Playing the Numbers: Some highlights from *Mathematics in Music*

In this talk, I want to present some excerpts from an honors course on math in music which I taught last spring. In particular, I want to show that music has connections with a surprisingly broad spectrum of undergraduate mathematics, including (among others) differential equations, probability, number theory and abstract algebra. (There is even a connection with geometry, but that will have to wait for another time.) We begin, though, with something simple: a piece of string.

Vibrations of Strings & Skins

When a string is pulled tight and then plucked, it vibrates back and forth, producing a sound. (Actually, the sound you hear is actually a wave in the air generated by the vibrating string; moreover, the string itself has so little power that the sound produced by the string alone is almost inaudible unless you're up close. In an instrument like a violin or guitar, it's the air in the soundbox that's enclosed within the instrument that generates the sound you hear.) We'd like to study the motion of the string as it vibrates, and we do this by producing a mathematical model---in the form of a differential equation---for that vibration, and studying the solutions of the equation. Applying Newton's second law of motion to the tension force on the string produces the differential equation

$$\partial^2 u/\partial x^2 = (1/c^2)\partial^2 u/\partial t^2$$

where u is the sideways displacement of the string, x is position along the string, and t is time, and c is the square root of the tension divided by the linear density of the string. This is the wave equation, for which we have the d'Alembert formula that gives all solutions

$$u(x,t) = f(x+ct) + g(x-ct).$$

If we take into account the fact that the string is fixed at two ends, which means that $u(0,t)=u(L,t)=0$ (where L is the length of the string), this determines g in terms of f, and implies that f is 2L-periodic:

$$g(x) = f(-x), f(x+L) = f(x-L).$$

Expanding f as a Fourier series (i.e., a linear combination of sines and cosines whose period divides 2L) shows that all the solutions are sums of terms of the form

$$(A_n \cos(n \pi c t/L) + B_n \sin(n \pi c t/L)) \sin(n \pi x/L),$$

where n=1,2,3,4,…

For a fixed n, one of these solutions has period 2L/(nc) in time, that is, the motion of the string repeats every 2L/(nc) seconds. When the motion u(x,t) of the string is described by just one of these terms, for a fixed n, that's called a *vibrational mode* of the string.

[show animation?]

In reality, u(x,t) will be a blend of many vibrational modes happening at the same time. The frequency of the *n*th mode, which is one over its period, is nc/(2L). (The frequency is how many times the motion repeats per second; so, a frequency of 440Hz means a vibration that repeats 440 times a second.) Notice that the frequencies of the various modes are all integer multiples of a lowest frequency c/(2L). This lowest frequency is called the *fundamental*, and is usually the loudest component of the sound heard, and the other vibrational modes are called *overtones*.

[see slide for spectrogram]

The vertical peaks in these spectrograms show the amplitude $C_n=\sqrt{(A_n^2+B_n^2)}$ of the nth vibrational mode. Research has shown that when our ears process musical sounds, they are actually doing Fourier decomposition, with individual nerve endings dedicated to detecting the amplitude of a sine wave with frequency in a particular range. (Our ears don't seem to be sensitive to the phase of the sine wave, though.) As the diagram shows, how much each mode contributes to the blend is one of the things that distinguishes the sound of one instrument from one another; just as important, though, are the attack and decay characteristics, i.e., how a note starts up and how quickly it fades away.

The vibrational modes of a stretched string are also known as *harmonics*, and it's from these harmonics that the standard system of musical intervals is formed. An interval means a pair of musical notes whose frequencies have a particular ratio; for example,
an *octave* corresponds to a 2:1 ratio,
a *perfect fifth* to a 3:2 ratio,
a *perfect fourth* to a 4:3 ratio,
and a *perfect third* to a 5:4 ratio.
(These terms don't make much mathematical sense, but they are traditional in music theory, because of how these intervals occur in the major scale.)
If two notes are sounded at the same time, and their fundamental frequencies are in one of these ratios, then the two notes will seem to resonate, since they will share a large portion of harmonics. This seems to have a pleasing effect on our ears, and such intervals are called *consonant*. For example, if two notes are an octave apart then all the harmonics of the higher note coincide with one half the harmonics of the lower note; for this reason, the octave is the most consonant interval. If two notes a perfect fifth apart, then half of the harmonics of the higher note coincide with one third of the those of the lower note, and so on. In general, the larger the integers involved in the frequency ratio are, the less consonant (and more dissonant) the interval sounds. However, our ears have some tolerance for error built in; as we'll see later, an interval that's close to 3/2 but is irrational may still sound consonant.

