THE GREAT COURSES®

Topic
Science & Mathematics

Subtopic
Mathematics

How Music and Mathematics Relate

Course Guidebook

Professor David Kung
St. Mary's College of Maryland

PUBLISHED BY:

THE GREAT COURSES
Corporate Headquarters
4840 Westfields Boulevard, Suite 500
Chantilly, Virginia 20151-2299
Phone: 1-800-832-2412
Fax: 703-378-3819
www.thegreatcourses.com

David Kung, Ph.D.

Professor of Mathematics
St. Mary's College of Maryland

Professor David Kung is Professor of Mathematics at St. Mary's College of Maryland, the state's public honors college, where he has taught since 2000. He received his B.A. in Mathematics and Physics and his M.A. and Ph.D. in Mathematics from the University of Wisconsin–Madison. Professor Kung's academic work concentrates on topics in mathematics education, particularly the knowledge of student thinking needed to teach college-level mathematics well and the ways in which instructors can gain and use that knowledge.

Professor Kung began violin lessons at age four. In middle school, he quickly progressed through the violin repertoire under the tutelage of Dr. Margery Aber, one of the first violin teachers to bring the Suzuki method to the United States. He attended the prestigious Interlochen music camp and, while completing his undergraduate and graduate degrees in mathematics, performed with the Madison Symphony Orchestra.

Deeply concerned with providing equal opportunities for all math students, Professor Kung has led efforts to establish Emerging Scholars Programs at institutions across the country, including St. Mary's. He organizes a federally funded summer program that targets underrepresented students and first-generation college students early in their careers and aims to increase the chances that these students will go on to complete mathematics majors and graduate degrees.

Professor Kung has received numerous teaching awards. As a graduate student at the University of Wisconsin, he won the math department's Sustained Excellence in Teaching and Service Award and the university-wide Graduate School Excellence in Teaching Award. As a professor, he received the Homer L. Dodge Award for Excellence in Teaching by Junior

Faculty, given by St. Mary's, and the John M. Smith Teaching Award, given by the Maryland-District of Columbia-Virginia Section of the Mathematical Association of America. His innovative classes, including Mathematics for Social Justice and Math, Music, and the Mind, have helped establish St. Mary's as one of the preeminent liberal arts programs in mathematics.

In addition to his academic pursuits, Professor Kung continues to be an active musician, playing chamber music with students and serving as the concertmaster of his community orchestra. ■

Table of Contents

Table of Contents

How Music and Mathematics Relate

Scope:

Great minds have long sought to explain the relationship between mathematics and music. This course will take you inside a fascinating subject filled with beautiful symmetries and simple mathematical explanations of musical sounds you hear every day. Exploring the connections between math and music, while assuming little background in either subject, will help you understand the seemingly simple sound of a vibrating string, the full sound of a symphony orchestra, an intricate Bach canon, how music is recorded, and even the voice of a loved one on the other end of the phone. Throughout the course, the central goal will be to reveal how mathematics helps us understand the musical experience.

Structured around the experience of listening to music, the course will start where all music starts: with vibrating objects. Whether a metal string, a tube of air, or a circular membrane, every instrument vibrates in particular ways, producing not only the frequency we notice most clearly but also a set of predictable overtones. The structure of these overtones, analyzed with mathematics appropriately called "harmonic analysis," leads to myriad fascinating topics: why different cultures choose particular musical scales, how to tune such scales, why some intervals sound more dissonant than others, how to fool the mind with auditory illusions, and why Western scales (and pianos) are always out of tune.

In addition to scales and intervals, rhythm forms a fundamental component of all music. The mathematics of rhythm is sometimes obvious, as in the time signature that closely resembles a fraction, but it sometimes hides itself well, as when a composer implicitly uses number theory to create a sense of instability in the music.

Self-reference sometimes plays a role in composition, as famously noted by Douglas Hofstadter in his classic book *Gödel, Escher, Bach*. We explore both mathematical and musical examples of self-reference, showcasing their sometimes mind-bending weirdness.

Having seen how mathematics helps explain the choice of notes and their lengths, we will turn to the ways composers use mathematics in writing compositions. Although modern composers, notably the 12-tone composers of the early 20th century and later avant-garde composers, explicitly used mathematical ideas from probability and other fields, musicians back to the Baroque period employed mathematical principles when manipulating melodies and harmonies. Understanding the probability and group theory behind their methods—many details of which were not written down by mathematicians until long after they were used in music—will help us understand the mathematical structures hidden in the music we hear every day.

After a composer completes a piece and an ensemble performs it, the final product is most often delivered to our ears digitally, via an MP3 file or a CD. In either case, mathematics plays a crucial role behind the scenes to make the listening experience an enjoyable one. In the case of MP3 (and similar technologies), the harmonic analysis of overtones helps in the compression of files that would otherwise require more lengthy downloads or larger drives. In the case of CDs and DVDs, error-correcting codes and other mathematical techniques are used not only to detect the errors that are unavoidable in the disc-writing process but actually to correct those errors! Incredibly, these mathematical algorithms ensure that the more than 50,000 errors that occur on a typical audio CD will be corrected before sound comes out of your stereo system! In fact, the digitization of music (and musical scores) allows us to accomplish tasks hardly imaginable a generation ago, including fixing out-of-tune notes on the fly and finding a composition knowing only a short melody.

In the final stop on our tour of the musical experience, we will delve into the available evidence for how the brain processes both mathematics and music. By examining similarities between the two subjects on many different levels, from infant development, to how the brain works with patterns, to the level of abstraction, to creativity and beauty, we will arrive at the ultimate connection between the subjects: that similar patterns of thought underlie both mathematics and music.

Throughout our journey, from the origins of single notes to the mental processing of music, the mathematical concepts that help explain musical phenomena will be illustrated with examples, primarily on the violin. Not only is the violin one of the most popular instruments in the orchestra, but it provides a way to visualize much of the mathematics in this course. The shifting in each lecture between interesting mathematics and engaging musical examples enables each subject to illuminate the other, helping us gain a better understanding of, and appreciation for, both mathematics and music. ■

Overtones—Symphony in a Single Note

Lecture 1

This course is structured around a single central question: How can mathematics help us understand the musical experience? Mathematics and music might seem to be separate topics, but our philosophy in this course is to show the connections between these two beautiful subjects. When we see mathematics, we will illustrate the math with musical examples. When we hear music, we will explain the underlying mathematics to help us understand the music better.

Frequency and Wavelength

- Instruments look different, but they all have something in common. When an instrument plays, something is vibrating, which causes a wave of pressure changes that travels through the air and reaches your ear. Mathematicians have studied vibrating objects extensively, and their work helps us understand the sounds produced by instruments.

- When we play a note, such as an A on a violin (440 A), what frequencies are produced? In this case, the answer is 440 hertz (Hz). The higher the frequency, the more times the waves are vibrating per second. For 440 Hz, the waves are vibrating 440 times per second.

- When we look at the spectrum for this note, we see that the peaks have different heights. The spectrum shows us that the string vibrates at many different frequencies, all at the same time. If we play a note at 100 Hz, we see frequencies of 100, 200, 300, and so on—the higher the frequency, the higher the pitch. The lowest frequency produced is called the "fundamental."

- In addition to frequency, we can also look at wavelength. Here, we're measuring from the peak of one trough to the peak of another.

- Sounds are actually pressure waves. Parts of the air are compacted, and parts of the air are rarefied. A 100-Hz note has a wavelength of about 3.4 meters.

- If we measure the wavelength of the overtones instead of the frequency, we get, in this example, 3.4 meters for the fundamental, 1.7 meters, 1.13 meters, and 85 centimeters. If we take that largest one, 3.4, as our measuring stick, we see that those are in a ratio of 1, 1/2, 1/3, 1/4, 1/5, 1/6, and so on.

- The frequencies and wavelengths satisfy a key equation: The frequency multiplied by the wavelength is a constant, the speed of sound.

- If we play a 440 A, what frequencies are produced? We can see the answer in the spectrum: 440 Hz, 880 Hz, 1320 Hz. These are all multiples of 440. The fundamental frequency is 440, and then we just add that repeatedly. Mathematically, we talk about this as an arithmetic sequence. The important thing for us is that it's additive. To get from one to the next, we simply add the fundamental frequency each time.

- A string vibrates at many different frequencies. Listening to a single string vibrating is like sitting in front of an orchestra of the jump ropes we've used here for illustration, each one vibrating in a particular mode.

- Note that there's a difference between pitch and frequency. Pitch is the perceived highness or lowness of a note, whereas frequency is this physical vibrating.

- The mathematical term "harmonic sequence" comes from the harmonics you can play on a string. The relative frequency of each overtone is in a ratio of 1:2, 1:3, 1:4, 1:5, and so on.

- Recall that the key equation here was that the frequency multiplied by the wavelength gives the speed of sound: $f\lambda = v$. If we look at the relative wavelengths versus the relative frequencies, we can see this equation working. If we multiply the frequency by 3, we have to divide the wavelength by 3 in order for them to multiply and get the same velocity, the speed of sound.

Differential Equations

- We need calculus (the study of change) in order to understand this, but we also need differential equations, which use calculus to predict the future. In particular, we need partial differential equations, which have more than one variable.

- There are three steps to understanding and using any mathematical model to predict the future.
  - The first step is to create a model: to use variables, talk about assumptions, and mathematize the situation.

  - The second step is to connect the variables, for which we use equations.

  - Finally, we have to solve the system. We have to use the equations to predict what will happen in the future, to get a function that represents what will happen for all time.

- If we do this well, then the result the model predicts closely follows what we see in reality. If we don't do it well, then the prediction doesn't closely match reality. We have to go back to step 1, change the variables, change the assumptions, tweak the model, and go through the whole process again.

- To construct a mathematical model for a vibrating string, we let x represent the distance along the string at the bridge (assume $L = 1$ unit), t represents time, T represents the tension on the string, and ρ stands for the mass per unit length (weight). The height of the string at position x and time t is $u(x,t)$.

- The next step is to connect the variables using equations. In this case, we will use Newton's equation $F = ma$. When we translate that to this particular situation, we get that the tension (*T*) multiplied by the second partial of *u* with respect to *x* is equal to the weight (ρ) multiplied by the second partial with respect to *t*. In mathematics, this is called a "wave equation."

- Taking into account boundary conditions and initial conditions, the next step is to solve the partial differential equation, and when we do so, we predict the future. In this case, given any initial condition *F* that we deform the string into, the mathematics will tell us what's going to happen for $t > 0$ for all time in the future.

- The power of differential equations is this: If you can model the forces that act on a system, the mathematics predicts what will happen in the future.

- The mathematics is telling us that the string is vibrating in a series of modes. Mathematically, we're getting a sum of sine waves; musically, we're getting a series of overtones.

The Universality of Mathematics

- The beauty of mathematics is that it's universal. Once we have a model for a situation in one field, it can be used in many different fields. In our area of interest, wind and brass instruments involve vibrating air columns, but the mathematics of the situation is almost identical to that for vibrating strings.

- Like a string, a tube of air resonates in particular modes, and what we hear is a combination of those modes—literally, a sum of those modes. If we swing a simple plastic tube, we hear a particular sequence of vibrating modes, perhaps 400 Hz, 600 Hz, 800 Hz, and so on. It's another arithmetic series.

- As we listen to an A played on a series of instruments, the spectrum shows us that we hear a sequence of different frequencies: the overtones.

- Notice that each of these instruments produces more than one frequency. There is a fundamental, and then there are many other peaks in the spectrum.

- Each instrument has essentially the same patterns of frequencies, which again, are additive. They are just multiples of the fundamental. The ratios are 1:2, 1:3, 1:4—the natural numbers.

- If we're looking at wavelengths, we get the reciprocals of those: 1, 1/2, 1/3, 1/4, and so on. The mathematics we use—differential equations—correctly predicts that we will get those particular sequences of numbers.

Visualizing Modes

- To get a different set of overtones, we have to move from one-dimensional objects—strings, vibrating columns of air—to something that's two-dimensional. We are essentially moving from the strings, brass, and winds in an orchestra to the percussion.

- When we play a drum head, we get different vibrating modes, just as we do on a string. But when we look at the spectrum of a timpani, we can see that it doesn't have the same structure as that of the other instruments. There isn't one fundamental with overtones above it. The peaks in the spectrum of the timpani aren't nearly as clear.

- We can use a frequency generator to see the different modes in which the timpani can vibrate.
  - Notice that when we change the frequency, we're not changing the amplitude, although the volume seems to change dramatically. That's not because the drum is vibrating in greater and greater amounts. Instead, it's simply because this particular drum has particular resonances.

  - Note, too, that when you hit a drum head, you don't hit the center (the nodal line), because the drum head doesn't move there. Instead, you hit near the side, away from the nodal lines.

- Once we understand the overtone sequence of an instrument, we can then start playing the instrument. We can vary the frequency of the fundamental. If we couldn't do that, music would be very uninteresting; it would have only a single note.

The Frequency of the Fundamental

- How does the frequency of the fundamental depend on such things as, in the case of a violin, the weight of the string, the tension, and the length?

- Qualitatively, we can say that if the weight of the string increases, the vibrations will slow down. We also know that if the tension increases, the speed of the vibrations will increase. And if the length increases, the vibrations will slow down.

- Differential equations give us a more quantitative answer. The formula here is: $f = \frac{1}{2L}\sqrt{\frac{T}{\rho}}$.

- In this formula, L is the length of the string, T is the tension, and ρ is the weight. We can see this formula at work. If we increase the tension on the string, the pitch goes up. That's because the tension, T, is in the numerator of that fraction. When T increases, the square root also increases, and the frequency goes up. We can also see the formula at work with the weight and length.

- How does the fundamental frequency depend on T, ρ, and L? To answer that, we do exactly what we did before: go through the three steps to make a mathematical model and then solve it. Once we've solved the equation, we can create melodies in predictable ways.

- As you listen to "Twinkle, Twinkle, Little Star," think about how each note is made of many different frequencies—the overtones—and how we're creating a melody by changing the frequency of the fundamental. The way to do that on a violin is by changing the length and the weight of the strings.

Changing Pitches

- To change the pitch on a vibrating column of air, we change the length of the column. We see this in a demonstration with the "Wonder Pipe 4000," an instrument consisting of a PVC pipe and a bucket of water. As we move the pipe lower into the water, the column of air that is vibrating gets shorter. If we raise the pipe almost out of the water, we get a much higher frequency.

© iStockphoto/Thinkstock.

The vibrato motion on a violin slightly changes the length of the string and lends a wavering sound to the pitch.

- Mathematics predicts this difference between a pipe in water and a pipe out of water.
  - The mathematics of a vibrating column of air, for instance, a flute, is similar to the mathematics of a vibrating string. When we looked at a string, $u(x,t)$ represented the height of the string at position x and time t. When we look at a column of air, $u(x,t)$ represents the difference in atmospheric pressure in this particular tube at position x and time t.

  - When we put this into equations, what we get is another version of the wave equation: c^2 (c = speed of sound) multiplied by the second partial of u with respect to x is equal to the second partial of u with respect to t. This comes from fluid dynamics.

- For a flute, the column of air is open at both ends; this fact tells us that the boundary conditions are 0 at both ends.
  - The key difference between a violin and a flute is that when a violin vibrates, there's a sort of middleman in the process. The string vibrates, and that vibrates the air.

- In contrast, when a flute is vibrating, there is no middleman. The air within the instrument is vibrating. The length of the instrument determines the wavelength of the sound.

- We can predict what the wavelength of sound will be from a flute exactly as we did for a violin.

- The lowest note on a clarinet is something around 160 Hz. That's much lower than the 240 Hz on a flute. To understand why clarinets play so much lower, the key is to understand the boundary conditions.
  - A flute, remember, is open to the air at both ends. There is atmospheric pressure at the end and at the mouthpiece.

  - In contrast, a clarinet is open at the bottom, but it's closed at the top. The clarinet player puts the reed entirely in his or her mouth. Pressure can build up against the reed, and that changes the mathematics of the situation.

  - Rather than having boundary conditions of 0 at both ends, we have 0 at the bell end—there is no pressure change—but at the closed end, there is maximum pressure change. There's nowhere for the air to go. It can't vibrate because it comes up against the edge of the reed.

  - The mathematics tells us that one end of the clarinet is like one end of the jump rope: It's held fixed. The other end of the clarinet has a place where the jump rope has to be flat. The mouthpiece, in other words, is located in the middle of the loop of rope when we think about this jump rope. We can use that to figure out the lowest wavelength that a clarinet can produce.

  - A clarinet is about 0.6 meters long. The lowest frequency produced by the clarinet should be, according to this mathematical model, the velocity, 340 meters per second, divided by 4 times the length. That gives us 2.4 meters. When we do the calculation, we get that the lowest frequency a clarinet should be able to produce is about 140 Hz, which is not far off.

- This same mathematics also explains the Wonder Pipe 4000. We realize that a flute cannot play very low because it's an open tube at both ends. A clarinet can play lower because it's closed on one end. When we put the Wonder Pipe down in the water, we're representing the closed mouthpiece, and we get a lower pitch. When it's out of the water, it acts like a flute and has a higher pitch.

- As we've seen in this lecture, a note played on almost any instrument includes many different pitches, different frequencies, called overtones or harmonics. For most instruments, those overtones are related to a harmonic series. In terms of wavelengths, that's 1, 1/2, 1/3, 1/4. In terms of frequencies, those are 1, 2, 3, 4, 5, and 6. We make different pitches by changing the vibrations in mathematically predictable ways.

Suggested Reading

Benson, *Music: A Mathematical Offering*, chapters 1, 3 (for the mathematically advanced).

Fletcher and Rossing, *The Physics of Musical Instruments* (for those with a significant physics background).

Harkleroad, *The Math behind the Music*, chapter 2.

Loy, *Musimathics*, vol. 1, chapters 1–2, 4–5, 7–8; vol. 2, chapter 7.

University of New South Wales, http://www.phys.unsw.edu.au/music/.

Wright, *Mathematics and Music*.

Questions to Consider

1. What is the overtone series of a vibrating object?

2. How is mathematics used to predict the overtone series for a vibrating object?

Timbre—Why Each Instrument Sounds Different

Lecture 2

In the first lecture, we learned that a note played on any instrument includes many different frequencies, called overtones or harmonics. For many instruments, these overtones are related to the harmonic sequence. In wavelengths, that's 1, 1/2, 1/3, 1/4, 1/5, 1/6; in frequencies, those are the ratios 1:2, 1:3, 1:4, and so on. We make different pitches on different instruments by changing the vibrations in mathematically predictable ways, whether it's shortening the string on a violin or putting holes in a flute to shorten the length of a vibrating tube of air. We know that if we play the same note on a flute and a violin at the same loudness, we would still know the difference between the two instruments. The reason we do is the subject of this lecture: timbre.

