383. These slides owe a lot to Avi Pfeffer, professor at Harvard, who inspired them with his lectures.

The main question that we'll be working on is, how do we move from sound that's made in the real world to a score that's printed out or displayed on the screen? What kinds of technologies are needed, and what kinds of problems need to be solved in order to do this really well?

Well, first, when we sing or play an instrument, we're making actual sound waves in the real world, so we need to solve the problem of how to most accurately transform sound into a signal. And this involves microphones and other devices, and it's kind of a pure engineering problem that we're not going to be focusing on in this class. But it's very complex and very valuable, and we offer some courses in the music department that will help you do this better.

But once we've got the sound from the real world into an electrical signal, we're working in the time domain, as we call it, and that is the amplitude of sound at a single moment-- forward, backwards, positive, or negative. And we usually record in something close to CD quality or 44,000 times per second. We're measuring the amplitude as a 16-bit signed integer. And so the sound, at this point, looks something like this if this was, I don't know, an attack of an instrument that's decaying, maybe a piano or something of that sort.

The problem here is, how do we convert from the time domain to frequency, to being able to say, OK, all these impulses add up to a particular frequency? How do we decode it from the signal and understand that, hey, in this sound, there's two principal frequencies that are being heard? And to do this, we use Fourier transforms, and the fast Fourier transform algorithm is something that is very good to know in order to work in this field.

So now we have frequencies, and we want to translate them to pitch and pitch connections. And that's not just thinking about, what is A, 440 hertz or something, but which of those frequencies that pop up are prolonged to make a particular pitch? Because even while I'm speaking, there are particular frequencies that are popping up, but they don't necessarily make pitches. The attack on a flute isn't necessarily the pitch that we want. So which ones are connected for this?

And once we have these particular frequency pitch connections, we want to move from there to pitch notes. As we saw with the piano example, there were two frequencies there. Does that mean two pitches, or is one of them a reverberation in the room? Or if you're listening to another instrument, maybe you hear seven pitches.

Six of them are higher and much quieter, and you might think, oh, that's the overtone series of that instrument. And so going from pitch connections to pitch notes involves an estimation of timbre. What is the timbre of the instrument? What is the overtone series that one expects and who we remove them. And from there, do we end up still with multiple pitches in a chord?

Now that we know the musical pitches that we're interested in, we want to go from there to understanding the duration of these pitches and the tempo of the piece. And so there are particular algorithms that we can use to estimate the duration. I mean, one thing is you just look at milliseconds and things and then say, OK, well, that's going to be the most common note. We'll call it a quarter note and so on. And from there, we can estimate the tempo.

Or we can do it in the other way. We can take all these pitches and try to estimate a tempo based on, what are the recurring onsets that seem to be beats? And from there, go and figure out the duration-- quarter note, half note, three beats long, something like that for each of the notes.

So which direction do MIR people go? It turns out that we usually create a sort of feedback loop. So we try to do both at once, and estimations of tempo lead to durations, which then can be rejected or accepted, and estimations of the duration lead to estimations of tempo. And so the technologies involved are beat tracking and self-similarity detection and others.

From the pitches, the duration, and the tempo, we can create what we call the full notes or just the notes. That is something from a pitch and a duration combined.

From all these full notes, we can construct what we call a virtual score, and that is the score as represented purely as collections of notes and other events. We might be estimating, at this point, crescendos and decrescendos, and maybe we're hearing certain notes are slurred, and we're keeping these all as objects inside memory. And so this involves the technology of music representation, which we will be getting to very soon in this class. And so how might we represent a score, maybe as a set of parts with measures and notes in between?

From the virtual score, we need to figure out, OK, how is this going to look on the page or on the screen? And so we perform layout estimation, and we create a laid-out score where it's not just notes inside of an array or a list or something like that but actual positions on a page. And this involves some technologies that we'll be getting to in the class. Here's an example of very poor layout estimation.

From the laid-out score, we choose what the glyphs that would be appropriate are and where are they going to be placed on the page? A sample of some of the things that might be involved in doing that.

And then once we have the laid-out score, the glyph selection, and placement, we move to pixel rendering, individual dots no longer connected to any semantic meaning but just, is this x, y coordinate going to be black, white, or something in between or something else?

And from there, we have particular technologies that we're going to need, such as screen drawing, old-school cathode ray tubes, LED display, things like that, or how to actually make a printer distribute ink on a page. And these, again, are engineering problems that are beyond the scope of this class but still very, very important.

So let's look, again, a summary of what we have here. We move from the time domain to frequency, frequency to pitch connections. We estimate the duration and the tempo, creating full notes. Create a virtual score or logical score. Some people separate these two out. Then we try to make the layout and to the output.

At each of these points in the process, we can involve music theory or musicology, and we can learn aspects about music theory or musicology from the data that we're getting at this point. So, for instance, the duration/tempo level, we can start to ask, well, what are the most commonly used rhythms in this piece? And from there, hey, let's create a backing drum track to accompany the soloist because we figured out what kinds of durations, what kinds of rhythms connect to this with a database, so on.

Or at the frequency to pitch connections level, we can now understand, based on which overtones were removed, what instruments were being used? Are they playing in their natural ranges? And from there, we might estimate the year of this recording based on particular sounds and sound artifacts or the year of the piece based on instruments that are used and so on. So these are places where music theory and musicology can get data from this MIR process.

But there's another side that's been much less talked about and it's a place of great potential that I think people in this class will be able to contribute to, and that is going from music theory or musicology back and using things that we know about theory or history of music or music studies to enhance the quality of the algorithms being used in these places. For instance, at the pitch connection level, we can say, well, let's look at the metadata, and it suggests that this is a piece by Mozart. He died in 1791. So let's look at those instruments.

Now, the frequency, overtone estimation was saying, OK, the most likely instrument is electric guitar. This is very likely wrong. I mean, I don't know. Maybe it's a modern arrangement, but it's pretty likely wrong. Why don't you try this overtone profile of a fortepiano, an early piano instead? And wow, you find out, OK, that was the second-best score but probably the best one when taking into account the metadata on the piece.

Or we can use music theory and musicology at the frequency-detection level. So we might say, hey, all of the notes you've been detecting around this note have been in E major. And then you hear an A sharp, and that's a little fishy. Oh, hey, there's a runner-up note of A natural that was 95% as likely. From a music-theory standpoint, I think that, more likely, we've detected the wrong note. So these are ways that music theory and musicology can enhance this process.

Most exciting at all, though, is when parts of the process feed into music theory and musicology algorithms and are enhanced by it. So we say, oh, hey, given these frequencies, pitches, durations, we get the style of the piece, and we suggest that it's from the swing era.

And then we can help later parts of the process by saying, if the style of the piece is from the swing era, when you do your virtual score, your logical score, probably all those triplets, those "yum, ba bum, ba bum, ba bum" should be written down and represented as eighth notes. That's what, conceptually, they are.

But we can not just move forward. We can, at the virtual score level, maybe it's beginning to lay out the score, and they're putting everything nicely into 3/4. And the musicology or music-theory algorithm says, well, these rhythms are highly unusual for 3/4 time. Let's go back and rerun some of the previous steps with that notion that that was highly improbable and try to make the score again so we can move backwards in time, redo a particular part of the process, rerun it, and then maybe come up with a different result. We'll put this in 6/8.

So here are some of the ways that music theory and musicology can feed back into the MIR sound-to-score process. Next time, sometime soon, I hope to say how we can go the opposite direction, how we can take a score and automatically perform it. And even though we're going to be going in the opposite direction, it turns out that there's a totally different set of problems that we have to solve.