Remarks. 1. The analysis of the vibrational modes of a string can be used to explain why it sounds better to bow a violin string than to pluck it. When you draw a horsehair bow across the string, the string is pulled by the bow, until the elastic resistance of the string overcomes static friction and the string quickly snaps back into place. (This happens hundreds of times a second.) This means that the graph of the displacement as a function of time, at a fixed location x_0 along the string, looks like a sawtooth. [draw on board]

Since our earlier formula gives
$u(x_0,t) = \sum C_n \sin(n \pi c t/L + \phi_n) \sin(n \pi x_0 /L)$,
the Fourier amplitudes of the sound produced by bowing the string are determined by approximating the shape of this sawtooth wave. As you may know, when you're trying to approximate a periodic function with a sharp corner, many high-frequency components come into play. This is known as *Gibbs' phenomenon,* and it is why the bowed violin produces a sound that's rich in high harmonics when. Moreover, one can show that the percentage of time that the string spends moving forward (as opposed to slipping back) is proportional to how far away from the bridge the bow is placed. For example, bowing quite near the bridge produces a sawtooth wave that's almost vertical, and for which the Fourier amplitudes C_n decay like 1/n. Most of the time, the bow is placed about 1/4 of the way along the string from bridge to neck, and the sound produced is still rich in harmonics, but more mellow.

2. Not all instruments produce overtones whose frequencies are integer multiples of the fundamental. For example, the vibration of a two-dimensional surface like the skin of a drum is modeled by the 2-d wave equation

$$\partial^2 u/\partial x^2+\partial^2 u/\partial y^2=(1/c^2)\partial^2 u/\partial t^2$$

When we rewrite the wave equation using polar coordinates r and θ, we find that the modes are of the form

$$u = J_n(\omega r/c) \sin(n\theta) \sin(\omega t), \text{ for } n\geq 0,$$

where J_n is a Bessel function of order n. Suppose that we require that u=0 at the edge of circle of radius R around the origin. Since each Bessel function has an infinite sequence of zeros $j_{n,k}$ for $k\geq 1$, setting $\omega R/c=j_{n,k}$ gives a mode with frequency $\omega/(2\pi)=(c/R)j_{n,k}$.

Thus there is a two-parameter family of vibrational modes for the head of a drum, indexed by the two integers n and k. If you took a high-speed picture of a drum vibrating in one of these modes, you'd see the drum head going up and down n times as you go along a circle around the center (moving in the theta direction), and going up a down k times as you move from one edge to the opposite side in the radial direction.

[see slide]

As you can see, the ratios of the frequencies are irrational, which is why the sound produced by a drum doesn't sound like a musical note. One exception is the tympani, or kettledrum. First, the modes with n=0 are not present because the air in the sealed drum can't be compressed, so that the fundamental is proportional to j_1,1. Moreover, tympanists have learned where to strike the drumhead so as to suppress the lowest modes that make a discordant interval with the fundamental.

Tuning Systems

Music gives an interesting example of equivalence relations. If I asked you to sing a note I give, not everyone sings that exact note; some people sing a note that is an octave up or an octave down. Two notes whose fundamental frequencies differ by a number of octaves are sometimes regarded as equivalent, and are said to belong to the same *pitch class*. Because this notion of equivalence is widely prevalent (even across different cultures), most musical scales are constructed in units one octave long. (The term "octave" comes from Western scales that contain 7 pitch classes, plus the starting note repeated an octave higher at the end.) But how are these scales constructed?

If you want to be able to tune, say, a stringed instrument to the notes in a scale, it would be nice if the intervals in the scale were in perfect, rational ratios. (This is because you can check if the strings are in tune with one another by hearing if their shared harmonics exactly line up.) So, it's desirable to use perfect fifths in constructing a scale. The scale constructed by Pythagorean school of natural philosophers contained 7 notes, corresponding roughly to the white keys on the piano keyboard, whose frequencies were derived by going up or down by perfect fifths from a starting note, and using octave

equivalence to put frequencies back into a single octave. If we give the starting note frequency f, the rest of the Pythagorean scale are:
f, 3/2 f, 9/8 f, 27/16 f, 4/3 f, 16/9 f, 32/27 f.
(Notice that I've divided or multiplied by 2's, so that the frequencies are within the octave above f.) Sorting these notes in increasing order (and taking f=1 for simplicity) gives the *Pythagorean scale*:
1, 9/8, 32/27, 4/3, 3/2, 27/16, 16/9, [2]
The intervals between two successive notes in this scale come in two sizes, either 9/8=1.125 or 256/243≈1.11, a big step and a small step. Since multiplying by 1.11 twice is approximately the same as multiplying by 1.125, the small step is called a halfstep (or *semitone*), the large one a whole step. (Note that since an interval corresponds to multiplying one frequency by a ratio to get another, two successive interval steps correspond to multiplying by the product of the corresponding ratios.) The scale consists of two half steps and five whole steps, arranged non-periodically.