Defining Timbre

- According to the *Grove Dictionary of Music*, timbre is: "A term describing the tonal quality of a sound. A clarinet and an oboe sounding the same note at the same loudness are said to produce different timbres. Timbre is a more complex attribute than pitch or loudness, which can each be represented by a one-dimensional scale (high-low for pitch, or loud-soft for loudness). … Timbre is defined as the frequency spectrum of a sound."

- A negative definition of timbre comes from the American National Standards Institute: "everything that is not loudness, pitch, or spatial perception." Spatial perception means that you can tell where a sound is coming from. If you hear a sound on your left side, how do you know it's on your left side? Your brain does an amazing calculation to give you this information.

- The mathematics we'll look at in this lecture is the Fourier transform, which breaks a complicated wave into simpler sine and cosine waves. The human ear does this sort of complicated mathematics in differentiating a flute from a violin.

Using Mathematics to Understand Timbre

- To begin, let's look at the sine wave (pressure variations) at 440 Hz.
    - The horizontal axis is time; the scale from one peak to the next is 1/440 seconds. The vertical axis is pressure. The 0 in the middle is atmospheric or ambient pressure.

    - Where we have a peak, that's where we have higher pressure, where the air molecules are compressed. Where we have a trough, that's below ambient pressure; the air molecules are more spread apart, and that's called "rarefaction."

- The *Grove* definition of timbre was the "frequency spectrum of a sound." What is the spectrum? It's the answer to the following question: How would you make this wave by adding up pure sine waves of different frequencies? In other words, the spectrum tells you what the "recipe" is to cook up a particular sound—a particular wave—using sine waves as the ingredients.

- What does it mean to add up sine waves? It turns out that sound waves add just like functions. To add a 440-Hz sine wave and a 500-Hz sine wave, at every point in time, we add the pressure values from the two waves. If they're both high, then when we add them, we get something even bigger. If one isn't a peak and one isn't a valley, then they cancel each other out, and we get something near 0.

- We can also do this process in reverse. What sine waves do we add to get a 440-Hz sawtooth wave?
    - Given the fact that it's a 400-Hz sawtooth wave, we know we need to add up sines that have the same period. In other words, we need sine functions that repeat every 1/400 seconds. That gives us some idea of which sine waves we should take. In fact, if f is the frequency, then the period is $1/f$, and we should take $\sin(2k\pi/ft)$ for $k = 1, 2, 3$, and so on. All of those have the correct period.

- We might try to add up and get a sawtooth wave $s(x)$ by taking $\sin(2\pi/ft) + \sin(4\pi/ft) + \sin(6\pi/ft)$. We might think about taking different amounts of each, and those amounts are a_1, a_2, a_3, and so on. Now, we've reduced the problem to a different question: How much of each ingredient should we take? What should a_k be for $k = 1, 2, 3\ldots$?

- Solving for the amount of each ingredient is a mathematical trick called "orthogonality." It turns out that if we take the $\sin 2j\pi t \sin 2k\pi t$ and integrate (that is, sum all the ingredients) from 0 to 1, we always get 0 unless j and k are the same number. Mathematicians say that those two functions, the $\sin(j\pi t)$ and $\sin(k\pi t)$, are orthogonal, which is a version of perpendicular.

- We can actually think of those functions as being vectors in some abstract function space, and those particular vectors are perpendicular. On the other hand, if $j = k$, then those two vectors are parallel.

The Fourier Series

- This is all the work of Joseph Fourier, a mathematician born in 1768. He was studying various problems when he came up with these Fourier series. He was looking for a recipe to write any periodic function as the sum of sines and cosines. Fourier did this work in the context of studying heat flow and metal plates.

The mathematics of the Fourier transform is key to virtually all signal analysis, including AM-FM transmission.

- His work is used to solve a wide variety of problems in differential equations, anything from signal processing to quantum mechanics. Most problems involving periodic functions use Fourier series or Fourier transforms.

- Let's return to the sawtooth wave problem. We figured out that the ingredients for the sawtooth wave were probably certain sine functions. Interestingly, unlike regular recipes in the kitchen, this recipe looks like it goes on forever. It's an infinite series. Now we're trying to answer the question: How much of each one of the ingredients should we take?

- That's where we turned to the orthogonality trick, and the answer we got from that is that we should take the first sine wave with amount 1, the next one with amount 1/2, the next one with amount 1/3, and so on. We've seen that before. That's actually a copy of the harmonic series!

- We can check this answer both graphically and with sound. As we add more terms (more ingredients in the correct proportions), the function looks and sounds more like the sawtooth wave.

- So far, we've talked about Fourier sine series, for which there are two important generalizations.
  - The first is that we could use sines and cosines but stay with even multiples of π for our frequencies: π, 2π, 3π. If we use sines and cosines with those as the argument, those are called Fourier series in general.

  - The second generalization we could make is to allow any frequency, not just integer multiples of π. If we do that, we get the mathematics called the Fourier transform. In all of these cases, the point is to break down a complicated periodic function into simple sine waves that we understand much better.

The Fourier Transform

- The mathematics of the Fourier transform is incredibly advanced. But for this lecture, what we need to know is that the Fourier transform takes a complicated wave and breaks it down into component sine waves of different frequencies. We call that the spectrum.

- It's also important to know that this is reversible. We can go from the wave form to the spectrum and back. Mathematicians call that going from the function side to the Fourier transform side and doing the inverse Fourier transform to get back. That's what our ears do: a Fourier transform.

Understanding Instrument Sounds

- Let's return to the simple sine wave A at 440 Hz. Again, peak to peak, we see a gap of 1/440 seconds. The spectrum of this, if we take the Fourier transform, shows a single peak at 440 Hz. In theory, it's a single infinitely tall point. It's a delta function. In practice, the computational issues smooth this out, so you see a peak sort of smoothing out to the sides.

- Does this sort of pure sine wave ever occur? Does anything vibrate in just a sine wave?
  - It turns out that some bird calls, such as that of a black-capped chickadee, are very close. Its wave form looks remarkably like a simple sine wave. Its spectrum is nearly a single peak at 3850 Hz. All of the other overtones are much smaller.

  - It's important to note that the vertical scale of the spectrum is logarithmic. Each line is 10 decibels, and the scale is multiplicative. When something is three lines below (30 decibels below), that's actually $10 \times 10 \times 10$ less. It's 1/1000 as powerful. The chickadee is singing almost a pure sine wave. All the other frequencies are much softer.

- Let's return to the spectrum of the sawtooth wave. We already figured out the recipe for this wave. When we look at its spectrum, it shows us visually the recipe we use to get it; we take each of the frequencies in smaller and smaller proportions. That's the harmonic series that we saw in our recipe: 1, 1/2, 1/3, 1/4, and so on. This confirms our idea of the correct recipe for the sawtooth wave.

- When we look at the wave form for an A on the violin, we see that it repeats, and the gaps are, again, 1/440 seconds. Its spectrum has peaks at multiples of 440. Here, we're concentrating on the heights of those peaks: How much of each harmonic are we hearing? The second harmonic, the one that's at 880 Hz, is actually 20 decibels below the fundamental. That means it's 100 times lower. On the other hand, the fifth harmonic is almost as loud as the fundamental.

- The spectrum of an A on a trumpet also shows peaks at multiples of 440, but the heights of those peaks form a different pattern from the spectrum of a violin. The clarinet spectrum shows that all the odd overtones of the clarinet A are much louder than the even ones. This relates to the fact that one end of the tube is closed in a clarinet.

- Why should we care about the spectrum or the Fourier transform? It's how we distinguish different instruments and different voices from different timbres. It has to do with the heights of the various spectrum peaks. How our ears do this has a fascinating mathematical component to it.

Resonant Frequency

- Deep inside the ear, on the other side of the ear drum, is an organ called the cochlea. It's conical-shaped—different sizes at different places. That means that different places resonate at different frequencies.

- If a particular sound comes in—if a particular sine wave comes in—there is some place in the cochlea where that resonates very loudly. The basilar membrane is inside the cochlea, and it picks up those vibrations and sends that message to your brain. In this way, your ear is figuring out a recipe for that sound.

- When an A is played on a violin, the sound wave hits your ear. Each overtone on the spectrum—each ingredient in our recipe—

resonates the cochlea in a different place and with a different amplitude, producing a different force on the basilar membrane. The ear then does exactly what the Fourier transform does. It separates a complicated wave into simple sine functions and sends those to your brain.

- Your brain has stored-up patterns of spectra from various instruments. Your brain knows that a spectrum of a particular pattern is a violin or a clarinet. The brain then does pattern matching.

- The caveat to all this is that the spectrum changes over time. The spectrum of the very beginning of the note (the "attack") is crucial. If we remove the attack electronically, it becomes difficult to distinguish the instrument. The timbre of the attack is particularly important.

- We hear four different ways of playing A 440 Hz on a violin: pizzicato, open A, A played on the G string, and ponticello. Looking at the spectra of these four sounds after the attack reveals differences. For example, the pizzicato is all about the lowest overtones. The attack would have higher overtones, but they die out fairly quickly. The open A has a rich set of overtones in most ranges.

Harmonics

- String harmonics are different from the harmonics that describe the modes of vibrations, the overtones. The idea here is that we lightly stop the string from vibrating at a particular point. If we do this in the fundamental mode, where the whole string is vibrating, the string will be completely disrupted.

- If we stop the string in the middle of the second harmonic, the first overtone, there's no disruption. The string can continue to vibrate on each side (each loop) even if the string is lightly stopped in the middle. Notice that all the even modes have nodes in the middle, which would mean no disruption. All of the odd nodes would be disrupted because they are moving in the middle.

- The spectrum of an A that is lightly stopped in the middle acts exactly like the mathematical description. The harmonic gives us only the even overtones; all of the odd ones are damped out. Notice, too, that this changes the fundamental. The lowest frequency is no longer 440 Hz; it's now 880 Hz, double the fundamental frequency of the original.

- If we stop the string two-thirds of the way up, we would hear only the modes where there's a node at that place. The fundamental doesn't have a node there, nor does the next mode. In fact, only the multiples of 3 will have nodes at two-thirds. We should hear only every third overtone.
  - When we do this, the timbre and the fundamental change. The timbre changes because we eliminate some of the overtones. The fundamental changes because the lowest frequency we hear is three times the original fundamental's frequency.

  - Stopping the string one-third of the way up sounds almost exactly the same as stopping it two-thirds of the way, again, because we're looking at only the multiples of 3 in the original overtone series of the A.

Pianos and the Seventh Harmonic

- If you're designing a piano, you have a choice about where on the string the hammer should hit. Different choices will give you different timbres.

- Remember that when we were looking at the overtones, we looked at the seventh of those. The seventh harmonic was not a note that was on our 12-tone scale. When we're making a piano, we can choose to put the hammer in a place so that the seventh overtone is less audible.

- Using partial differential equations, we can actually predict how much of each overtone we will hear for a given hammer position. To avoid the seventh harmonic, we position the hammer exactly one-seventh of the way up the string.

- To understand this, think back to the jump rope. The seventh harmonic has seven loops and will have a node one-seventh of the way up the string. If we hit it there, we will disrupt that mode, so we won't hear any of the seventh harmonic. This is actually the way that pianos are made.

How Composers Use Timbres

- The primary way that composers use timbre is by choosing different instruments for different parts. Think about the piece *Peter and the Wolf* by Prokofiev. The bassoon represents the grandfather, the oboe represents the duck, and so on. These are different timbres for different instruments because of the storyline.

- A more subtle way of doing this is not by having different instruments play different things because of the difference in timbre, but by having a single instrument play in different ways—plucking, bowing, or playing harmonics.

- This takes us back to the opening music for this lecture, Bach's "Air on the G String" from his Orchestra Suite No. 3 in D Major. A German violinist, August Wilhelm, adapted this piece just for violin and piano. He changed the key to C, brought it down an octave, and had the violinists play entirely on the G string, which gives the music a darker quality. We perceive the lower, darker overtones via a Fourier transform.

Suggested Reading

Benson, *Music: A Mathematical Offering*, chapter 2 (for the mathematically advanced).

Fletcher and Rossing, *The Physics of Musical Instruments*.

Harkleroad, *The Math behind the Music*.

Loy, *Musimathics*, vol. 1, chapters 2, 6, 8; vol. 2, chapters 3, 6.

University of New South Wales, http://www.phys.unsw.edu.au/music/.

Wright, *Mathematics and Music*, chapter 10 (an excellent, less mathematically technical discussion).

Questions to Consider

1. How is the timbre of a note related to the spectrum, overtone series, and Fourier transform?

2. In what sense does the ear perform a Fourier transform before sending information about a note on to the brain?

Pitch and Auditory Illusions

Lecture 3

Thus far, we've learned that a note on almost any instrument produces many different frequencies, called overtones or harmonics. We've also learned that how much each overtone is produced is called timbre. And we've learned to break down a wave form into its constituent frequencies, figuring out what recipe goes with a particular sound. The mathematics of that is the Fourier transform. In this lecture, we will see how these ideas lead to our brains being tricked with auditory illusions. We'll also learn why pitch is not as simple as a low-to-high continuum.

Defining Auditory Illusions

- An auditory illusion is similar to an optical illusion, but rather than a visual stimulus, it's a sound that tricks the brain.

- Consider a male voice versus a female voice on a cell phone. A man's voice vibrates at around 100 Hz; a woman's voice is much higher, perhaps 350 Hz. But the speaker on a cell phone has a range of only 350 Hz to 4000 Hz. Your brain thinks you hear the low frequencies of a male voice, but the speaker on the cell phone can't produce frequencies that low. It misses the fundamental in the first couple of overtones.

- Why is your brain tricked? Let's go through exactly what happens when you hear a male voice on a cell phone.
  - Vocal cords, like all the other one-dimensional vibrators we've discussed, vibrate at a particular sequence of overtones. When you hear a male voice at 100 Hz, his vocal cords are also vibrating at 200 Hz, 300 Hz, and so on. When he speaks into his phone, those vibrations are digitally encoded and sent as 0s and 1s, and the wave form is decoded by your phone.

  - At this point, you're still not missing the fundamental on the first few overtones. The entire signal is present, but when it's played back on your phone's speaker, you lose some of the lower overtones.

- Here's where your brain gets involved. Your ear, first of all, does the Fourier transform and sends the spectrum to the brain. The brain recognizes that if you were hearing a 400-Hz sound, the overtones would be 800, 1200, 1600, and so on. But the sound you're hearing also has 500, 600, 700, 900, and so on. In fact, three-fourths of the sounds that are coming into your brain don't fit the pattern of a 400-Hz sound.

- The pattern matches closer to what your brain knows as a 100-Hz sound. All that's missing are the first three overtones, 100, 200, and 300.

- In your brain, the idea of a low G means a particular set of neurons all firing at the same time. When you hear a 100-Hz note, there's actually a neuron firing 100 times per second. When there's something vibrating at 200 times per second, there's another neuron that's firing at exactly 200 times per second. Thus, the idea of a low G is simply a set of neurons firing at the same time. When you remove just three of those, you're still firing the same set of neurons. To your brain, the pitch low G is just a particular pattern of neurons firing together.

- We need to remember here the difference between pitch and frequency. Both relate to how high or how low a note is. But pitch is the perceptual attribute—the "psychoacoustical attribute" of a sound—whereas frequency is the physical attribute of the wave form.

- The missing fundamental tells us that pitch and frequency are not the same. When a male speaks at 100 Hz, the pitch is a low G, but the lowest frequency that comes through on the phone is 400 Hz.

Musical Notation

- On a piano keyboard, the higher pitches are to the right and the lower pitches are to the left. The white keys are called A, B, C, D, up through G, and then that pattern repeats. The reference place on a piano is A, 440 Hz. The black keys are the sharps and the flats, the incidentals. For instance, the black key between A and B is A-sharp and B-flat.

- The interval between two notes is the distance in pitch from the lower one to the higher one. The smallest interval on a piano is to go from one key to the next, black or white, or from white to white.
  - Going from A to A-sharp would be a half-step; going from G-flat to G would be a half-step; and going between two white keys, B and C, is also a half-step.

  - Going from one key to the next of the same name is called an octave. Thus, going from an A to another A or going from a D-sharp to the next sharp is an octave.

- If we need to distinguish As, we number them: A^0 is the lowest A on the piano, followed by A^1, A^2, A^3. A^4 is at 440 Hz. There's a bit of an oddity about this: You increase the number not at A but at C. Thus, the bottommost notes are A^0, B^0, and C^1, and if you go up to 440, that's A^4 and then B^4 and C^5.

- A sharp symbol means to go up one half-step. That really means to go up one key on the piano, and a flat symbol means to go down one. Thus, B-sharp is actually the same thing as a C; an F-flat is the same thing as an E. A B-double-sharp is the same thing as a C-sharp, and an F-double-flat is the same thing as an E-flat, at least on a piano.

- An octave is made up of 12 half-steps. If you start at one key and go up 12 keys, you'll get to another key of the same name. How do these notes correspond to frequencies? For example, what are the frequencies of the As in the octaves above and below 440 Hz? Starting at 440 Hz, the next A is at 880, and the next one is at 1760.

We have to double the frequency every time we go up an octave. To go down from an A at 440, we halve it, so we get 220, 110, and 55.

- The key observation here is that octaves are multiplicative. Remember that overtones are additive. This difference between the multiplicative system of intervals and an additive system of overtones has numerous implications, some of which are very surprising.
  - Suppose a piano key sounds at some fundamental frequency x. What are the frequencies of the octaves above and below x? For the octaves above, the answer would be $2x$, $4x$, $8x$, and $16x$. For the octaves below, the answer would be $1/2x$, $1/4x$, and $1/8x$.

  - In general, if we want to go n octaves away from x, the formula would be $2nx$, and that works for both positive and negative n.

  - What about the overtones of x? Those are the multiples $2x$, $3x$, $4x$, $5x$, $6x$, and $7x$. In general, the end-harmonic above x will be nx. Note that this works only for positive n. There are no harmonics below the fundamental, at least in natural sounds. To go from one to the next, we are just adding x again. The general formula here is $n \times x$, and it's multiplication because multiplication is just repeated addition.

Comparing Overtones and Octaves

- Let's compare overtones and octaves played at frequency x. We hear the overtones x, $2x$, $3x$, $4x$, $5x$. The octaves above that are x, $2x$, $4x$, $8x$, $16x$. All of the octaves are in the overtone series, but not all of the notes in the overtone series are octaves.