[see circle diagram on slide]

One of the features of this asymmetrical arrangement is that if you want to write a melody that ranges over a different octave, say a few steps higher or lower, then the scale you get starting on that note sounds very different. The scales obtained by picking different starting points are called *modes*, and were prevalent in music up until the late Middle ages. (The scale given above is called the Dorian mode, and corresponds roughly to a white-note scale beginning on D on the piano keyboard.) Only two modes are regularly used today, the major and minor scale. The major scale is two whole steps, followed by one halfstep, then three whole steps, and another halfstep. If you want to begin a major scale on any note within the octave, you'll need to fill in the 5 gaps in the Pythagorean scale (where the whole steps are), giving a 12-note scale that allows you to take whole steps and half steps in the right order to make a major scale. The trouble is, filling in the gaps can't be done in a way that's consistent with using only perfect fifths. For, as you go up an increasing number of perfect fifths from the starting note, the notes you generate to fill in the gaps do so more and more asymmetrically. In the end, if you go up 6 perfect fifths and go down 6 perfect fifths from the starting note, you end up with two ways of filling in the gap between G and A which differ by about 10% of a whole tone.

The reason for this "near miss" is that going up 12 perfect fifths is approximately 7 octaves:
$(3/2)^{12} \approx 2^7$, since $3^{12}/2^{19} \approx 1.014$
The modern system of *equal temperament* comes about by replacing the (3/2) ratio of the perfect fifth by an "imperfect fifth" that has the property that 12 successive imperfect fifths is exactly 7 octaves. If we do this, then each whole step becomes exactly two half-steps, with exactly 12 halfsteps in the octave. Thus, the halfstep ratio is given by $2^{(1/12)} \approx 1.0595$ (cf. Pythagorean $256/243 \approx 1.0535$).
This has the drawback of being irrational, so that instruments can't be tuned by listening for when *any* harmonics exactly line up. And, the fifths are audibly different!

[Mathematica demo]

There is much more to the story of tuning systems than the Pythagorean scale and the system we use today, for example:

- the construction of "just intonation" systems that use a combination of perfect fifths and perfect thirds to fill out the 12-note scale
- mean temperament, that distributes the Pythagorean discrepancy over all the fifths, in a way that the thirds come out perfect. (This is still used in performing Renaissance music.)

For a mathematical audience, the fact that equal temperament is a reasonable compromise can be explained in terms of rational approximations to irrational numbers. If we take the base 2 log of both sides of $(3/2)^{12} \approx 2^7$ we get $\log_2(3/2) \approx 7/12$. In fact, if you expand $\log_2(3/2)$ as a continued fraction, the sequence of rational numbers you obtain are
1, 1/2, 3/5, 7/12, 24/41, 31/53, 179/306, …

This raises the possibility that we can get a better approximation of the perfect fifth by using equal temperament, but dividing the octave into a larger number of small steps. For example, we can use a 41-note scale, in which the perfect fifth is approximately 24 steps, or a 53-note scale in which the perfect fifth is approximately 31 steps. (Some musical instruments were built in 19th century using this scale, which also closely approximate the perfect third.)

[see slide for picture of Bosanquet harmonium]

Transformations in Music

Here, I want to pursue an analogy between geometrical transformations like translations and reflections and the transformations used in composing music. Musical space has (at least) two dimensions, time and pitch being the most easily described. Because of the way musical notation works, pitch is regarded as the vertical dimension and time as horizontal. It only makes sense to consider transformations that preserve those axes:

--translation in time (a.k.a. repetition)
--translation in pitch (a.k.a. transposition)
--reflection in time (a.k.a. retrograde, playing something backwards)
--reflection in pitch (a.k.a. inversion, playing something upside down)

Repetition is not so interesting, so we'll consider the effect of the other three transformations on musical motives. (There are other transformations composers have used, for example in canon and fugue forms, but we don't have space to discuss them here; instead, see Lecture .)