- What about the notes that aren't octaves? We hear a full G at 200 Hz and then we hear each overtone isolated in turn. The fundamental is a G^3. The first overtone (the second harmonic) at 400 Hz is a G^4. The third harmonic, now at 600 Hz, is no longer a G; it's a D^5. It shouldn't be surprising that the fourth harmonic at 800 Hz is a

G again because we doubled the 400 Hz we just heard. The fifth harmonic at 1000 Hz is a B^5.

- There are many different representations of musical notes: the sound itself, numbers (frequencies), keys on a piano, notes on a staff, and peaks on a spectrum. We see these different representations added as we listen to the first 10 overtones of F^2.

- In all these representations, the octaves fall on the powers of 2, as we've discussed. Note, too, that the third harmonic is a C. That's the first note that isn't an F. It's the first note that isn't in the octave scale, and the C is a fifth above F. In other words, C is the fifth note on the F-major scale; that's how it gets the name "fifth."

- For our purposes, the two most important intervals are the octave, going from one note to the next of the same name, and the fifth, the fifth note on the major scale of the lower note. Another way to think of a fifth is going up seven half-steps.

- A key mathematical observation is that going up an octave is multiplying the frequency by 2. Going down an octave is dividing by 2. Going up a fifth is multiplying the frequency by 3/2; therefore, going down is dividing by 3/2. We can walk through a numerical example with the fundamental 87 Hz.

- In general, we see that this works, but there's an important complicating fact here.
    - The first overtone above an F^2 is an F^3. But the overtone series of an F^3 is not the same as the overtone series of an F^2.

    - The third harmonic above the F^2 is C^4, but when we play a C^4 on the piano, we hear *its* overtone series.

    - Each note comes with its own symphony of overtones. If the overtones of two notes match up, the result is a pleasant sound, what musicians call "consonance." If the overtones don't match up, the result is an unpleasant sound, "dissonance."

- The reason the overtones get closer together as we go up the piano keyboard is that the intervals are multiplicative. If we go from one octave to the next, we're multiplying by 2.
    - The normal number line has equally spaced points; the differences between numbers are the same.

    - But on a keyboard, it's the ratios that are the same, not the differences. It's a logarithmic scale: Every time you go up, you're multiplying, not adding.

Deconstructing the Missing Fundamental

- As we said, when the lower overtones are removed, the brain reconstructs the fundamental and fools you. We can hear this when we progressively remove the overtones from a G^3. The timbre changes, but because of the missing-fundamental illusion, musicians will agree that the fundamental stays at G^3.

- What happens if we add overtones? We get a bigger and more complete picture of the sound, and gradually, we start to actually hear the fundamental.

- Organ makers use this auditory illusion. The lowest note needed in organ music most of the time is incredibly low; it's at 16.4 Hz. To produce a note that low requires a 32-foot pipe, which is too large for many churches. The solution is to use the missing-fundamental illusion to make people think the 32-foot pipe is present. This is

The lowest note produced in most organ music is at 16.4 Hz, which requires a 32-foot pipe.

done by adding smaller pipes with higher frequencies to give the illusion of a lower frequency.

- A 32-foot pipe vibrates at 16.4 Hz, which we'll call x. The overtones of a correct 32-foot pipe should be $2x$, $3x$, $4x$, $5x$, $6x$, and so on. Which smaller pipes could we use to simulate that 32-foot pipe?

- The idea here is to produce the first and second overtones of what would be a 32-foot pipe so that listeners will hear the fundamental. What pipe would have a fundamental of $2x$? The answer is a 16-foot pipe, exactly half the length of the 32-foot pipe. The wavelength is half, so the frequency is doubled.

- However, a 16-foot pipe will produce overtones only at $2x$, $4x$, $6x$, $8x$, $10x$, $12x$. We are missing $3x$, so we need to add another pipe with a fundamental of $3x$. We're trying to triple the frequency, which means we cut the wavelength by a factor of 1/3, and that's the size pipe we need: a pipe that measures 10 2/3 feet.

- When we put all the overtones from those two pipes together, we're missing only a few, and that gives the illusion of something playing at x Hz.

The Scale Illusion

- Let's begin by listening to the full sound of the scale illusion. If you're like most people, in one ear, you hear the sound go down and up in pitch, and in the other ear, it seems to go up and down.

- In fact, there were no descending or ascending scales at all. The sounds jumped around, but your brain is so used to hearing scales that go up and down that it mixed the sound up and decided you must be hearing a scale.

- Tchaikovsky used this illusion in 1893 in his Sixth Symphony, *Pathétique*. The scale illusion comes in the opening of the last movement. The melody is not really present in the first or second violin parts; instead, the melody emerges from both parts (the first and second violins) coming together with the scale illusion.

Shepard Tones, Falling Bells, and the Tritone Paradox

- Shepard tones are reminiscent of an optical illusion known as the Penrose staircase. They always go up but somehow manage to get back to where they started. How can we make an endlessly rising note? The key is that each note is a symphony of different frequencies.
  - If you sit down at a piano and play A, A-sharp, B, C, and C-sharp, by the time you get to the next A, you will have played all 12 notes and gone up exactly one octave.
  - Now imagine sitting down at a keyboard that goes infinitely low and infinitely high. You play not just one A but every single A on the keyboard, and then every single A-sharp, and then every single B, C, C-sharp. By the time you get up to A, you're playing the same note you started with because you're playing every single A on the piano.
  - At every stage, you're going up one half-step, but you manage to get back to the beginning. In other words, Shepard tones are actually in a circle.
- Related to Shepard tones is an illusion called falling bells. These are notes that seem to rise, but they don't come back to where they started; they fall over time. The idea is to take Shepard tones, again, on an infinite piano, and slowly move the "envelope" of tones downward. From one note to the next, the pitch is definitely going up, but over time, the envelope drags the pitches lower and lower.
- In the tritone paradox, we hear four pairs of notes. Interestingly, people disagree on whether those pairs go from higher notes to lower notes or vice versa.
  - A "tritone" is a musical term for a particular interval of six half-steps. From A to D-sharp or from C to F-sharp would be a tritone.

- In this paradox, two Shepard tones are played that are six half-steps apart. Two notes that are a distance of six half-steps apart are across from each other on the circle.

- If you think the intervals go up, that means you're thinking that you are going one way around the circle. If you think the notes are going down, you're thinking that you are going the other way around the circle.

- When you go from playing every A on an infinite piano to every D-sharp, you might have gone from every A up to D-sharp, or you might have gone from every A down six keys to D-sharp. Because it's every A and then every D-sharp, you cannot distinguish between the two.

- Each note is made up of many different frequencies. You can compare two individual frequencies, but once we manufacture them and put them into these notes, you cannot compare the entire note.

- In Lecture 2, we learned that timbre is more complex than pitch because pitch can be represented on a one-dimensional scale of low to high. The Shepard tones and other illusions show that it's not that simple. Here, notes are constructed that cannot be compared in pitch; neither is really higher or lower than the other. Frequency—the physical attribute, not the perceived one—is a one-dimensional scale, but pitch—the perceived attribute—is much more complicated.

Suggested Reading

Benson, *Music: A Mathematical Offering*, chapter 4.

Deutsch, "Diana Deutsch's Audio Illusions."

Loy, *Musimathics*, vol. 1, chapter 6.

Questions to Consider

1. How does our knowledge of how the ear works explain the missing-fundamental illusion?

2. How did Shepard use the fact that we interpret multiple frequencies (in a harmonic series) as a single note to create an endlessly rising note?

How Scales Are Constructed

Lecture 4

Vivaldi's *Spring* and *Lady Meng Jiang* are from two very different musical traditions, Western and Chinese. But what is it that makes them different? Many differences exist between music from different cultures, such as instruments, notations, and so on, but one of the key differences is that they use different musical scales. They make different choices about notes. In this lecture and the next, we'll look at how the harmonic series informs our choice of scales and how we tune those scales.

Defining a Scale

- A scale is a collection of notes, increasing or decreasing in pitch. For our purposes, we can think about them as covering a range of one octave, from one frequency to its double (e.g., from A at 440 Hz to 880 Hz).

- Remember, an octave is from one key on a piano to the next of the same name, 12 half-steps. As we know, the keys on a piano have names—A, B, C-sharp—but when we talk about scales, we also give them numbers—the fundamental, the second up through the seventh, and the octave.

- The number of notes on a scale is key. We talk about the number of notes on a scale as being the number of notes before you get to the next octave. When you play a scale, you usually start on the bottom and go to the next octave, but if you have a seven-note scale, a "heptatonic scale," you actually play eight notes because you include the eighth note on the top end.

- The piano's keys are sort of a fixed reference point; the names of the keys never change. However, the numbers—the fundamental, second, third, and so on—do change. When you're in the key of C, C is where you measure from, and when you're in the key of G, G is where you measure from.

- When we talk about the key of the scale, we're not talking about a piano key. We're talking about the lowest note of the scale, usually the first. When we talk about the key of some composition, that's the note that's most important, and pieces tend to start and end on that particular note.

- An interval is the distance in pitch between one note and another. It's related to scales; in numbers, the relation is as follows: The interval of the fifth is the distance between the fundamental and the fifth note on the major scale when starting on the fundamental.

Constructing Scales

- Scales can have different moods (bright and happy versus darker and sad), but one of the things they all have in common is that they all have the fifth and the octave. Those are the second and third harmonics we hear when we're vibrating just a single note or a single column of air. Nearly every musical tradition on earth contains both the fifth and octave.

One of the key differences between Western music and music from other cultures is that they use different musical scales—they make different choices about notes.

- Let's review what we need to construct scales.
  - o We know that objects vibrate in different modes (overtones), and we know that the wavelengths of the most common overtones are in a ratio of 1/1 to 1/2, 1/3, 1/4. The harmonic series and the frequencies are just multiples, 1, 2, 3, 4; those are arithmetic series.

  - o What we learn from the overtone series is that to go up an octave, we multiply frequencies by 2, and to go down an octave, we divide frequencies by 2. We also learned about a fifth—seven half-notes on a piano is a fifth—and to go up a fifth, we multiply the frequencies by 3/2. To go back down, we divide by 3/2.

- Let's look at a numerical example: How do we go up an octave from a 440 A? We multiply by 2 and get 880 Hz. If we start back at A and want to go up a fifth, that would be the fifth note on the A-major scale, E^5. To figure out what the frequency of that is, we multiply 440 Hz by 3/2 and get 660 Hz. The octaves and the fifths are the key ingredients we need to construct scales.

Choosing Notes on a Scale

- Let's start by building a scale with a 100-Hz note, roughly a G^2. The next octave is 200 Hz, a G^3. We can try building a 5-note scale by putting the notes equally spaced between 100 and 200 Hz (i.e., 100, 120, 140, 160, 180, 200). The resulting scale sounds nonstandard.

- Let's now take this pattern up an octave and go from 200 to 400 Hz (between a G^3 and a G^4). Again, our pattern is that we're going up 20 Hz at a time, which means that this scale has 10 notes, not 5. These notes sound much closer together in pitch.

- Going 20 Hz at a time from 400 to 800 Hz (from a G^4 to a G^5), there are 20 notes in the octave. This would be strange because it would mean that higher voices would have more notes available to them in their octaves. In other words, some of the notes for sopranos might not exist for the tenors or basses.

- The key problem is this: These intervals and octaves have a multiplicative structure. We shouldn't be adding the same number each time; we should be multiplying by the same number each time.
    - Sticking to this multiplicative system puts corresponding notes in each octave for each voice, and it reduces the problem to looking at a single octave.

- If we can figure out where the notes go in a single octave, we can use the multiplicative structure to get the same notes in higher octaves, for sopranos, and in lower octaves, for basses.

Just Tuning

- For our first attempt to choose scale notes for a single octave, we'll try using the overtone series of the bottom note to determine the higher notes. This is called "just tuning." Let's find the notes for an A-major scale.

- We can figure out the overtones and the notes for an A^4 at 440 Hz: A^4, A^5, E^6, A^6, C-sharp^7, E^7, sharp of an F-sharp^7, A^7, and B^7. The problem is that most of those notes are not in the octave we want, which is between A^4 and A^5.

- Instead of working with 440 Hz, let's talk about relative frequency; let's treat the A^4 as 1. We're looking to get relative frequencies between 1 and 2, and to do this, we're going to go down an octave, which is dividing by 2 in frequency.
    - One of the notes we have in the overtone series of the A is an E^6, but that isn't the note we want. We actually want an E^5, and to get from an E^6 to an E^5, we're just going down an octave, which is dividing by 2. An E^6 is the third harmonic above A, so it has a frequency three times A; we divide that in half, and we get our E^5 of 3/2.

    - We can do the same thing with a C-sharp. C-sharp is the fifth harmonic, so it's five times the frequency of A, but it's two octaves too high. We divide by 2 twice, and we get 5/4 for our C-sharp.

    - We can take as many of the overtones of A as we want and bring them down an octave; the result is a just scale.

A Pentatonic Scale

- To get a just-tuned pentatonic scale, we have to do only the notes we just discussed.

- We have an A (the fundamental), and the second one is also an A. The third one is an E; that's a new note. The fourth is an A again. The fifth is a C-sharp. The sixth is an E again. The seventh is an F-sharp (at least that's the closest note on our scale). Finally, the ninth note, the B, is the fifth note we have.

- To get a pentatonic just scale, we take harmonics 1 through 9—because of the duplication, we need all of those first nine to get just five notes—and we get those in the right octave by dividing by 2.

• The nice thing about this just pentatonic scale is that everything is perfectly in tune with the A. The downside of it is that the notes are not necessarily in tune with each other. Let's look at what this means; in particular, let's look at the B versus the F-sharp.
  - In terms of frequency, we arrived at a 495-Hz B. That B has its own overtone series, so when we play the B at 495 Hz, we also hear its first overtone (its second harmonic) at 990, and its second overtone (its third harmonic) at 1485. Roughly, those are a B (the fundamental), the octave (B^5), and the next overtone (F-sharp^6).

  - But the F-sharp that we've decided on for our scale, the F-sharp^5, has its own overtone series. The fundamental frequency for that F-sharp is 770 Hz, and its first overtone is 1540, but its first overtone should be an octave above; it should be F-sharp^6.

  - This is a problem because one of the overtones of the B is an F-sharp^6 at 1485 Hz, and one of the overtones of the F-sharp is an F-sharp^6 at 1549 Hz. Those are very different; the two notes would produce dissonance.

• Let's look at this B to F-sharp problem a different way. The F-sharp should be the fifth note above B, exactly seven half-steps above B. Therefore, it should have a frequency that is 3/2 times the frequency of B. That's what it means to go up a fifth. The frequency of B is 495; multiplying by 3/2, we get 742.5 Hz. That's very different from 770 Hz.

- This tells us that there's a real problem with tuning in general. We can get the B and F-sharp to be perfectly in tune with A, but we can't get them to be perfectly in tune with each other.

A Heptatonic Scale

- To go from a five-note scale to a seven-note scale, the philosophy is, again, that we're building on the overtones of the fixed reference point of the A and bringing those back into the correct octave by dividing by 2. That way, we'll get an A-major just heptatonic scale.

- However, the two notes we added when we went from a pentatonic to a heptatonic scale are not quite right. The D and the F-sharp can't have come from this process.
  - When we look at the harmonics of an A, we're multiplying by n, and for the octaves below that, we're taking one of those and dividing by 2. That tells us that every note we get in this process has to have the form $n/2^k$. There's no way of getting 4/3 or 5/3 from this process because 4/3 and 5/3 are not of the form $n/2^k$.

  - The trick that's used here to get a just heptatonic scale is the fact that D does not appear on the overtone series of A, but on the overtone series of a D are other As; thus, we can tune the Ds so that the overtone series of the D matches up with the A.

Musical Results of Just Tuning

- Great music is built on this philosophy of just tuning. The key question to ask to determine whether a musical tradition should use just tuning is whether or not modulation (switching keys) is used.
  - If we modulate "Twinkle, Twinkle, Little Star" from A to C, it's still recognizable as the same melody; it just starts on a different note.

- o Remember that the key is the bottom note of the scale, and it's the note that a melody returns to. It's usually the first and last note of the melody; switching keys is just starting on a different note.

- o If you're not going to do any transposing and you're not going to play multiple notes at once—only notes played with the fundamental—then just tuning is wonderful.

- o But if you want to modulate—if you want to be able to switch keys or to play chords with two notes that are not the fundamental—then just tuning leads to problems.

- Just tuning is used with the bagpipe in Scottish culture. The bagpipe has a drone that is always heard underneath the music. The chanter is a recorder-like part that plays the higher notes, one note at a time. There's no problem like the B and F-sharp that we just saw, because the chanter simply plays one note at a time; it never plays two of those notes. Thus, a bagpipe is properly tuned as just tuned.

- Indian music, played with such instruments as the sitar and the tabla, also always stays in one key and has a drone that plays a low note throughout an entire piece. The melody is then played on top, one note at a time above the drone. Each scale is matched to different overtones of the fundamental, and different sets of overtones are chosen depending on which scale is being used.

Bootstrapping

- A second way to choose notes on a scale is bootstrapping, that is, going from one to the next. Here again, we'll use the overtone series 1, 1/2, 1/3, 1/4 and frequencies of 1, 2, 3, 4, 5.

- The first philosophy was to take A as a fixed reference point and build on the overtones of A. The second philosophy is to take A and use that to get a note, and then to take that as the new reference point and get another note, and so on.

- The idea is that the overtones of each note should be included in the scale. It's not possible to include all of them, but we're going to try for most audible.

- The most audible intervals we hear are the octave and the fifth. The second harmonic is the octave and the third harmonic gives us the fifth, so if we want to use this philosophy, we have to include the octave and the fifth.

- We start with A; the third harmonic of an A is an E. E is the fifth note on an A-major scale. We then add the third harmonic of the E, a B. Once we bring that into the correct octave, we have another note. Next, we add the fifth note above B, an F-sharp. Then, the third harmonic of the F-sharp is a C-sharp, which gives us a pentatonic scale.

- The notes in this pentatonic scale sound consonant, but the problem here is that we've ended with a C-sharp. The second overtone above a C-sharp is a G-sharp, and that's not on our pentatonic scale. Still, a pentatonic scale can be used to produce interesting music.

- For example, Chinese music, such as the piece we heard at the beginning of this lecture, *Lady Meng Jiang*, uses a pentatonic scale. In China, this piece would be played on a two-stringed instrument called an *erhu*.

- A melody in the second movement of Dvořák's *New World Symphony* is based on a D-flat major pentatonic scale. Dvořák chose the English horn to play this melody, the instrument in an orchestra whose spectrum is closest to a human voice.

Why Pianos Are Never in Tune

- The bottom note on a piano is A^0. The idea is that we will start there and bootstrap up a fifth repeatedly, tuning the fifths exactly, until we reach A^7. To go up a fifth, we multiply the frequency by 3/2.

- A^7 is seven octaves above A^0, and we know that to go up an octave, we multiply the frequency by 2. Checking our work, we find that the fifths we got were $27.5 \times 3/2^{12}$ and the octaves were 27.5×2^7. Obviously, those are not the same numbers.

- How should we tune this top A^7? Should we put it in tune with all the other fifths, or should we put it in tune with the octaves? The fact that we can't do both explains why no piano is ever in tune.

Suggested Reading

Benson, *Music: A Mathematical Offering*, chapters 4–6.

Forster, *Musical Mathematics*, chapters 9–11.

Harkleroad, *The Math behind the Music*, chapter 3.

Loy, *Musimathics*, vol. 1, chapter 3.

Wright, *Mathematics and Music*, chapters 4–6, 11–12.

Questions to Consider

1. How can you use the harmonic series of a single note to create a scale with as many notes as you wish?

2. What makes a 12-note scale a natural choice?

3. Why is a piano never in tune?

How Scale Tunings and Composition Coevolved

Lecture 5

The main point of the last lecture was that mathematics, especially the math of the harmonic sequence, informs which notes to include in a scale. In this lecture, we will continue to discuss scales, in particular, how mathematics informs the tuning of those notes—exactly which frequencies we choose to put on a particular scale. We'll also discuss the profound impact these small changes have on composition, because this lecture also relates to the coevolution of scale tunings and musical composition in Western classical music.

Brief Review

- In the last lecture, we used math in the overtone series to inform our choice of notes on a scale. We used two key methods for choosing notes: First, we talked about choosing notes based on the overtones of a single fundamental, a fixed reference point. The second method was using overtones of one note to inform the next note and then using the overtones of that note to get the next—bootstrapping.

- Remember that objects vibrate in different modes, and we now know very well that those frequencies are in a ratio of 1 to 2, to 3, to 4, to 5, and the wavelengths are in a ratio of 1, 1/2, 1/3, 1/4. What we learned from the overtone series is that to go up an octave, we multiply the frequencies by 2, and to go down an octave, we divide by 2. We also know that to go up a fifth, we multiply the frequencies by 3/2.

Just Tuning Revisited

- Let's start by finding the frequencies in A-major scale. Where exactly should you put your fingers on the strings to play in A-major scale? We're going to come up with three different answers, two of which we discussed in the last lecture: just tuning and Pythagorean tuning, which is the bootstrapping idea. The third answer is equal-tempered tuning, which is modern piano tuning.

- As we said, just tuning is great for some instruments, such as bagpipes and sitars. It's used in some musical traditions where there were no modulations, no key changes. Pianos are never in tune, in part because we try to modulate. Essentially, bootstrapping is modulation. We start at the low A and we take the fifth of that, and then we modulate into that key to find the fifth note in that.

- The problems with tuning pianos gave rise to a slow evolution of tunings that mirror changes in composition. But if we use just tuning, no evolution of tunings is needed. We can always just tune things exactly with the fundamental.

- We can calculate the exact frequencies on a just scale using the overtones. As we've seen, we can generate as many notes as we want by taking the overtones and bringing them into the correct octave. In this way, we generate a seven-note scale. The problem with this scale is that although the notes are in tune with A (the fundamental), they aren't in tune with each other.

- Let's look at the F-sharp. If you remember, the F-sharp we got originally for a pentatonic scale was a different F-sharp; it was not in our 12-note scale. This F-sharp is in our 12-note scale.
    - The just-tuned F-sharp is related to the ratio of 5/3. The frequency should be 5/3 multiplied by the fundamental. That's about 1.667.

    - The overtones of B include an F-sharp, and we've figured out the frequency for B. B is related to the ratio 9/8; the first overtone of B will be 2 × 9/8, which is 9/4. The second overtone will be an F-sharp. It's actually an F-sharp^6, and that is 3 times the fundamental B, or 3 × 9/8, or 27/8.

    - To get an F-sharp^5, we divide the F-sharp^6 by 2 and get 27/16, or about 1.688.

- This is why playing in an ensemble is difficult. If you see an F-sharp on your page, then you actually need to play a different note if everyone around you is in the key of A (F-sharp at 1.667) than if everyone is in the key of B (F-sharp at 1.688).

- What's going on here is a fundamental problem. We can't create the scale so that the overtones of one note exactly match up with the other notes. Even with only three notes, the additive structure of the overtones conflicts with the multiplicative structure of intervals.

- This isn't just a problem with pianos or other fixed-tuned instruments. The mathematics tells us that no instrument can play A, B, and F-sharp so that they're all in tune with each other.

Pythagorean Tuning

- Of course, Pythagorean tuning is named after Pythagoras of Samos. Followers of his belief system thought that all numbers could be written as fractions (rational numbers). They didn't believe in the existence of such numbers as $\sqrt{2}$, which cannot be written as a fraction. The Pythagoreans further believed that musical notes were pleasing together if the ratio of frequencies was a fraction with small numbers. Thus, they prized the octave, with its 2:1 ratio, and the fifth, with its 2:3 ratio. The fifth is the key in Pythagorean tuning.

- The goal in Pythagorean tuning is to keep the fifths exactly in tune. When we tune Pythagorean scales, what we're doing is walking around the "circle of fifths." From one note to the next, we're going up a fifth, which is seven half-steps or seven keys on a piano. That's also going from the fundamental to the fifth note on the major scale.

- Let's take a trip around the circle of fifths: We start on A, and the fifth note on the A-major scale is E; the fifth note on that scale is B, then F-sharp, and then C-sharp. We can also go backwards. The fifth below A is D; in other words, an A is the fifth note on the D-major scale.

- Again, to go up an octave is multiplying by 2, down an octave is dividing by 2, and up a fifth is multiplying by 3/2. If we start out on A^4 at 440 Hz, let's think of that as 1 and work in relation to that frequency. We're looking to put scale notes in between A^4 and A^5 to tune our scale. Really, that's between 440 Hz and 880 Hz, which is the octave above an A^4, but in terms of our fundamental at 1, we're thinking of getting numbers between 1 and 2.

- As an example, let's see if we can find the C-sharp on our scale. To get a C-sharp, it looks on our scale as if we have to go up 4/5. To do that, we need to multiply by $3/2^4$. That doesn't get us to the correct C-sharp; it gets us to C-sharp^7.

- We have to go back two octaves to get down to C-sharp^5, the one we want. That's dividing by 2 each time. Now that we've found our C-sharp, we should multiply by $3/2^4$ and then divide by 2^2. The answer comes out to about 1.266.

- If we work through all the details, we can come up with an A-major scale where the fifths are tuned exactly perfectly. For comparison, the C-sharp on the just scale was 4/5, or 0.8, whereas the C-sharp on the Pythagorean scale is 64/81, or about 0.79. Those are quite different.

- If we think about this on a piano, we could start tuning at A and go up to 12/5. As we saw in the last lecture, the problem with piano tuning was that going up 12/5 was different from going up seven octaves, although it ended up at the same key: $3/2^{12}$ was not the same as 2^7. That gap between 12/5 and seven octaves is called the "Pythagorean comma." Even though the gap is small, it has led to significant changes in composition.

"Tempering" the Gap

- There are multiple ways to "fix" the gap, although there are no exact solutions. We can, for example, put the entire gap in the last fifth. We can spread the gap out among more of the fifths so that they're not perfect 3/2 ratios but the last fifth doesn't have such a large gap.

- When we spread the gap out more, keyboards will sound good in a greater number of different keys. Spreading the gap out gave composers more flexibility, resulting in an evolution in music—from the Baroque, to the Romantic, to 20th-century music.

- By 1900, modern tuning on a piano spread the gap out completely evenly. Every fifth on a modern-tuned piano, or an equal-tempered piano, is equally out of tune. Remember, though, the problem here is that $3/2^{12}$ is not the same as 2^7.

- Mathematically, we can solve this by finding a solution to $r^{12} = 2^7$. We take the 12th roots of both sides to get an answer of $r = 2^{7/12}$. When we plug that into a calculator, we get about 1.4983, close to 1.5. That value of r is irrational; it cannot be written as a fraction. The Pythagoreans didn't think that such numbers existed.

- Let's look at another perspective: If we divide the octave into 12 equal half-steps (multiplicatively, not additively), then each one should have a ratio of $2^{1/12}$. If each half-step had a ratio of $2^{1/12}$, then seven of them will have a ratio of $2^{7/12}$, which is the number we just got.

- Let's use this number to construct the scale. Again, we start with a 440-Hz A. If we want to find C-sharp, that's up four fifths and down two octaves. We go up four fifths from the 440-Hz A—that's our key number r^4—and then we divide by 4 to go down two octaves. When we do that calculation, we get roughly 554.36 Hz. Remember that a just C-sharp was at 5/4 of the fundamental, and that gives us a 550-Hz note.
  - How different are the just C-sharp, the Pythagorean C-sharp, and the equal-tempered C-sharp?

  - With the A at 440, the just C-sharp was 550 Hz exactly. The Pythagorean C-sharp was roughly 556.89, and the equal-temperament C-sharp was 554.36. The Pythagorean C-sharp sounds sharper and higher in pitch. The equal-tempered C-sharp is in between the just and the Pythagorean.

Why 12 Keys per Octave?

- Why does the equal-temperament system use 12 notes and not a different number? Would another number work better?

- Let's build a scale with equal spacing, which gives us maximum flexibility. If we have equal spacing, we can modulate into any key we want. The notes that we definitely want in our scale include the octave, the fifth, the fourth, and the major third.

- Suppose the fundamental is at 100 Hz. In just tuning, which sounds best, the octave will be double that, the fifth will 3/2 of that, the fourth will be 4/3, and the major third will be 5/4. Those are the notes that are perfectly in tune with the overtones, and those are the key fractions that we need to look at: 3/2 is 1.5, 4/3 is $1.\overline{3}$, and 5/4 is 1.25. For what values of n will a scale with n equally spaced notes include notes that are close to these key fractions?
    - Let's work by trial and error. If we want to add a note between the fundamental and the octave, that note needs to be exactly halfway in between. If we do the calculation, we get a "tritone," the augmented fourth.

    - If we want to add two notes between the fundamental and the octave, we get close to the major third, but we don't get close to the fourth or fifth. In fact, as we continue to add notes, we get close to some of the key ratios but not all three.

    - It's not until we reach 12 notes that we get very close to the three key ratios we want to hit. If we want equally spaced notes and we want to match up with the most prominent overtones (the major third, the fourth, and the fifth), we need at least 12 notes.

What about *n*-Note Scales?

- Let's try to resolve this in a different way, using some mathematical theory. If we were to place n equally spaced notes (multiplicatively spaced), they would have relative frequency of 1 (the fundamental) and then $2^{1/n}$, $2^{2/n}$, $2^{3/n}$, up to $2^{n/n}$, which is 2, and we would be at the octave. Remember for our 12-note scale that each half-step was $2^{1/12}$; this is just a generalization of that for n notes.

- Let's say that the k^{th} one of those is close to the fifth: $2^{k/n} \approx 3/2$. We can multiply both sides of that equation by 2, then we take the $\log_2$ (log base 2), and what we get is that the $\log_2$ of 3 needs to be very close to $k/n + 1$. In other words, we need a fraction k/n that is very close to $\log_2(3) - 1$. We're looking for a rational approximation of the $\log_2(3)$.
    - o Here, "rational" means "ratio" or "fraction," and "rational approximation" means that we're trying to get close to an irrational number using just rationals (fractions). There's a sort of astonishing connection here to something called "continued fractions" that we'll discuss more in a later lecture.

    - o The $\log_2(3)$ is about 1.58496. To get an approximation of this, we can find the continued fraction. Then we ask the question: Where should we cut off this fraction?

The most famous continued fraction is the "golden ratio," which appears in art and architecture and throughout nature.

- If we stop after two layers, we get 3/2 (1.5), which is not a very good approximation of 1.58496. If we stop after three layers, we get 1.6, which is a little bit better. When we stop after four layers, the value of the continued fraction is 19/12, or 1.5833…, quite close to the actual value of the $\log_2(3)$.

- What would it mean musically to take this as a rational approximation?
  - It would mean that we are approximating $\log_2(3)$ with 19/12, and when we invert the log, we're taking exponentials, so that would mean $2^{19/12}$, or approximately 3. We can divide both sides of that equation by 2, and we get $2^{7/12}$, or approximately 3/2.

  - That's saying that one way to approximate 3/2 is to use $2^{7/12}$. That was the equal-tempered fifth. In other words, this continued fraction just gave us the equal-tempered 12-note system of Western music today.

- What would our musical system look like if we had taken one more layer of fractions? When we do that, we get 65/41, and that gives us $2^{24/41} \approx 3/2$. Musically, we would need 41 keys between one octave and the next, and the 24th of those would be very close to a perfect fifth.

- Here's what the mathematics is telling us: If we want to get a better and better approximation—if we want a note that is closer and closer to the pure fifth of 3/2—we need to keep taking more layers in our continued fraction. As we do that, the advantage is that some notes are much better in tune; the disadvantage is that we have too many notes in each octave.

Why Do Tunings Matter?

- Just tuning, Pythagorean tuning, and equal-tempered tuning aren't actually very different, so why do we care? The answer is that tunings matter because composers use keyboards to compose music, and what sounds good on a keyboard depends on the tuning. Before we had equal-tempered tuning, composers chose different keys for different moods, because when the gap wasn't spread out evenly, different keys actually sounded different.

- In Western music, from about 1500 to 1900, the Pythagorean comma was spread out more and more equally. Tuning systems included the meantone and quarter-comma meantone, Werckmeister III, Kirnberger III, well-tempered, quasi-equal tempered, and finally, equal tempered. Pianos were tuned equal tempered by the late 1800s.

- Interestingly, the fact that guitars have frets that go across all of the strings made it necessary to use equal-tempered tuning back in 1500, when guitars were first made. There is almost no piano-and-guitar music in the period 1500–1900, partially because a guitar and piano used different tuning systems and would have been out of tune with each other.

- A quick tour of the history of Western classical music shows a coevolution of tunings and composition. Tunings moved from not spreading out the Pythagorean comma at all to evenly spreading it out. As that happened, composers gained access to increasing numbers of keys that sounded good, and over time, they lost the devotion to a single key and moved toward all keys being equal. Understanding Western classical music requires understanding the mathematics of the tunings that underpin the compositions.

Suggested Reading

Barbour, *Tuning and Temperament.*

Benson, *Music: A Mathematical Offering*, chapters 4–6.

Duffin, *How Equal Temperament Ruined Harmony*.

Dunne and McConnell, "Pianos and Continued Fractions."

Forster, *Musical Mathematics*, chapters 9–11.

Harkleroad, *The Math behind the Music*.

Loy, *Musimathics*, vol. 1, chapter 3.

Wright, *Mathematics and Music*, chapters 4–6, 11–12.

Questions to Consider

1. How do subtle changes in tuning mirror changes in compositional styles, and how does mathematics inform those changes?

2. Explain the choices a piano tuner has to make when tuning the fifths on a piano.

Dissonance and Piano Tuning

Lecture 6

We now understand a single note and how it vibrates, and we understand the sequence of notes called a scale. In this lecture, we'll talk about two or more notes played at the same time—a chord—such as some of the chords we hear in Bach's Chaconne. Not all combinations of notes sound good, and that's also the topic of this lecture: dissonance, the rough, slightly unpleasant sound we hear when music is played out of tune. We'll look in particular at the mathematics of dissonance.

Defining Dissonance

- Dissonance is a discordant sound that's produced by two or more notes played together that sound displeasing or rough. It's a sort of tonal tension. Sometimes, composers will build in dissonance for a tense moment and then release it by finding consonance.

- Dissonance has both a physical and a cultural component, and what's considered dissonant changes over time. The Pythagoreans thought that notes in consonance had a ratio of frequencies with small numbers and that dissonance was anything but that.

- Herman von Helmholtz, a German mathematician and physicist, gave a mathematical description of dissonance. For him, it involved the overtones, as well as the fundamental notes. His explanation was that dissonance was "beats."
  - The *Grove Dictionary of Music* calls this kind of beats "an acoustical phenomenon useful in tuning instruments, resulting from the interference of two sound waves of slightly different frequencies."

  - What we mean by that is a throbbing sound produced when two notes are played slightly out of tune with each other.

- We hear an example in two Ds, one played on an open D-string and one played on a G-string. The timing of the throbbing sound can be changed by moving the notes closer to each other. As the notes get closer, the beats slow down. As the notes get further apart, the beats speed up. If the notes are exactly the same or very far apart, we hear no beats at all—either just one note or two separate notes.

The Beat Equation

- The beat equation is a trigonometric identity:

$$\sin(a)+\sin(b)=2\sin\left(\frac{a+b}{2}\right)\cos\left(\frac{a-b}{2}\right)$$

- Let's look at an example to make sure we understand this equation. We'll let a be $5t$ and b be $3t$. Now the equation is:

  $\sin(5t) + \sin(3t) = 2\sin(4t)\cos(t)$

  - Looking at a graphical representation of this equation, notice that we use point-wise addition to solve the left side. We pick some x value and look at the height of $\sin(5t)$, and then for the same x value, we look at the height of $\sin(3t)$. Adding those two heights gives us the solution for $\sin(5t) + \sin(3t)$.

  - For the right side of the equation, we do point-by-point multiplication. For a particular value of x, we look for the height of $\sin(4t)$ and the height of $\cos(t)$, and we multiply those values. If either of those functions is 0, then when we multiply, the product will be 0. The last thing we have to do on the right side of the beat equation is multiply by 2 to get the solution for $2\sin(4t)\cos(t)$.

  - The graphs for the left and right sides of the equation are exactly the same.

- We can also explore a musical example. Here, let's take *a* to be (450 × 2π × t) and *b* to be (440 × 2π × t). These represent a 450-Hz sine wave and a 440-Hz sine wave.
  - According to the beat equation:

$$\sin(450 \times 2\pi \times t) + \sin(440 \times 2\pi \times t) = 2\sin(445t \times 2\pi) + \cos(5t \times 2\pi)$$

  - For any particular *x*, when we add the value of the 450-Hz wave and the value of the 440-Hz wave and we listen to the result, we can hear dissonance.