Transposing means starting on a different note, but using the same interval steps (e.g. transpose E-G-C to B-D-G).

Retrograde just means playing the theme backwards (e.g. reverse E-G-C to C-G-E). To do an inversion, you turn all the steps up into the same length steps down, and vice-versa. Every inversion has a "fixed point", a pitch which doesn't change, and which you can think of as the "mirror pitch" (e.g., invert E-G-C-D to C-A-E-D, with D as mirror).

Now, suppose we regard these transformations as operating not on actual musical themes, but on sequences of pitch classes. That is, we don't care about the duration of a note, and we ignore octave transpositions. Then the only transpositions that matter are those of between 1 and 11 semitones up. Label these as T_1 through T_{11}. These transformations, together with the identity T_0, form a cyclic group of order 12, labeled Z_{12}. If we let I_0 be inversion with middle C on the piano as "mirror", and let I_k be the inversion about a note k semitones about middle C, then the inversions all have order 2: $(I_k)^2 = T_0$. Just as in geometry the product of reflections in parallel lines is a translation through twice the distance between the lines, the same is true for a product of inversions:

$$I_k \, I_0 = T_{2k}$$

Using this, you can show that inversions and transpositions don't commute; instead we have, for example

$$T_k \, I_0 = I_0 \, T_{-k}.$$

Together, these relations imply that the group generated by transpositions (modulo octaves) and inversion is a dihedral group of order 24. If we throw in retrograde, which commutes with all of the above, then we have a group of order 48.

This group is highly relevant to the study of 20th century music. Around 1920, the Viennese composer and teach Arnold Schoenberg invented a new way of writing highly original and complex music called the *twelve-tone method*. The method is, write down a sequence of 12 pitch classes (called a tone row) and begin writing music until you've used all pitch classes in the order you wrote down (either in chords or in melodic fragments; the rhythms are up to you—for now!). Then let the 48-element musical group act on your tone row, and use one of the transposed, inverted or retrograde forms that result. Keep doing this until you're finished the piece.

[play sound samples]

Schoenberg's pupil Anton Webern liked using tone rows that had internal symmetries, like the last 6 notes being a retrograde inversion of the first. Enumerating tone rows with such symmetries has been the subject of some papers in the Monthly. [see slide]

Still later in the 20^{th} century, the composer Pierre Boulez wrote pieces where all aspects of each note---duration, volume, articulation---were controlled by series. [play sound sample from Structures IA]

Ringing Changes

You could argue that you don't need to know much about group theory to use musical transformations as a composer. However, a surprising amount of the theory of permutation groups was discovered by musical amateurs who liked to practice change-ringing.

Another application of group theory to music is *change-ringing*, which refers to the process whereby people playing church bells can ring the bells in every possible order. The practice of change-ringing developed in England, where there can be as many as 6, 7 or 8 bells in a church tower, all tuned to different pitches. Closer to hand, there is a peal of 10 bells installed at Grace Episcopal Church on Wentworth St.; the largest bell in the set weighs almost a ton. They are tuned to the notes of an E-flat major scale, plus an extra F and G at the top.

Each of the bells is operated by pulling on a rope, which causes the bell to swing until it is nearly upside down, at which point the clapper inside the bell hits the bell and it rings. When the bell is nearly upside down, a little tug on the rope adds enough momentum to ensure that, in a few seconds, it swings all the way in the other direction so that it's nearly upside down again. From then on, the bell can be kept ringing with comparatively small tugs on the rope at the appropriate times. So, each bell goes through a cycle of two rings, one on the handstroke and one on the backstroke. If the ropes are pulled in the right way, there will be enough time between the two rings for all the other bells to ring.

When all the bells are rung in a certain order, that's known as an *extent*. It's traditional to label the bells as 1,2,3, ... from smallest to largest, so an extent on N bells can be notated by arranging the numbers 1 .. N in some order; of course, there are N! ways of doing this. The fact that only small adjustments can be made in the timing of when each bell rings means that the position of a bell in the extent can only be changed by one place at a time. In other words, the physical constraints of bell-ringing imply that the only way we can change the order of bells is by swapping the positions of two bells that are one after the other. For example, with three bells the change 123 -> 312 is not allowed because the "3" bell moved more than 1 place; however, the change 123 -> 213 is allowed.