  - If we zoom out from the graph, we can see the beats in the wave form.

  - On the right side of the equation, when we multiply $\sin(445t \times 2\pi)$ and $\cos(5t \times 2\pi)$, it seems almost as if the cosine becomes an envelope into which the sine wave fits.

  - The sine wave is going back and forth between +1 and –1. When it gets to +1 and we multiply by the cosine, we simply get the value of the cosine. When it gets to –1 and we multiply by the cosine, we get the negative value of the cosine. That's why this function appears to be wavering back and forth inside the cosine envelope. The cosine term is producing the beats and telling us how fast the beats are.

- The beat equation tells us that if we have *a* and *b*, where *a* is slightly larger than *b*, we should get $a - b$ beats per second. But from the formula, it looks like we should get $(a - b)/2$ beats per second. Why the difference?
  - It's true that the cosine has a frequency of $(a - b)/2$. But every full wavelength of the cosine has to go down and back up, and when it does that, that single wave form of the cosine produces two beats.

  - The sound is beating twice during one full cycle of the cosine. That tells us that our conjecture was correct. It's not $(a - b)/2$

beats per second, as the cosine would indicate. It's double that, giving us a beat frequency of $a - b$.

- What happens when $a = b$? When we solve the equation, we find that the cosine term disappears, and there are no beats.

- In the last lecture, we learned that subtle changes in pitches have led to big changes in composition. In practice, how exactly do we make subtle changes in pitch? The answer is by using beats. A demonstration of piano tuning helps us understand.
    - When you tune a piano, you need to adjust the tension on the strings. As we know, adjusting the tension affects the frequency at which the strings vibrate.

    - As you bring two strings into tune—as the two frequencies get closer—the beats slow down. When the two frequencies are equal ($a = b$), we're left with two copies of the sine, and the beats disappear. This process is called "tuning a unison."

A Question about Beats

- If you hear two notes played that are close in frequency, your brain hears the right side of the beat equation. If you hear two notes played that are far apart, your brain hears the left side of the equation; it hears two notes, but you actually can hear beats when an octave is played out of tune.

- We originally thought that when the frequencies were far apart, we wouldn't hear beats, but we do, and it seems as if the beat equation doesn't explain this.
    - If we played an A at, say, 442 Hz and another at 220, the beat equation tells us that we should hear the difference between them. We should hear 222 beats per second if we're hearing the right side of the equation. But we're not hearing the right side; we're hearing the left side. We're hearing two notes, so why are we hearing beats?

- To answer this question, we have to go back to the overtone series. If we're playing one note at 220 Hz, we're hearing its overtones, that is, all the multiples of 220: 440, 660, 880, and so on. If we're playing the higher note, 442, we're hearing its overtones: 884, 1226, and so on.

- The beats aren't coming from the fundamentals, which are 200-some Hz apart. The beats are coming from one of the harmonics of the lower A (440 Hz) and the fundamental of the higher A (442 Hz).

- We're hearing both of those frequencies, which are very close together, and because of those frequencies, we hear beats.

- This problem of beats with notes played far apart doesn't appear only with octaves; it also happens with other intervals. The key observation is that if two of the overtones of two notes are close, we will hear beats.

- Let's look at a fifth: an A at 440 Hz and a note at 293 Hz that is close to a D. The A is the fifth note on the D-major scale, and that's why it's an interval of a fifth. If we play A 440, we hear the fundamental at 440; we hear the second harmonic at 880, and then 1320, and 1760. For the note that's close to the D, we hear the fundamental at 293; we hear the second harmonic at 586 and the next one at 879, which is close to the 880 in the first overtone of the 440 A. Listening to those notes, we could hear beats.

- At what frequency should we put the D in order to completely eliminate the beats?
  - Let's put the D at x Hz. If we play something at x, we know its overtones will be $2x$, $3x$, $4x$, and so on. It's the $3x$ that we want to try to match up with the 880, the first overtone of the A. That gives us an equation to work with: $3x = 880$. Solving, we find that we should put the frequency of this D at 279 1/3 Hz.

- We know that a perfect fifth should have a ratio of 3/2. If we divide 440 Hz by the lower frequency, which is 293 1/3, we get 3/2.

Tuning Fifths

- What we just calculated was this: If we want a Pythagorean or a just-tuned tuning system—if we want the fifth to be perfect—we should aim the D for 293 1/3 Hz. But on a modern piano, we don't want the fifths to be perfectly in tune; we need them to be slightly narrower. Instead of 3/2, the ratio needs to be $2^{7/12}$.

- We need to determine the correct frequency for the D, and then, we use that and the beat equation to figure out how fast the beats should be when we play that D with a 440 A. We then use that answer to tune the D.

- The first step is to find the frequency for a correctly tuned equal-tempered D. That means adjusting the fifths so that the octaves will work out when we get to the top—spreading that Pythagorean comma completely evenly around all the fifths.
    - Remember, the half-steps were tuned to $2^{1/12}$, and the fifths were tuned to a ratio of $2^{7/12}$. If the A is tuned to 440, then the D should be tuned to some frequency z so that $z(2^{7/12}) = 440$.

    - Solving, we find that the D should be tuned approximately to 293.66 Hz.

- The next step is to figure out how fast the beats will be if we tune the D exactly to that value.
    - If we play the 440 A, we'll hear all of its overtones, including the pure 880. If we play the equal-tempered D at 293.66 Hz, its second overtone (third harmonic) will be at approximately 880.99 Hz.

    - The beat equation told us that the difference between those two would give us the frequency of the beats. That difference is about 0.99, just slightly under 1 beat per second.

- The math tells us that if we want equal-tempered fifths, we should tune the fifths so that they're not perfect, but they're just narrow enough that we hear 1 beat per second.

- Essentially, we now know everything we need to know in order to tune a piano. Of course, many piano tuners today use electronic tuners, and in fact, some of them don't tune in an equal-tempered manner because some performers don't want equal-tempered tuning for particular pieces.

© iStockphoto/Thinkstock.

Piano tuning is an art, and the truly great tuners work by ear, not with an electronic tuner.

- The octave, the fifth, and the fourth are called perfect intervals, and greater precision is required in playing them. If they are played a little bit out of tune, the beats can be heard.
    - This becomes important in some music. For example, much of the music of Aaron Copland invokes an open American spirit. The way he achieved that open feeling was by using open chords—these perfect intervals.

    - An excellent example can be heard in a piece written by Copland called the *Fanfare for the Common Man*.

A Mathematical Coda

- Virtually everything we've done in this lecture relies on a single equation, the beat equation. We can walk through a proof of this equation that requires just a few trigonometric facts and some basic geometry.

- The proof implies that when two notes are just slightly out of tune, we should get dissonance. From this, we know why musicians have to be more careful with some intervals than with others and why piano tuners use dissonance to tune in an equal-tempered system.

Suggested Reading

Benson, *Music: A Mathematical Offering*, chapter 1.

Fischer, *Piano Tuning*.

Forster, *Musical Mathematics*, chapters 2, 5 (on the subtleties of vibrating piano strings).

Loy, *Musimathics*, vol. 1, chapter 6.

Questions to Consider

1. How does mathematics explain the "beating" phenomenon we hear when two notes of similar frequencies are played?

2. Why is it more important for musicians to play octaves, fifths, and fourths in tune than other intervals?

Rhythm—From Numbers to Patterns

Lecture 7

In all of our lectures so far, we've talked about a single topic, pitch. In this lecture, we will turn to rhythm. The term is difficult to define, but we can think of it as a regularly ordered pattern of durations and strengths of notes. Imagine music without rhythm or, more precisely, music without any change in rhythm. There would be no drama, no suspense, no life to the music! At the other end of the spectrum, American composer Steve Reich's work *Clapping Music* shows that it's possible to have music with only rhythm. In this lecture, we'll see how composers use rhythm as a musical tool to add interest and emotion to their work.

Rhythm in Poetry

- Rhythm plays a role not only in music but also in poetry because words form a rhythm. Indian poetry, for example, provides some interesting connections between rhythm and mathematics.

- Indian poetry was traditionally written in Sanskrit, and it has two types of words: those with short syllables and those with long syllables that are exactly twice as long as the short ones.

- Pingala was an Indian poet who lived several centuries before the Common Era. He asked an interesting question: How many different ways are there to put short and long syllables together to get a line of a given length? If we're thinking musically, we could ask: How many different ways can we put one- and two-beat notes together to get a rhythm n beats long?

- Interestingly, the answer to this question follows the pattern of the Fibonacci numbers. For a one-beat rhythm, there is just 1 choice; for two beats, 2 choices; for three beats, 3 choices; followed by 5, 8, 13, and so on. Why is it that the next number in the pattern is the sum of the two previous numbers?

- Here's another question: To get the answer for *n* beats, do you have to figure out all the choices before it? The answer is no. Binet's formula gives us the number of ways of using short and long syllables to form an *n*-beat phrase.

Western Musical Notation

- In Western musical notation, a whole note represents four beats; a half note is two beats; a quarter note is one beat; an eighth note is half a beat; a sixteenth note is a quarter of a beat; and so on. Music also uses rests for silences of the same length of any of these notes. Further, there is notation for notes that aren't fractions with a power of 2, such as triplets or quintuplets. A dot added after a note or a rest symbol adds half the length of the note or rest.

- The time signature, which looks suspiciously like a fraction, tells us how many notes are in a particular measure. For example, with 3/4 time, each measure has three beats (the top number), and each beat is one quarter note long. The time signature 4/4 is called "common time." It's four beats per bar, and each beat is a quarter note long.

- There are many standard time signatures: 3/4, 2/4, 2/2, 6/8, and so on. Notice that the top number can be any natural number, but the bottom number must be a power of 2. Modern composers sometimes make use of stranger options, such as 7/16.

Hemiolas

- The time signatures that involve groups of six, such as 6/8 time, are mathematically interesting. The number 6 is the product of two prime numbers, 2 and 3. We can think of 6 as two groups of 3 or three groups of 2.

- Musically, 6/8 sometimes feels like triplets (1-2-3, 1-2-3) and sometimes feels like three 2s, (1-2, 1-2, 1-2). The song "I Like to Be in America" from Leonard Bernstein's *West Side Story* groups the beats in both ways.

- A "hemiola" is a particular musical figure heard in a piece in which every six beats are usually grouped into two groups of 3, and then the hemiola comes when those six beats are grouped into three groups of 2. Hemiola means "one and a half," and that's the ratio of the length of the groups, groups of 3 to groups of 2.

- The hemiola is used extensively in Western classical music, especially by Handel but also by Brahms and Dvořák. The effect of using a hemiola is that it interrupts the normal flow, catching the listener a bit off guard. We hear an example from Handel's *Water Music* of 1717.

Polyrhythms

- Polyrhythms occur when two parts are playing in different rhythms. Chopin's *Fantasie Impromptu*, which we heard at the beginning of this lecture, uses polyrhythms to give the music an agitated, unsettled feeling.

- We hear a polyrhythm of four versus two written in 4/4 time. We can think of this as four quarter notes in one part and two half notes in the other. With an example of three versus two, we hear that the two parts don't quite fit together. This is sort of like a hemiola, except instead of being played sequentially, the notes are played simultaneously.

- One of the more famous examples of polyrhythm comes in the third movement of Tchaikovsky's Piano Concerto No. 3. The time signature is 3/4, three quarter notes per bar. The piano part is played in 3s (1-2-3, 1-2-3, 1-2-3), but the part for the strings is played in 2s (1-2, 1-2).

Polyrhythms are used throughout Western music and in African drumming.

- Why do composers use polyrhythms? They lend a sense of instability to music. Listeners get the feeling of turning around or being unsettled before the music resolves back into a normal rhythm. The two rhythms act as two competing forces, introducing rhythmic tension to the music. Finally, that rhythmic tension is released when the rhythms fall back into line.

Calculating Polyrhythms

- Calculating polyrhythms is similar to adding fractions. With a rhythm of three versus two, what we're doing in one part is dividing something into three equal parts. In the other part, we're dividing the same thing into two equal parts. To do this, we need a sort of common denominator. We need some sort of equal-spaced piece that we can divide the measure into.

- With two versus three, we divide the bar into six equal parts. That's the right-sized piece that allows us to split the bar into groups of three and groups of two. With a rhythm of p versus q, we need to divide the time into the least common multiple of p and q. That's exactly what we would do if we were adding the fractions $1/p$ and $1/q$.

- Let's consider a rhythm of three versus four. The easiest time signature to use for this rhythm would be 12/8, but we could also write the same pattern in 6/8. We could even use triplet notation to write it in 4/4 or 2/4.

- Chopin's *Fantasie Impromptu* is written in 4/4 time. There are four quarter notes in each bar. The right hand in this piece plays sixteenth notes (1-2-3-4, 1-2-3-4, 1-2-3-4 in every beat), and the left hand plays triplets (1-2-3, 1-2-3, 1-2-3, 1-2-3 in every beat). The tempo marking for this piece is *allegro agitato*, and we can hear the agitation caused by this polyrhythm.

Complicated Polyrhythms

- Why and how would you play such rhythms as three versus five or four versus seven? For three versus five, you would need 15 divisions. For four versus seven, you would need 28 divisions.

- Chopin used some of these more exotic polyrhythms. In a span of just five bars in his Nocturne Opus No. 3, he uses three versus five, three versus seven, three versus eight, three versus one, and more.

- Another extreme example comes from the Grieg Piano Concerto, composed by Edvard Grieg. This example is in the first movement, in the cadenza, where the orchestra drops out and leaves just the solo pianist. In this section, the left hand is making runs of seven notes per beat, and the right hand is playing a melody plus eight notes per beat. If we wanted to do this exactly, we would have to divide each beat into 56 pieces.

- The second movement of Charles Ives's Fourth Symphony takes polyrhythms to their logical extreme: Two different parts of the orchestra are in two completely different rhythms. In fact, two conductors are usually used to perform this piece.

Combining Rhythm and Pitch

- So far, we've heard rhythmic patterns, but if we add pitch, we can do more interesting things. Let's return to a single instrument and think about playing a rhythmical pattern that is five notes long.

- We could add to those notes a pattern of three pitches. Notice that the notes and the pitches don't match up. As these patterns are played together, the music sort of turns around and takes a while to get back to the beginning, almost like two gears of different sizes turning together. We know that the pattern repeats every 15 notes (three measures) because 15 is the least common multiple of 3 and 5.

- George Gershwin's piece *Rhapsody in Blue* makes use of these repetitious phrases. We can hear the sense of turning and instability, followed by a resolution when the rhythms and pitches line up again.
  - Another example from *Rhapsody in Blue* has six notes per bar with an eight-note scale; the pattern repeats every 24 notes (four measures).

  - Notice that Gershwin builds up rhythmic tension with this pattern of unmatched rhythms and pitches; he then pauses and hits listeners with a chord full of dissonance.

- There are many examples in music where we can pull out these kinds of rhythmic patterns. For example, in the opening of *Till Eulenspiegel's Merry Pranks* by Richard Strauss, there's a seven-beat phrase played by the horn, but the piece is in 6/8 time; the phrase doesn't repeat for 42 notes. In Olivier Messiaen's *Quartet for the End of Time* is an extreme example, with competing patterns that are 17 and 29 notes long.

A Musical Proof

- Let's close by looking at a mathematical question we can answer with musical notation. Does the following sum go to infinity: 1/2 + 1/4 + 1/8 + 1/16 + 1/32…? The answer is no. If we keep adding numbers like this forever, we get 1. We can prove this answer with rhythms.

- Let's think about writing a measure of music in 4/4 time; that's four quarter notes per measure. We start with a whole note, which takes four beats, so we have one measure. Mathematically, we have one note that has four beats that equals one measure, so that's 1 = 1.

- We replace that whole note with two half notes, which are two beats each. We still have one whole measure, but now, mathematically, we're looking at two half notes equals one measure: 1/2 + 1/2 = 1.

- Next, we replace the second half note with two quarter notes, which are one beat each. We still have one whole measure, but now mathematically, we have 1/2 + 1/4 + 1/4 = 1.

- The pattern here is as follows: Each time, we replace the last note with two notes, each of which is half as long as the one we replaced. That leaves the total length of the rhythm exactly the same—one measure long. If we continued this process mathematically into infinity, we would prove 1/2 + 1/4 + 1/8 + 1/16 + 1/32 + … = 1.

- Mathematically, this is called a "geometric series." Each term is equal to the previous one multiplied by a fixed constant, in this case, 1/2.

Suggested Reading

Magadini, *Polyrhythms*.

Wright, *Mathematics and Music*, chapters 2, 8.

Questions to Consider

1. How is the mathematics of adding fractions (common denominators) related to the musical idea of polyrhythms?

2. How do composers use the idea of least common multiples to create a sense of instability or turning?

Transformations and Symmetry

Lecture 8

So far in this course, we have discussed the mathematics of the musical experience. We started by coming to understand a single note and varying its pitch; then we combined notes into chords and scales; and finally, we added rhythm. Now, we're ready to start composing music. In this lecture, we will discuss parallels between musical and mathematical transformations. A transformation is the process by which one expression is converted into another that is equivalent in important respects but differently expressed or represented. Transformations add beautiful structure to both mathematics and music.

Geometrical, Functional, and Numerical Transformations

- Geometrical transformations are, perhaps, the simplest to begin with. Let's consider any figure in a plane and let $r(x)$ be the reflection of that figure over the x axis. We actually have two transformations here: reflecting the figure over the x axis (the reflection transformation) or doing nothing (the zero, identity, or "do nothing" transformation).
  - o The operation in this case will be one transformation followed by the next (composition). If we do nothing and then do nothing again, the net result is doing nothing. If we do nothing and then reflect, the net result is a reflection. If we reflect and do nothing, again, the result is a reflection. But if we do a reflection and another reflection, the net result is the identity transformation.

  - o A "group table" is an array listing all the possible outcomes from our operation of one transformation followed by the next. In our case, this will be a 2 × 2 table.

- Some functional transformations are similar to geometrical transformations. For the function $f(x)$, the identity transformation would be doing nothing, but we could also do a negation—we could look at $-f(x)$. This flips the function over the x axis. Again, we have two transformations here—the identity transformation and the negation transformation—and the operation is composition.

- Of the possible four pairings with this transformation, the interesting one is taking the negation and the negation again, which gets us back to the identity.

- Note that the transformations are not the graphs but how the graphs are changing. The transformation isn't the picture but how the picture changes.

- Let's think about addition with regard to the set of even and odd numbers. The set consists of all the even numbers and all the odd numbers. We can construct a group table for transformations of this set that tells us: even + even = even, even + odd = odd, odd + even = odd, and odd + odd = even. The identity here is even because it operates on all the other elements without changing them. Both evens and odds are their own inverses.

- Mathematicians call this group Z_2, and we think of it as addition modulo 2. The question to think about here is: If we divide by 2, what is the remainder? The answer is that the remainder is always either 0 (for evens) or 1 (for odds). We can represent all even numbers with 0 and all odd numbers with 1. The identity in this group is 0, and each of the elements is its own inverse.

Musical Transformations

- With a musical transformation, we could have no change in a melody (the identity), or we could do an inversion. We can think about this as a group because we have two transformations, and we can look at the structure of the group. Remember, we are not looking at the melody itself but how the transformation changes the melody.

- Each of Bach's 14 canons on the Goldberg ground was written using transformations. The canons give us insight into how Bach used transformations to create increasingly complicated music. The canons are based on an aria from the *Goldberg Variations* and are all written on a single sheet, totaling perhaps six or seven minutes of music. The underlying ideas are very similar to group theory.

- Geometrically, an inversion is flipping something over the x axis; functionally, it's multiplying by a negative sign; and musically, it's flipping all the intervals up. In canon 3, Bach puts this together as follows: the original melody (the identity transformation); the inverted melody, four notes after the melody; and a repetition of the sequence.

- In addition to musical inversion—playing a melody upside down—we can also have a retrograde—playing a melody backwards. If we construct a group table, we again have two transformations: the original melody (the identity) and the retrograde. The retrograde followed by the retrograde results in the original.
    - The corresponding transformations in geometry are reflecting around the y axis instead of the x axis.
    - In terms of functions, instead of looking at $-f(x)$, this transformation looks at $f(-x)$.
    - Bach put together retrograde transformations in canon 1. We hear the theme twice, then the retrograde, then both played together.

- Can we do both the retrograde and the inversion? There's a problem when we try to construct a three-element group. We don't have a transformation for a melody that is played retrograde and then inverted. Mathematically, we would say that this operation is not closed. We can fix the problem by adding an element that is exactly a retrograde inversion.

- In canon 2, Bach puts together not the original melody but an inversion of the melody and then adds the retrograde inversion. We hear the inversion played twice, the retrograde inversion played twice, and both played together.

- In canon 5, Bach introduces a new melody and its inversion. If you listen closely, you can hear two baselines in this piece. One is the main theme, and the other is the main theme's inversion.

Group Theory

- A group is defined as a set and an operation. An operation takes two things in the set as input and produces one thing in that set as output. For us, the operation was composition—doing one transformation and another.

- A group must also have four other properties: (1) It must be closed (the output must always be in the set); (2) it must have an identity (one element acts like 0 does in addition; it leaves all the other elements alone when we do the composition); (3) it must have inverses (there must always be a way to get back to the identity); and (4) it must have associativity (the grouping of elements doesn't matter).

- Let's consider addition with the natural numbers: 1, 2, 3, ….
    - We can construct a closed group table for this set, but there's no identity. There's no way to solve the equation $a + x = a$. We can fix this problem by adding 0 to the set, which means we're working with the set of whole numbers.

    - The table associated with the whole numbers is closed, has an identity, and has associativity, but there's a problem with inverses. There's no way to solve the equation $a + x = 0$. We can fix this by adding negative numbers to the set, which means we're working with the set of integers.

    - With addition of integers, we get a group; in fact, we get what's called a "commutative group."

- Let's now return to group Z_2. We've seen many copies of this group: geometrical (reflections), functional (negating functions), evens and odds (addition mod 2), and musical (inversions). All of the tables for these groups have the same structure; we simply change the names of the columns and rows. These are called "isomorphic groups."

- What is Z_2? To answer this, let's look at an easier question: What is 2?
  - The idea of 2 is an abstraction. When you're looking at two apples, what's there? Apples, not 2. In some sense, 2 is what the sets of two eyes, two ears, two fingers, and two apples have in common.
  - The same is true of Z_2. None of the group tables we saw is Z_2; Z_2 is what all of those examples have in common. Both the number 2 and group Z_2 are abstract mathematical concepts—separate from the physical world.

Musical Transformations Revisited

- The four musical transformations we identified earlier (identity, inversion, retrograde, and retrograde inversion) fulfill all the properties of a group. But rather than a 2×2 chart, this is a 4×4 chart; it's clearly not Z_2.
- Considering that this is a four-element chart, we might guess that it would be Z_4. Recall that we're looking here at remainders when we divide by 4. The possible remainders are 0, 1, 2, and 3, but our table has a different structure. In the musical group we're looking at, every element is its own inverse. If we do the inversion and then the inversion, we get the identity. But that's not true in Z_4. If we add the element 1 to itself, we don't get back to the identity, 0. That tells us that these two groups are not isomorphic.
- Let's consider another representation of the same group in geometry. Here, the four elements are: the identity transformation, reflection over the x axis, reflection over the y axis, and reflection over both, or rotation of 180 degrees. That table has the same structure as the one for our musical group.
- In functions, the four elements would be: the identity transformation, $-f(x)$ (reflecting over the x axis), $f(-x)$ (reflecting over the y axis), and $-f(-x)$ (reflecting over both, or rotating 180 degrees). Again, this table has the same structure as the one for our musical group.

- To see the correct numerical example, we have to look at something called $Z_2 \times Z_2$, the "Klein four group." $Z_2 \times Z_2$ is not just numbers; it's actually ordered pairs of numbers, for example, 0 and 1 in one coordinate and 0 and 1 in the other coordinate. The operation here is coordinate-wise addition modulo 2.
    - Let's try (0, 1) + (1, 1). We add the first coordinates, 0 + 1, to get 1. The second coordinates are both 1, but in modulo 2, 1 + 1 = 0. When we add these two elements (0, 1) + (1, 1), we get (1, 0). Is this a group?

    - The answer is yes; this is the same group table as the musical one.

- Interestingly, there are only two mathematical groups with four elements: Z_4 and the Klein four group.

Transpositions

- In music, a transposition is moving up or down in pitch. For example, instead of starting on the key A, we could start on D. That would be transposing up a fourth because D is the fourth note in the A-major scale. When we do this in music, it's easily recognizable as the same tune, played a bit higher.

- A corresponding geometrical transformation would be taking a figure and translating it up. For functions, we could think about adding c to $f(x)$. It's just a vertical translation when we're looking at a graph.

- The group structures of these transformations are interesting. We can find the inverse if we transpose up (we just transpose back down), but transposition has infinitely many elements. In theory, we could keep transposing up as many intervals as we want. At any point, if we wanted to go back down, we could transpose infinitely many steps. The table for this group is infinitely large.

- One way to keep the table from being infinitely large is to work by note names, that is, to work what's called "modulo octaves." This approach treats every A on the piano as if it's the same and every C-sharp as the same.

- Think about transposing up one half-step. If you transpose in this way 12 times, in terms of note names, you get back to where you started. You may start on A and go up 12 times, but you end at A. This is similar to adding 1 to itself but then somehow getting back to the beginning.
  - We actually see something every day that does exactly this: a clock, with the 12 acting as 0. This group is called Z_{12} or "clock arithmetic." In Z_{12}, if we add 10 + 3, that's the same as adding three hours to 10 o'clock; we don't get 13 o'clock but 1 o'clock.

  - Transposition by half-steps is exactly the same thing. We go up 12 half-steps and get back to the beginning.

- What if we transpose up a major third, which is four half-steps? What group would we generate if we did this? This group (M_4) has three transformations: do nothing, up four half-steps, and up eight half-steps. The table for this group has the same structure as Z_3; it's addition modulo 3.

- What happens if we combine M_4 with an inversion? This group is not commutative. The M_4 transformation followed by an inversion is not the same thing as the inversion followed by the M_4 transformation.
  - To find a geometrical representation of this requires us to look at symmetries on an equilateral triangle. The group table for these transformations in geometry is isomorphic to the musical group M_4.

  - Mathematicians call this group the "dihedral group of order 6," and it is the smallest noncommutative group we can construct.

- Some mathematical researchers are working to construct musical versions of these mathematical groups.

Symmetry

- Group theory is seen as the language of symmetry in mathematics. Mathematicians categorize objects by their symmetry group.

- The outline of a violin has only one symmetry. It can be reflected over the x axis, but there is no other way to transform that picture and get back to the same picture. In contrast, the outline of a barbell has three symmetries. It can be reflected over the x axis, reflected over the y axis, or rotated 180 degrees. That means the symmetry group for the barbell is a copy of the Klein four group.

- Among the interesting objects that can be categorized by their symmetry groups are frieze patterns. There are just seven different symmetries on frieze patterns and no more. There are exactly 17 different symmetry groups for a wallpaper form.

- In chemistry, group theory is used to classify crystal structures. In physics, it's used to study subatomic particles.

© iStockphoto/Thinkstock.

All 17 wallpaper symmetry groups appear in the wall mosaics of the Alhambra in Spain.

Augmentation and Diminution

- Yet another transformation in music that has this group structure is augmentation and diminution, that is, stretching and shrinking by a factor of 2. The augmentation of a melody is twice as slow, and the diminution is twice as fast.

- We can think about the group structure involved in these transformations as follows: Augmentation is taking the inverse of diminution. We can also augment as many times as we want, each time stretching the notes twice as long; thus, this group is infinite.

It actually has the same structure as the integers, where the integers serve as powers of 2: 2^n.

- Augmentation and diminution are among the transformations Bach was working on in his 14th canon.
  - This canon has four parts: The top voice is supposed to play the music as written; the next voice is supposed to play it augmented (twice as slow as the original) but also inverted and transposed; the next voice is supposed to play it augmented again (four times as slow as the original) and transposed again; and the bottom voice is supposed to play it augmented yet again (eight times as slow as the original) and also inverted and transposed.

  - Amazingly, the bottom voice is the original melody from the Goldberg variation.

  - As we listen, notice that all the parts are the same in some sense, but they're transformed from the original melody.

Suggested Reading

Benson, *Music: A Mathematical Offering*, chapter 9.

Harkleroad, *The Math behind the Music*, chapter 4.

Questions to Consider

1. What are groups, and how are they used to study symmetry transformations in mathematics?

2. What are the musical transformations, including inversion, retrograde, retrograde inversion, transposition, and augmentation/diminution?

3. How are musical transformations and group theory connected?

Self-Reference from Bach to Gödel

Lecture 9

We can point to many different versions of self-reference: talking about yourself, looking in a mirror, and using references to a work of art in the work itself. Self-reference creates beauty and a bit of strangeness in both mathematics and music, and the self-reference in these two subjects seems somehow similar. In this lecture, we'll discuss different levels of self-reference, from basic, to intermediate, to advanced, looking at both musical and mathematical examples. As we progress on this continuum, we'll see that infinite loops begin to appear, and these loops take on stranger meanings.

Basic Self-Reference

- Basic self-reference in classical music is fairly common. Western composers like to take snippets of earlier parts of a piece and play them later to sort of refer listeners back and keep the parts of the piece connected.
    - One of the most popular pieces of all time does this: Beethoven's Ninth Symphony, just before the "Ode to Joy." After an opening raucous blast, Beethoven intersperses a vigorous cello line with references to earlier movements. The piece quotes earlier parts of itself.

    - Many similar examples of basic self-reference can be found in the Western concert music repertoire.

- Mathematicians use basic self-references all the time in the form of functions—sequences that refer to themselves. Think, for example, about the Fibonacci numbers, a sequence that recurs everywhere—in pinecones, sunflowers, Indian poetry, and much more. Each term in the sequence refers back to earlier terms. In mathematical notation, the sequence is $f_n + 1 = f_n + fn - 1$. The name for this sort of self-reference is a "recursively defined function."

- If we have a function defined on a small interval, say 0, 1, and we want to repeat it—we want it to go on and on and do exactly the same thing—we write $f(x) + 1 = f(x)$. What is $f(2)$? It refers back to $f(1)$; $f(2)$ is $f(1) + 1$, and that should be $f(1)$ according to our equation. What is $f(1.5)$? That refers back to $f(0.5)$; $f(1.5)$ is $f(0.5) + 1$, which is the same thing as $f(0.5)$.

Intermediate Self-Reference

- Basic self-reference in music is a piece referencing itself. Intermediate self-reference in music might be a composer referencing himself or herself. One of the most famous examples of intermediate self-reference in music is Sir Edward Elgar's *Enigma Variations*.

- To understand intermediate self-reference in the work of Bach, we need to know a little bit about music in Germany, in particular, the system for naming notes. In Germany, our B-flat is indicated with a B, and our B is indicated with an H. Thus, in the German system, Bach was able to encode his name in his music. We hear examples in *The Art of the Fugue* and the *Brandenburg Concerto* No. 2.

- Other composers have honored Bach in the same way, by writing BACH into their music. Still others, including Robert Schumann, Franz Schubert, Johannes Brahms, and even Dmitri Shostakovich, have also written their names into their music. In some cases, if the letters don't match up with the notes, they use tricks of musical nomenclatures or puns to accomplish the self-reference. Shostakovich shortened the German form of his name to DSCH, used the fact that S in German notation is E-flat, and arrived at D, E-flat, C, B.

- Here's a fun math version of intermediate self-reference: If you pick an answer at random from the following choices, what is the probability that you will be correct? (A) 25 percent, (B) 50 percent, (C) 0 percent, or (D) 25 percent. We can see the self-reference here: The answer is the probability of picking the correct answer.

- Each answer is equally likely, so that is a 25 percent chance of picking each one. But the 25 percent appears twice; if we pick at random, there is a 50 percent chance of picking 25 percent. Perhaps the answer should be 50 percent. But 50 percent appears only once, and the chance of picking 50 percent is 25 percent.

- Neither 25 percent nor 50 percent could be correct, so the correct answer should be 0 percent. But 0 percent appears only once, and there is a 25 percent chance of picking it. No answer is the correct percent chance of picking that answer.

- A sneaky way to answer this question would be to write in a fifth option: (E) 20 percent.

- In some sense, all differential equations are a form of self-reference. Consider, for example, the differential equation $y' = 0.5y$. This is among the simplest differential equations we could write. The equation asks: How quickly does y change? The answer to that question is: The change is half the size of y.
  - This equation models something like exponential population growth. Think about starting with 100 rabbits, and at every time step, half of them reproduce, so at the next time step, we have an additional 50 rabbits. Then, with 150 rabbits, at the next time step, we get an additional 75 rabbits.

  - The key self-reference here is that the function appears on both sides of the equation; it is a function referring to itself.

- In an earlier lecture, we looked at the wave equation, which is another example of a differential equation (a partial differential equation). Recall that $u(x,t)$ represented how far above the median line the vibrating string went at position x and time t. Notice that $u(x,t)$ occurs on both sides of the equation. It is a form of self-reference; u is related to itself in this particular way.

- Recall that the golden ratio (phi) we discussed in Lecture 5 is like pi, a number with special properties, and it is also rational; it cannot be written as a fraction. We usually think of the golden ratio not in terms of its continued fraction but as 1 + √5/2, the ratio of the sides of pi.

- If we have a rectangle with sides that are exactly on that ratio and we remove the largest possible square, the remaining rectangle has the same proportions as the original. From that, if we remove the largest possible square, the remaining rectangle has the same proportions as the original, and so on.

- If we inscribe a curve inside each one of these squares—roughly a quarter circle in each square—we get a logarithmic curve.

- We have seen that the golden ratio is 1 + √5/2, but we have also seen it as a continued fraction with all 1s. How do we know those two are the same?
  - Let x be the continued fraction with all 1s. The denominator is just another copy of x. It is just 1 plus a fraction with an infinite sequence of further fractions all involving 1s; that is what x is. That means that the whole expression x is actually equal to $1 + 1/x$, the entire expression again. The value x appears on both sides of the equation.

  - To solve, we multiply both sides by x, and we get $x^2 = x + 1$. We then subtract $x + 1$ from both sides, which leaves us with a quadratic equation: $x^2 - (x - 1) = 0$.

  - We then apply the quadratic formula: $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, which gives us phi: 1 + √5/2.

- A geometrical version of self-reference is the Möbius strip. To construct one, we take a long strip of paper, which appears to have two sides—a front and a back—and we twist it once and fasten the ends together. Now, the front side continues on what was the back side, which continues back to the front side. The strip has just one face and one edge.

- If we cut the Möbius strip down the middle and keep cutting all the way around, we end up with a single piece of paper that has two faces and two edges.

- If we cut it in thirds, we get two separate loops that are interlocked.

Self-reference creates beauty and a bit of strangeness in the everyday world and in mathematics and music.

Advanced Self-Reference

- Musical examples of advanced self-reference are rare. One of the few is a "crab canon." This is a duet formed by playing a piece of music forward and backward simultaneously. Bach wrote one in response to a musical challenge posed by Frederick the Great of Prussia in 1746.
  - Written out, a crab canon has two parts, but the second part is the same as the first played in reverse. In the language of transformations, the second part is simply the retrograde of the first part. In Bach's crab canon, the second part reads the music retrograde, or from the end to the beginning.

  - In a "table canon," the second part plays retrograde inversion from the end to the beginning but inverted—upside down. If we were to write out the table canon on one long musical staff, we could then represent an infinitely long musical score by forming a Möbius strip.

  - Canons such as these and those we heard in earlier lectures represent the pinnacle of self-reference in music.

- For advanced self-reference in mathematics, we turn to Kurt Gödel.

- In 1910, Alfred North Whitehead and Bertrand Russell published *Principia Mathematica*. Their goal was to provide axioms that would give a stable base for mathematics, in much the same way that Euclid had done for geometry, except they were trying to do this for arithmetic and all of the theorems that come with arithmetic.

- Whitehead and Russell, along with David Hilbert, sought axioms that were consistent (so that both a statement and its opposite could not be proved) and complete (axioms that proved all the true statements). If a statement was true in their system, they wanted to make sure it could be proved using just the axioms they provided.

- This is a particular view of mathematics that with the right axioms, there might be only two kinds of statements in the world: true statements, all of which would be provable, and false statements, none of which would be provable. Gödel's work ruined the dreams of these mathematicians.

- The liar's paradox is a tricky version of self-reference. Consider the statement s = This statement is false. It can't be true because the statement says that it is false, and that's not possible. But if s is a false statement, then "This statement is false" is false, so the statement must be true. Thus, s can be neither true nor false. Gödel used this kind of self-reference to burst Whitehead and Russell's dream.