The mathematical study of bell-ringing comes down to the question of how can one systematically run through all N! extents only by exchanging bells in adjacent positions?
Note that we are not allowed to repeat an extent until the end, and moreover the position of one bell shouldn't remain the same for more than two successive extents.
For example, here is a solution for three bells,

123 ->213->231->321->312->132->123,

in which we alternate between swapping the bells in the first two positions and swapping the bells in the last two positions. Say we label the positions within an extent as A,B,C,D, etc. To denote operation of swapping the bells in positions A and B, we'll write (AB). So, the above sequence of permutations on 3 bells could be written as
123 –(AB)-> 213 –(BC)-> 231 –(AB)-> 321
–(BC)-> 312 –(AB)-> 132 –(BC)-> 213

What about 4 bells? Here we could consider the operation of simultaneously swapping two adjacent bells (i.e., the transformation (AB)(CD)). Here's the beginning of a bell-ringing pattern called *Plain Bob Minimus*:

1234 -> 2143 -> 2413 -> 4231 -> 4321 -> 3412 -> 3142 -> 1324 ...
where we have alternated between applying (AB)(CD) applying (BC). Notice that one more (BC) swap will bring us back to the beginning; in other words, $((BC)\ (AB)\ (CD))^4$ = the identity.
But there are 4!=24 possible orders in all, and we've only gone through 8 of them, so it's too early to repeat. In fact, no matter how many times you apply these two transformations you'll only go through these 8 permutations. This means that the two transformations generate a proper *subgroup* H of order 8 inside the full group G of 24 permutations. (This group of permutations on 4 positions is usually denoted S_4.)

The trick used in Plain Bob Minimus to generate the rest of the permutations is to toss in an extra swap that's not in this subgroup, namely the swap (CD). Then we can go ahead and apply the original pattern of alternating
(AB)(CD) with (BC) again. This means that the next 8 extents are
-> 1342 -> 3124 -> 3214 -> 2341 -> 2431 -> 4213 -> 4123 -> 1432 …
Again, because ((BC) (AB) (CD))^4 is the identity, we don't apply that last (BC) swap. Instead, we insert another (CD) swap as before:
-> 1423 -> 4132 -> 4312 -> 3421 -> 3241 -> 2314 -> 2134 -> 1243 …
and then a final (CD) swap brings us home to the original extent 1234.

A mathematical way of understanding the structure of Plain Bob Minimus is to think in terms of decomposing G into right *cosets* of the subgroup H:

$$G = H \cup H\,k \cup H\,k^2, \qquad k = (CD)(BC)=(CBD)$$

So, the first 8 extents arise by applying permutations in H (including the identity) to the initial 1234; the next 8 come from applying permutations in the coset H k to 1234; and the last come from applying permutations in H k^2 to 1234.
Where does k = (CD)(BC) come from? When we have finished the first 8 extents, one more (BC) operation would bring us back to 1234.

Instead, we apply (CD). So, the first extent 1342 of the middle set can be thought of as obtained by applying $(CD)(BC)^{-1}$ to 1234. And, since (BC) is its own inverse, this is the same as applying k=(CD)(BC) to 1234. Everything else in the middle set of 8 extents is obtained by applying permutations in the coset Hk to 1234. Similarly, the last 8 extents come from applying permutations in coset $H k^2$ to 1234.

Designing change-ringing patterns on larger numbers of bells (5,6,7 or 8) requires similar tricks. For example, many change-ringing patterns are based on the "Hunt" pattern, which on N bells alternates between α=(AB) (CD) (EF) ... and β=(BC) (DE) (FG)....
These two operations generate a subgroup of size 2N. To generate all N! permutations in a pattern that's easy to understand (and teach to a group of ringers) you have to find extra moves outside the subgroup to insert between the sequence of Hunting changes. For example, when N=5, to run through all 5!=120 possible extents the change-ringing pattern called *Grandsire Doubles* inserts the operations
(AB) (DE) and the "bob" (DE) to split the group S_5 into 12 cosets of size 10. This method dates back to at least 1600. A different method of splitting the symmetric group S_5 is used in *Stedman's Doubles*, which uses a decomposition into 20 cosets of size 6. Various mathematical historians and bell-ringing enthusiasts have made the case that Fabian Stedman, who published the bell-ringing treatise *Tintinnalogia* in 1671, was the first group theorist.
==

References

Music: A Mathematical Offering, David Benson (Cambridge)
The Math Behind the Music, Leon Harkleroad (Cambridge)
Measured Tones: The Interplay of Physics and Music, Ian Johnston (Institute of Physics Press)

Lectures and other material from my course are available at
http://iveyt.people.cofc.edu/mathimusic.html