- Given a set of axioms strong enough to prove basic arithmetic truths about the natural numbers, Gödel gave a way of producing within that system the statement g = This statement is not provable.

- Is g itself provable? It cannot be provable because then the statement would be both provable and not provable. That would introduce inconsistencies; thus, g is true, exactly because it is not provable.

- Kurt Gödel put this in his 1931 masterpiece, *On Formally Undecidable Propositions in Principia Mathematica and Related Systems*, and with this, he destroyed the dreams of Whitehead, Russell, Hilbert, and many other mathematicians. His first theorem states that no theory strong enough to do arithmetic is complete. There are true but unprovable statements. His second theorem states that no theory strong enough to do arithmetic can prove its own consistency. A system might be consistent, but we cannot prove within the system that it is consistent.

Suggested Reading

Hofstadter, *Gödel, Escher, Bach*.

Nagel, Newman, and Hofstadter, *Gödel's Proof*.

Questions to Consider

1. In what ways is self-reference used by classical composers in their music?

2. How did Gödel use self-reference in his work that revolutionized mathematics?

Composing with Math—Classical to Avant-Garde

Lecture 10

In the last few lectures, we've been talking about composing music. Scales, rhythms, transformations—these are all important aspects of composition, but in each case, the mathematics is sort of embedded; it's implicit. The composers aren't necessarily thinking about the math, although now you know that they really are doing some math. What if the mathematics were explicitly used in composition? What would a math-based composition sound like? In this lecture, we'll explore ways to explicitly use mathematics and the tools of mathematics to write and analyze music. We'll take a chronological tour through explicit uses of mathematics in composition, going from 1600 up to modern times.

Musical Dice

- The waltz by Mozart we heard at the beginning of the lecture is an example of algorithmic music, produced by a set of rules. It actually stems from an 18th-century European tradition called Musikalisches Würfelspiel, "musical dice games."

- The idea was to compose a large number of measures, all in the same key, and then to make the measures harmonically interchangeable, sort of like puzzle pieces. The players would then write out a chart with instructions on exactly how to pick from these measures, and musicians would play the music based on the roll of a pair of dice. Today, there are online sites that allow you to play a similar game.

- To count the number of waltzes Mozart could compose with this game, we need to use something called the "multiplication principle."
    - o A simple version of this is as follows: If you have 3 shirts to choose from and 4 pairs of pants, that gives you 12 possible outfits ($3 \times 4 = 12$). If you have 3 shirts, 4 pairs of pants, and 2 belts, you have 24 different outfits.

- With Mozart's algorithmic waltzes, there are 11 choices for each of the measures (because the dice can roll anything from 2 to 11), and there are 16 different measures to choose. That's like having 16 different parts of an outfit; thus, we have 11^{16}, which is roughly 4.5×10^{16}, or 45 quadrillion.

- But Mozart did not actually have all those possibilities. Look closely at the eighth measure. If you roll a 2, you get measure 30. If you roll a 3, you get measure 81. If you roll a 4, you get measure 24. But when you look at those measures, they are all exactly the same. The roll doesn't matter.

- You'll always get the same thing for measure 8. In other words, there aren't 11 ways of choosing measure 8; there is really only 1 way to choose measure 8. Similarly, with the 16th measure, there are really only 2 choices.

- To incorporate that information into our analysis, we need to revise the original number of choices, 11^{16}. Two of those 11s become a 1 and a 2, which leaves us with fourteen 11s. The number we get is 2×11^{14}, or about 759 trillion.

- Is each one of those 759 trillion options equally likely? The answer is no because the odds of rolling different combinations of dice are different. For example, you have about a 3 percent chance of rolling double 1s but an 11 percent chance of rolling a 9 because there are more ways to roll a 9 (4 and 5, 5 and 4, 3 and 6, 6 and 3). Thus, you're much more likely to get some measures than others.

- Mozart wasn't the first to compose algorithmic music; the first composer to do so was probably Giovanni Andrea Bontempi.
    - In 1660, he wrote something called "A New Method for Composing for Four Voices, by means of which one thoroughly ignorant of the art of music can begin to compose." The idea was to give people with no composition skills tools to compose music. Other composers who used this method included Joseph Haydn and C. P. E. Bach, Johann's son.

- These Baroque and classical examples of math-based composition are much more tame than what we'll hear later on.

The Progression toward Atonality

- The 19th century continued the progression toward atonality. As you'll recall, this means having no tonal center, no main key on which to begin and end and to return to throughout a composition. Composers moved away from reliance on a single scale or multiple related scales. By 1900, composers were challenging the notions of what music is.

- In 1885, Franz Liszt wrote *Bagatelle sans tonalité*, "Bagatelle with no tonality." In 1894, Claude Debussy wrote *Prelude to the Afternoon of a Faun*, which again, pushed the boundaries of avoiding tonality.
  - This music was not always well-received. In 1913, the premiere of Igor Stravinsky's *Rite of Spring* sparked a riot in Paris.

  - That composition represents almost the height of this increase in dissonance, that is, using intervals that are not prevalent in the overtone series.

- The use of dissonance was taken to its logical extreme by Arnold Schoenberg with 12-tone music and atonality.
  - Schoenberg was born in Vienna in 1874. He played the violin and cello, and by 1909, he had arrived at his ideas for what he called "pantonal" music.

  - Schoenberg was working around the same time that cubism was emerging in the art world, with the work of Picasso and Georges Braque. He was equally revolutionary in terms of his composition and performance.

  - In 1920, he teamed with two of his students, Alban Berg and Anton Webern, forming the Second Viennese School.

- All music theory up to that point had assumed tonality—a foundational note or scale—and by design, some of those notes were more prominent than others. Schoenberg had to search to find a replacement for the structure that tonality provided, and he found it in math. His work used the mathematics that we saw in Lecture 8, especially the Z_{12} group, as well as transformations, inversions, retrograde, and retrograde inversions.

- Schoenberg's goal was to avoid any sense of tonality, and his solution was to force himself to use all the notes with equal frequency.
  - He started with a tone row using each piano note once. We're thinking about modulo octaves, so he's using each of the 12 notes once, but he hasn't decided which octave A might come from; C could be from any other octave. And he does this in a serial way. The tone row is sort of the foundation of a series of tones.

  - In 12-tone music, no tone dominates. That's why it's called "atonal music." Schoenberg did this by structure intentionally. He actually preferred the word "pantonal," not atonal, meaning a synthesis of all of the keys, not avoiding any one key.

- Let's see if we can crudely demonstrate Schoenberg's methods with a simple piece. Instead of working with 12 notes, we'll work with 5 notes. Mathematically, we're going to think about working modulo 5, that is, looking at the remainders when we divide by 5, and those would be 0, 1, 2, 3, and 4.
  - We start by randomly ordering those 5 notes—2, 3, 1, 0, 4—then, we transpose. We start at 0 and add 3 to each one. Working modulo 5, we get $3 + 2 = 0$, $3 + 3 = 1$, $3 + 1 = 4$, $3 + 0 = 3$, and $3 + 4 = 2$. That's our tone row: 0, 1, 4, 3, 2. We then write down the inversion of this, the retrograde, and the retrograde inversion, and then we have to translate the numbers into notes. To do that, we think of A being represented by 0, B by 1, and so on up through E, 4.

  - Next, we have to choose such elements as note lengths, the octave from which we will pick the notes, and the rests.

That gives us a piece of music composed using some of Schoenberg's ideas, although he used all 12 notes and we've used only 5.

- We hear the opening of the fifth piano piece from Schoenberg's Opus 23, *Five Pieces for Piano*. Remember, his goal was to explicitly avoid tonality—the predominance of any single key or scale. He did this using a mathematically based tone row system to ensure that in every group of 12 notes, each half note appears exactly once. Such music places heavy demands on both the performer and the audience.

- Schoenberg's system always referred to a starting point, 0. An alternative method is to pick a starting note, go up four half-steps, use that as the reference point, go up two half-steps, use that as a reference point, and go down three half-steps, similar to bootstrapping. With this system, it's difficult to tell whether notes have been duplicated.
    - A Russian math student and music lover named Vladimir Viro encoded the musical parts in this alternative to the Schoenberg system. He then used a music database to digitize a great number of classical compositions and make them searchable.

    - Viro's database searches only changes and pitch, and it can find melodies no matter the key. His mathematical encoding, which is very similar to what Schoenberg did, makes the search possible.

Aleatory, Spectral, and Computer-Programmed Music

- In the 1950s, music took a turn back toward mathematics, reintroducing randomness into concert music. This is called "aleatory music" or "chance music." The most famous composer of chance music, and possibly the most influential composer of the 20th century, was John Cage.

- Cage was known for his radical experiments. He wrote pieces for "prepared piano," where he placed bolts and screws on top of the strings to alter the sounds as he played. He composed for dance accompaniment with Merce Cunningham's group. He composed based on James Joyce's *Finnegan's Wake*. He composed for melting ice sculptures. He composed a piece called *Organ/ASLSP* ("as slow as possible"); one performance of this piece began in 2001 and is scheduled to end in 2640.

Wind chimes play a version of aleatory music, in which pitches are chosen at random by the wind.

- Perhaps Cage's most famous piece is *4'33"* ("4 minutes and 33 seconds"). This is a piece with no notes. It's written for piano in three movements, and it's timed. The pianist is supposed to sit at the piano for three movements that add up to 4 minutes and 33 seconds. This work seems to be a musical version of the mathematical concept of 0.

- Cage began composing chance music around 1950, making a transition from creation to acceptance—accepting the results of chance. He also gave the performer a degree of uncertainty and randomness. Some compositions require the performer to do random things and play based on the results.
    - This kind of composition introduces mathematical questions. What kind of randomness is Cage using? What is the sample space? When we were talking about Mozart's dice game, there was a sample space where 7 and 9 were more likely than 2 or 12. One of the sample spaces Cage used was a Chinese text called the *Yijing*, the "*Book of Changes*."

  - Cage would compose based on the charts of the *Yijing*, choosing pitches, rhythms, tempo, and so on. One of the results was a 1951 piece called *Music of Changes*.

  - In 1961 and 1962, he used the randomness of star charts to compose *Atlas Eclipticalis*. In 1983, he used the particular rock formation in the Zen temple in Kyoto, Japan, the Ryoanji Temple.

  - Other composers stretched the idea of randomness even further. Charles Dodge, for example, composed a piece called *Earth's Magnetic Field*, in which the pitch changes were taken from the changes in the earth's magnetic field resulting from solar winds.

- Spectral music comes from a European tradition of trying to create the spectrum. Recall from Lecture 2 that the spectrum showed us how much of each overtone we would hear. Mathematically, it was the Fourier transform of the wave form that gave us the spectrum. In Gérard Grisey's *Périodes*, seven instruments were used to re-create the spectrum of a trombone.

- Many programmers are using computers these days to compose program music. One of the most prolific is David Cope, who has done extensive work in this field, including work to emulate Mozart. The algorithms he uses range from fairly simple to highly complicated, bordering on the field of artificial intelligence.

Using Math to Analyze Music

- Music theorists have used mathematics in many different fields to analyze music. Some of the most interesting results here come from the recent work of Princeton's Dmitri Tymoczko, who is working on the geometry of music, in particular the geometry of two-note chords.

- The geometry of a one-note chord is fairly simple. Remember that the notes run from A all the way through G-sharp, and then the next A is really the same thing as the original A if we're thinking modulo octave. Thus, the geometry of a one-note chord is a circle.

- If we had two notes, we would get two coordinates, so we would have A, A-sharp, B laid out on the *x* coordinate and A, A-sharp, B on the *y* coordinate. Each two-note chord would be a spot on a 12 × 12 grid.
    - If we had a chord with G and C, physically, we would have many options for playing that because we could pick the G and C from any octave we wanted, but those would all be examples of a G, C two-note chord.

    - What happens if we wrap this into a circle, as we did for the one-note chords? At first, we might think that the two-note chords form a torus, but when we eliminate redundancy in the chords, we find that we get a Möbius strip.

- Tymoczko's work shows us that the space of two-note chords has the geometry of a Möbius strip, and any sequence of two-note chords is really just a path along the Möbius strip. This is a new way to visualize and study music.

Suggested Reading

Forte, *The Structure of Atonal Music*.

Harkleroad, *The Math behind the Music*, chapters 5–6.

Lewin, *Generalized Musical Intervals and Transformations*.

Loy, *Musimathics*, vol. 1, chapter 9.

Perle, *Serial Composition and Atonality*.

Wright, *Mathematics and Music*, chapter 6.

Questions to Consider

1. How have the ideas of probability been used by composers, both in the classical era and in modern times?

2. How did the atonal and 12-tone composers ensure that their compositions didn't accidentally become tonal?

The Digital Delivery of Music

Lecture 11

In the course of these lectures, we've talked about vibrating objects and their overtones, and we've seen how to construct scales and cords. We then used that knowledge to find some compositional techniques that use mathematics. But how does music get to our ears? In this lecture, we will talk about an important but underappreciated subject: the digital delivery of music and the mathematics of that process. Specifically, we'll look at three ways that mathematics has changed the delivery of music in our digital world: the notes we hear, the number of songs we can fit into a smaller space, and how much cleaner music sounds now than it did in the early days of recording.

Delivery of Music

- The original delivery of music was only in person. Before about the time of the U.S. Civil War, no music had ever been recorded to be played back later. People traveled to hear concerts and had only two or three chances in their lifetimes to hear such works as Beethoven's Ninth Symphony.

- The first known musical recording was not from Edison but from Édouard-Léon Scott de Martinville singing "Clair de la Lune" in 1860. It was written to paper by something called a "phonautograph." This recording was not heard in sound until 2008, when it was reconstructed from lines drawn on this paper. The first music recorded onto a replayable cylinder was Handel's *Israel in Egypt*, recorded in 1888 for Edison.

- The move from cylinders to flat discs around 1890 made it possible to produce multiple copies of a recording. With this development came the rise of gramophone companies. Radio broadcasts of live music started around 1906 to 1910, and recording and playback quality quickly improved.

- Earlier technologies for recording sound resulted in degradation of the recorded media. Modern CDs don't degrade over time, but the manufacturing process for CDs introduces more than 10 errors per second on a disk—50,000 errors! Even so, most CDs sound fine when we play them.

Victrolas and other gramophones were entirely mechanical, requiring no electricity to play.

© iStockphoto/Thinkstock.

The Mathematics of Pitch Correction

- Auto-Tune was invented by Antares Audio Technologies in 1997. It was first heard widely on Cher's album *Believe*. The idea behind Auto-Tune was to provide a digital fix for out-of-tune singing. The process involves two steps: pitch detection and pitch correction.

- For pitch detection, there are several options. Looking at the Fourier transform, we can see where the first peak is; that should be the fundamental frequency. The distance between the peaks also gives us the fundamental frequency, because they represent the fundamental frequency added to itself. If we wanted to stay on the wave form side, we could look at the distance between the wave repeating; that period should give us 1/frequency.

- Once we have done the pitch detection and we know a singer is singing at, say, 430 Hz, we can ask whether the frequency should be something else. In other words, is the singer out of tune? If we know that an A natural should be about 440, we can correct that note upward a bit.

- But we can't just shift the graph over 10 Hz because if the singer is singing at 430 Hz, we are also hearing the overtones at 860, 1290, 1720, and so on, when we should be hearing 880, 1320, and so on.
    - o The key to remember here is that intervals are multiplicative, not additive, and that tells us how to get our solution.

    - o To do the pitch correction, we should multiply, not add. To get from 430 to 440, we need to multiply by 44/43.

- Here's the process for auto-tuning we have so far: We start with a wave form. We compute the spectrum, we detect the frequency, we compare that with a table of correct frequencies, we multiply by a constant, and then we invert. We have to take the inverse Fourier transform to get back to the wave form. Now, we can simply play that corrected sound. In theory, we could adjust the tune of anything with this process, although in practice, it's a bit more complicated.

The Mathematics of Audio Compression

- If we recorded sound at its highest fidelity on a CD, the CD would hold less than two minutes of music. The problem here is that recorded audio and video simply contain too much information. The goal of audio compression is to reduce this amount of stored information while still minimizing the effect on the listening experience.

- The Victrola used a groove that was modeled on the wave form, so that when the needle went through the groove, it vibrated exactly as the wave form did. Digital audio uses 0s and 1s, but it's necessary to convert from a continuous wave form to discrete points that can be represented in this way. That process is called "sampling."

- Sampling a wave form involves identifying a number of points and deciding how tightly to space the points on the wave. More points gives a better sound, but fewer points reduces the size of the file for more storage.

- The mathematics needed for this sampling is called the Nyquist theorem, proved by Harry Nyquist in 1928. What Nyquist proved is that if we sample at frequency *f*, so that we are putting points every 1/*f* seconds, we will save information on all waves that have a frequency of less than *f*/2. If we think about this in terms of the Fourier transform, we will retain all the information on frequencies that are less than half of the frequency we are sampling.

- The limit of human hearing is about 20,000 Hz. The Nyquist theorem tells us that if we sample at more than 40,000 Hz, we will accurately reproduce all sounds that are less than 20,000 Hz. The common sampling rate for audio CDs is actually 44,100 times per second.

- We still need to decide what vertical levels we will sample. How many levels can we actually distinguish? The rate that has been determined for CDs is 16 bits, which doesn't sound like much. But if we have sixteen 0s and 1s, that gives us 2^{16} different levels, or 65,000 levels at which we can sample.

- Once we know how many samplings per second we need to do and what the possible output levels are, that information gives rise to the bit rate—how much data per second of music. If we want CD-quality sound, we need 1400 kilobits per second, that is, 1.4 million 0s and 1s in each second of music.

- To make the file size even smaller, MP3 compression is used. MP3 uses perceptual coding, retaining the parts of the data that people are likely to notice and dropping the parts that people are unlikely to notice. This is an area called psychoacoustics, which looks at such issues as the threshold of hearing. Perceptual audio compression is a very complicated subject. It's also a great example of the intersection of different fields, in this case, math, music, psychology, and computer science.

The Mathematics of CD Encoding

- As mentioned at the beginning of the lecture, even a new CD can have about 50,000 errors, yet it still sounds fine when we play it.

What's used to address the error issue in manufacturing CDs is the Cross-Interleaved Reed-Solomon Code (CIRC).

- A sales representative taking an order on the phone from a customer knows immediately whether or not the customer has read his or her credit card number correctly. How?
    - The representative's computer performs a series of calculations that must result in a multiple of 10 to verify that the correct credit card number has been entered. Using 10 as a credit card check digit catches 70–80 percent of all errors made in entering credit card numbers.

    - Similar check digits are used on airplane tickets, UPCs, bank routing numbers, ISBNs on books, and vehicle identification numbers. Internet communication includes check digits in the packets of information sent back and forth.

- Once errors are detected in this way on a CD, they need to be corrected. To understand the error correction process, we need to go back to 1947 and the work of Richard Hamming at Bell Labs.
    - Hamming developed a system to detect and correct errors on a very early computer. In his system, only four out of every seven pieces of data were true data; the other three were check digits.

    - Imagine that we are sending a digital message of 1, 0, 1, 1, and we are going to append three digits to the end. We put the original four digits into four regions on a chart, and then we add digits in regions 5–7 in such a way that each circle has an even number of 1s. That gives us our check digits, 0, 1, and 0, and we can now put all seven digits in order; the message block reads: 1011010.

    - The person on the receiving end of the message can put those numbers into the same chart to determine whether or not there was an error in transmission and, if so, how to correct it.

- Not every sequence of seven 0s and 1s is a valid block, but how different are two valid blocks? What is the distance between two valid blocks as measured by the number of digits by which they differ? The answer is always three or more. In our example, the distance between what was sent and what was received was only one, and because of that, we knew the message had to be wrong.

- The 7-4 Hamming code has a message length of four set inside a block length of seven. That gives an information rate of 4/7; about 57 percent of the digits that are sent are actual data, and the others are check digits. This particular system can check one error and correct one error. An extended Hamming code enables checking for two errors and correcting for one.

- The next step in encoding a CD is called interleaving. A scratch on a CD may corrupt a number of 0s and 1s in a row, and the solution to avoid losing that bit of music is to intersperse the data from any one moment in the music in a number of different places on the CD. We see an example of interleaving with the message "Math is my favorite." A further example shows how the Hamming code and interleaving can be used to correct a message with a transmission error rate of 20 percent.

- Instead of using a Hamming code, manufacturers of CDs actually use what's called a Reed-Solomon code. This code replaces the single digits that we were just talking about with groups of eight digits. Because of that, a Reed-Solomon code is really working in a particular group, a field called Z_{256}. The resulting Reed-Solomon code can detect up to three errors and correct up to two errors.

- How good is this system? As we said, it results in about 50,000 errors on a disk—but that's out of about 20 billion bits!

- Mathematics is encoded somewhere in all of the ways that we deliver music digitally now. And these ideas are not just used for

CDs but for communicating in deep space missions, receiving satellite TV, and storing data on your hard drive.

Suggested Reading

Benson, *A Mathematical Offering*, chapter 7.

Loy, *Musimathics*, vol. 2, chapter 1.

Pohlmann, *Principles of Digital Audio*.

Questions to Consider

1. How are mathematical tools, such as the Fourier transform, used to correct singers' pitch and compress digital music?

2. What are error-correcting codes, and how do they ensure that scratched CDs still play without problems?

Math, Music, and the Mind

Lecture 12

Throughout these lectures, we have looked at a central question: How can mathematics help us understand music? And we've seen a great deal of evidence for connections between math and music: vibrations, scales, compositional techniques, and so on. In this final lecture, we're going to look at deeper connections; we'll refer back to everything we've learned—all the details—but to make a larger point: The connections we make between mathematics and music are in our minds.

Differences between Math and Music

- Mathematics and music are clearly not the same thing. What aspects of music are completely nonmathematical?

- Although computers can generate melodies using mathematical algorithms, real musical composition is not mathematical. The results of computer composition have no themes or coherence.

- The Fourier transform can explain how the overtones of an oboe are different from those of a clarinet, but orchestrating a melody—deciding which instruments should play which parts—is a nonmathematical art.

- Mathematics can guide us in tuning a piano and predict the sounds that will come out when we hit the keys, but math cannot tell a pianist how hard to strike each note or how the tempo of a piece should ebb and flow.

- Of course, there are also some things that math does that music simply can't. Virtually all scientific advances rely on some form of mathematics, and music cannot make that claim at all. Music can and should be seen as the pinnacle of civilization's creative works, along with other arts, but it isn't useful in the same way that mathematics frequently is.

- Music also seems much more accessible than mathematics. Everybody likes some form of music, even if they don't understand it or can't play it themselves. Of course, the same can't be said for mathematics.

Infant and Child Development

- Almost from birth, infants start to think both mathematically and musically. For example, when babies are just three or four days old, they can distinguish three dots from two dots. This is evidence of "subitizing," that is, instantly counting without counting individual items.

- By about five months, infants can do basic addition, recognizing that 1 + 1 = 2. In experiments at the Infant Cognition Center at Yale University, babies have been shown to look longer at a screen that shows only one doll when two were expected. Older infants differentiate both between the items shown (a doll versus a block) and the number of items shown.

- Some mathematical capabilities seem to be innate or nearly innate, but what about music? Conditioned head-turning experiments in the United Kingdom have shown that infants seem to have early preferences for fast, loud, and familiar music. Additional research has shown that even after a year, babies preferred music they had heard in the womb over similarly styled and tempoed music.

- Infant brains are structured in ways that allow them to process these fundamentally important aspects of both math and music—even before they can use language or walk. The brain somehow comes wired to process, to remember, and maybe even to understand both math and music.

Patterns and Prodigies

- Our brains are marvelous at pattern matching and pattern predicting, and these abilities are at the core of both mathematics and music. One area in which we see the importance of pattern matching is with prodigies, children who perform at an adult level in a given field. Note that we often associate prodigies with three fields: music, math, and chess.

- o In music, we have such prodigies as Mozart and, more recently, the cellist Yo-Yo Ma and the violinists Hillary Hahn and Sarah Chang.

- o In math, we have Srinivasa Ramanujan, who grew up in poor conditions in India and eventually moved to England to work with the most renowned mathematicians in the world. John von Neumann, another famous mathematician, was also a great prodigy. The Teaching Company's own Art Benjamin was something of a child prodigy, as well.

- o In chess, the most famous child prodigy was Bobby Fischer, but other children have become grand masters at even younger ages than Fischer.

- What is it that math, music, and chess have in common that seems to engender prodigies? The answer may be: patterns.
  - o When you're listening to music, your mind is continuously predicting what is next, and it does that based on what you have just heard in that piece, what you have heard before in pieces of the same style, and so on.

  - o Mathematical patterns are usually more explicit. Some of them we know well—1, 2, 3, 4, 5—and we know what comes next. The Fibonacci sequence was a pattern we needed in order to understand octave: 1, 1, 2, 3, 4, 5, 8, 13…. The pattern 2, 3, 5, 7, 11, 13, 17… is the primes in order, the numbers that have only 1 and themselves as divisors.

  - o Sometimes, simple patterns lead to difficult mathematics. For example, Goldbach's conjecture says that every even number can be written as the sum of two primes. This statement has been checked with computers up to extremely large numbers, but we do not know that it is always true.

- The combination of musical and mathematical patterns results in something like the work of David Cope, which is at the boundary between music and artificial intelligence.
    - o Cope has written computer programs that will compose melodies in the style of certain composers or predict melodies based on rules about continuing patterns. In emulating Mozart, Cope's program gets between 64 and 71 percent of the notes correct.
    - o This work tells us that when we predict the end of a musical phrase, we are doing mathematical thinking. It's a mathematical-style algorithm that does this prediction of patterns.

Practice

- The need to practice is another feature that is shared by both math and music. The psychologist K. Anders Ericsson at Florida State University has noted that it takes about 10,000 hours of practice to become an expert in any endeavor.

Psychologists estimate that it takes about 10,000 hours of practice—3 hours a day for 10 years—to become an expert in any endeavor.

- One of the keys to Ericsson's theory is that it is not just practice that is required—not just mindless repetition—but deliberate practice, time spent breaking down, assessing, and refining one's craft.

- Again, it seems as if people can imagine musicians practicing much more easily than they can mathematicians. For most mathematicians, practice consists of pondering puzzles and mathematical ideas, playing games that require strategic thinking or geometrical reasoning, and asking questions of their favorite teachers. Practicing math is not so constrained as other types of practice.

- Of course, there are limits to the argument about gaining expertise through 10,000 hours of practice. Most of us couldn't get in the NBA, even with 1 million hours of basketball practice!

Creativity

- Both mathematics and music have elements of creativity, and this creativity in the two disciplines seems similar. In both cases, practitioners work within structured systems that have patterns.
  - On the musical side, the system includes scales, keys, and tempos. On the mathematical side, it includes definitions and logic.

  - Further, both mathematicians and musicians try to construct something new and original, and if they succeed, what they create is studied and emulated by others.

- One of the implications of creativity in these fields is that we will never run out of math or music. In fact, we can actually use mathematics to quantify the fact that we will never run out of music.
  - How many 10-note melodies can we create with just 12 notes and three note lengths? The answer is 36^{10} different melodies.

  - It would take 1 million songwriters writing 1,000 melodies a day for more than 10,000 years to write out all of those melodies.

- Which endeavor is more creative? Of course, we can't answer that question. But whether you are playing well with a string quartet or attacking a math problem from a new direction, the sense of creativity is exhilarating.

Abstractness

- In addition to patterns and creativity, math and music share the curious trait of abstractness. Both can be expressed intrinsically, and in both, there is no necessary reference to the natural world, although that reference may be present.

- The fact that 5 is a prime number is independent of any part of our physical reality. So is the fact that Beethoven's Fifth Symphony exists. Even if all the copies of Beethoven's Fifth were destroyed, it would still exist in our minds.

- In some sense, music is the most abstract of the arts and math is the most abstract of the sciences. The arts, with some exceptions, largely refer to the human experience. The sciences study physical objects. But both mathematics and music are built around abstract patterns.

Beauty

- Musical beauty seems fairly identifiable. When you hear a soprano soaring through an aria, it evokes a sense of beauty that somehow reaches in and touches your soul. Many people also know that there are different styles of musical beauty: Renaissance music, Baroque, classical, Romantic, and so on.

- What most people don't know is that mathematics has different styles also; it has different aesthetic sensibilities.
  - Mathematicians who study logic—sets, relations, and so on—are sort of like Baroque purists, playing period instruments and making sure that the A they use is not today's 440 A but the A of that particular period.

- Topologists, the mathematicians who study rubber-sheet geometry, give loose, imprecise proofs, similar to the fluid rhythms of Chopin or Debussy.

- Those who study abstract algebra—group theory—and write amazingly perfect proofs are more like Beethoven, who spun his melodies into perfect symphonies.

- The category theorists are the most abstract of mathematicians. Their thinking about such abstract concepts as functions is akin to the work of the atonal and pantonal composers, who work in their own cerebral worlds.

- We end with the Bach Chaconne that we discussed in Lecture 1. As you listen, think about the fact that math tells us what sounds will emanate from a vibrating string. Think about the chords constructed in part because of the mathematics behind the tuning systems used in Bach's time. Think about how scales are constructed using mathematical principles and Bach's use of ideas akin to group theory to transform melodies and put them together. Listen, in other words, for how mathematics informs the musical experience.

Suggested Reading

Harkleroad, *The Math behind the Music*, chapter 9.

Lakoff and Núñez, *Where Mathematics Comes From*.

Levitin, *This Is Your Brain on Music*.

Rothstein, *Emblems of Mind*.

Sacks, *Musicophilia*.

Questions to Consider

1. In what ways are the types of thinking done in mathematics and music similar? In what ways are they different?

2. In what ways are the abstractness and beauty of mathematics and music similar? In what ways are they different?

Bibliography

Five of the books in this bibliography have been written as textbooks for Math and Music courses or “complete” views of mathematics and music and provide excellent additional resources for many of the topics covered in this course. Of these, Wright’s *Mathematics and Music* and Harkleroad’s *The Math behind the Music* are the most accessible, with the former reading more like a textbook (it assumes no mathematical or musical background but explains concepts quite quickly) and the latter as more of a general-interest book.

Three other volumes are significantly more technical and assume mathematical knowledge at roughly an undergraduate-degree level. Of these, Forster’s *Musical Mathematics* is the least mathematically technical because the author brings the perspective of an instrument maker and focuses on the physics of instruments and the scales that can be played as a result. Unlike Forster’s book, Benson’s *A Mathematical Offering* does delve into mathematical compositional techniques in relatively readable manner. Loy’s two-volume *Musimathics* is the most complete reference in existence for connections between mathematics and music from a mathematics or physics perspective; it includes significant material on signal processing and the electronics of digital sound production.

In terms of these five texts, here is a quick guide to how the lecture material matches up with the chapters:

	Author/Chapter				
Lecture No.	**Wright**	**Harkleroad**	**Forster**	**Benson**	**Loy**
1–2	10	2	1–2	1–3	V1: 1–2, 4–8 V2: 2–3, 6
3				4	V1: 6
4–5	4–6, 11–12	3	9–11	4–6	V1: 3
6			2, 5	1	V1: 6
7	2, 8				
8		4		9	
9					
10	6	5–6			V1: 9
11				7	V2: 1
12		9			

Barbour, J. M. *Tuning and Temperament: A Historical Survey*. Mineola, NY: Dover, 2004. Originally published in 1951, this is the bible of piano tunings and temperaments, including technical analysis of subtly different “microtunings.”

Benson, D. *Music: A Mathematical Offering*. Cambridge, UK: Cambridge University Press, 2006 (available for free at http://www.abdn.ac.uk/~mth192/html/maths-music.html). An excellent overview of mathematics and music, written for someone with a mathematical background equivalent to an undergraduate degree in mathematics. Includes sections on synthesized sound not covered in this course.

Deutsch, D. “Diana Deutsch’s Audio Illusions.” http://philomel.com/musical_illusions/. One of the world’s preeminent researchers in psychoacoustics maintains a website with examples of auditory illusions.

Duffin, R. W. *How Equal Temperament Ruined Harmony (and Why You Should Care)*. New York: W. W. Norton & Co., 2008. A popular treatise that argues for the less-equal temperaments popular during the pre-20th-century eras.

Dunne, E., and M. McConnell. “Pianos and Continued Fractions.” *Mathematics Magazine* 72, no. 2 (1999): 104–115. A complete, mathematically rigorous account of the mathematics of equal-tempered scale systems of *n* notes and the connection with continued fractions.

Fischer, J. C. *Piano Tuning: A Simple and Accurate Method for Amateurs*. New York: Dover Publications, 1907. The classic text on piano tuning written in simple language, allowing the amateur tinkerer to adequately tune a piano.

Fletcher, N. H., and T. D. Rossing. *The Physics of Musical Instruments*. New York: Springer Verlag, 1998. A thorough look at the ways in which different musical instruments produce sound through vibrating strings, air columns, solid beams, hollow cylinders, and circular membranes. Assumes significant knowledge of physics and mathematics.

Forster, C. *Musical Mathematics: On the Art and Science of Acoustic Instruments*. San Francisco: Chronicle Books, 2010. A thorough reference written from the perspective of an instrument maker (which Forster is), including the use of English units (inches, feet, slugs, and so on). Of all these references, this is the most complete description of the vibration of wound strings (like most piano strings) and of non-Western scales.

Forte, A. *The Structure of Atonal Music*. New Haven, CT: Yale University Press, 1973. A thorough examination of the (sometimes mathematical) principles behind atonal music. Requires a significant background in music theory and some knowledge of set theory.

Harkleroad, L. *The Math behind the Music*. New York: Cambridge University Press, 2006. Like this course, a gentle introduction to the connections between the two subjects, assuming little knowledge of either subject.

Hofstadter, D. R. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books, 1979. The groundbreaking, Pulitzer Prize–winning, fancifully constructed tour through three seemingly disconnected masters. Sections of Socratic dialogue are interspersed with more technical discussions of the types of self-reference and strange loops seen in mathematics, music, and art.

Lakoff, G., and R. E. Núñez. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books, 2000. A linguist and a cognitive scientist combine forces in an attempt to explain how abstract mathematics arises starting with bodily experiences. Early chapters include overviews of the mathematical abilities of infants.

Levitin, D. J. *This Is Your Brain on Music: The Science of a Human Obsession*. New York: Dutton Adult, 2006. A popular book exploring what the emerging field of neuroscience says about how humans process music.

Lewin, D. *Generalized Musical Intervals and Transformations*. New York: Oxford University Press, USA, 2010. Originally published in 1987, this standard work of music theory sets out an ambitious agenda of analyzing music with mathematical tools. Assumes a significant knowledge of music theory.

Loy, G. *Musimathics: The Mathematical Foundations of Music.* Volumes 1–2. Cambridge, MA: MIT Press, 2006, 2007. The most complete and current encyclopedia of math/music connections. Assumes a significant knowledge of both mathematics and music theory. The material throughout the two-volume set is meticulously and thoroughly referenced.

Magadini, P. *Polyrhythms: The Musician's Guide*. Milwaukee, WI: Hal Leonard Corporation, 2001. A guide to playing and practicing polyrhythms, primarily for percussionists.

Nagel, E., J. R. Newman, and D. R. Hofstadter. *Gödel's Proof*. New York: University Press New York, 1958. A popular and accessible overview of Gödel's groundbreaking work on the incompleteness of axiomatic systems of mathematics. Coauthored by a leading philosopher and the future author of *Gödel, Escher, Bach.*

Perle, G. *Serial Composition and Atonality: An Introduction to the Music of Schoenberg, Berg, and Webern*. Berkeley: University of California Press, 1991. A detailed look at the theories behind atonal music. Not for the musical novice or the mathematically faint of heart.

Pohlmann, K. C. *Principles of Digital Audio*. 6th ed. New York: McGraw-Hill/TAB Electronics, 2010. A thorough, mathematically dense tour through all aspects of digital audio, including CD encoding and various types of compression.

Rothstein, E. *Emblems of Mind: The Inner Life of Music and Mathematics*. New York: Times Books, 1995. A philosophical look at the connections between the subjects, focusing on the similarities for the participant. Both the musical analysis and mathematical arguments are written for the non-expert but remain complicated and impenetrable to the novice.

Sacks, O. *Musicophilia: Tales of Music and the Brain*. New York: Alfred A. Knopf, 2007. The bestselling physician, author, and neurologist takes us on a tour of unusual neurological conditions and how they affect people's ability to process music.

University of New South Wales, http://www.phys.unsw.edu.au/music/. The best online resource for the latest research in the physics of sound. Includes detailed explanations of technical acoustical phenomena.

Wright, D. *Mathematics and Music*. Volume 28. Providence, RI: American Mathematical Society, 2009. This textbook is appropriate for an undergraduate course in mathematics and music, covering a handful of important topics without assuming much knowledge of either subject.