

Thu Sep 15, 2016



Music theory for nerds

Not *music* nerds, obviously.

I don't know anything about music. I know there are letters but sometimes the letters have squiggles; I know an octave doubles in pitch; I know you can write a pop song with only [four chords](#). That's about it.

The rest has always seemed completely, utterly arbitrary. Why do we have twelve notes, but represent them with only seven letters? Where did the key signatures come from? Why is every Wikipedia article on this impossible to read without first having read all the others?

A few days ago, some of it finally clicked. I feel like an idiot for not getting it earlier, but I suppose it doesn't help that everyone explains music using, well, *musical notation*, which doesn't make any sense if you don't know why it's like that in the first place.

Here is what I gathered, from the perspective of someone whose only music class was learning to play four notes on a recorder in second grade. I stress that I don't know anything about music and this post is terrible. If you you so much as know how to whistle, please don't read this you will laugh at me.

Sound and waves

Music is a kind of sound. Sound is a pressure wave.

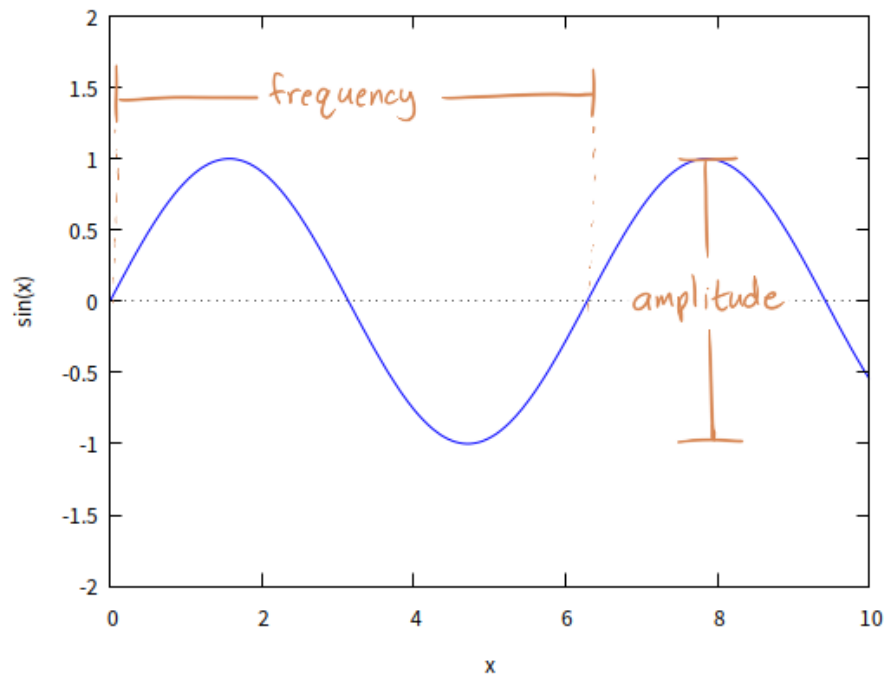
Imagine what happens when you beat a drum. The drumhead is elastic, so when you hit it, it deforms inwards, then rebounds outwards, then back inwards, and so on until it runs out of energy. If you watched a point in the center of the drumhead, its movement would look a lot like what you get when you hold a slinky by the top and let the bottom go.

When the drumhead rebounds outwards, it pushes air out of the way. That air pushes more air out of the way, which pushes more air out of the way, creating a 3D ripple leading away from the drum. Meanwhile, the drumhead has rebounded back inwards, leaving a vacuum which nearby air rushes to fill... which leaves another vacuum, and so on. The result is that any given air molecule is (roughly) drifting back and forth from its original position, just like the drumhead or the slinky.

Eventually this pressure wave reaches your eardrum, which vibrates in exactly the same way as the drumhead, and you interpret this as music. Or perhaps as noise, depending on your taste.

I would love to provide an illustration of this, but the trouble is that it would look like ripples on a pond, where the wave goes *upwards*. Sound happens in *three* dimensions, the movement is directed towards/away from the source, and I think that's a pretty important distinction.

Instead, let's jump straight to the graphs. Here's a sine wave.



It doesn't matter what a sine wave *is*; it just happens to be a common wave that's easy to make a graph of.

In graphs like this, time starts at zero and increases to the right, and the wave shows how much the air (or your eardrum, or whatever medium) has moved from its original position. Complete silence would be a straight line at zero, all the way across.

All sound you ever hear is a graph like this; nothing more. If you open up a song in Audacity and zoom in enough, you'll see a wave. It'll probably be a bit more complicated, but it's still a wave.

Waves are defined by a couple of things: frequency, amplitude, and shape. The particular sound you hear — the thing that distinguishes a guitar from a violin — is the shape of the wave, which musicians call *timbre*.

▶ 0:00 / 0:05

A sine wave sounds something like this:

Amplitude is the distance between the lowest and highest points of the wave. Or, depending on who you ask, it might be half that — the distance between the highest point and zero. For sound, amplitude determines the *volume* of the sound you hear. This seems pretty reasonable, since in physical terms, amplitude is the furthest distance the medium moves. If you tap a drum lightly, it only moves very slightly, and the sound is quiet. If you wail on a drum, it moves quite a bit, and the sound is much louder.

Frequency is, quite literally, how frequent the wave is. If each wave is very skinny, then waves are more frequent, i.e. they have a higher frequency. If each wave is fairly wide, then waves are less frequent, and they have a lower frequency. Musicians refer to frequency as *pitch*. Non-musicians would probably just call it a *note* or *tone*, which musicians would scoff at, but what do they know anyway.

Frequency is measured in Hz (Hertz), which is a funny way of spelling “per second”. If it takes half a second to get from one point on a wave back to the same point on the next wave, that’s 2 Hz, because there are two waves per second. The sound above has a frequency of 440 Hz. (The graph, of course, does not; it’s a completely unchanged sine wave generated by wxMaxima, so its frequency is $1/\tau = 1/(2\pi)$.)

A critical property of the human ear, which informs all the rest of this, is that if you *double* or *halve* the frequency of a sound, we consider it to be “the same” in some way. Obviously, the result will sound higher- or lower-pitched, but they “feel” very similar. I can take a few guesses at the physical reasons for this, but we can treat it as an arbitrary rule.

Compare these three sine waves, if you like. The first is the same as the sine wave from earlier; the second has 1.5× the frequency of the first; the third has 2× the frequency of the first. The first and third sound much more related than the second sounds to either.

- 440 Hz: 
- 660 Hz: 
- 880 Hz: 

Notes and octaves

The difficulty with music is that half of it is arbitrary and half of it is actually based on something, but you can’t tell the difference just by looking at it.

Let’s start with that fact about the human ear: doubling the pitch (= frequency) will sound “the same” in some indescribable way. For any starting pitch f , you can thus generate an infinite number of other pitches that sound “the same” to us: $\frac{1}{2}f$, $2f$, $\frac{1}{4}f$, $4f$, and so on. (Of course, only so many of these will fit within the range of human hearing.) All of those pitches together have some common quality, so let’s refer to a group of them as a *note*. (“Note” can also refer to an individual pitch, whereas *pitch class* is unambiguous, but I’ll stick with “note” for now.)

If we say 440 Hz produces a note called A, then 880 Hz, 220 Hz, 1760 Hz, 110 Hz, and so forth will also produce a note called A. An important consequence of this is that *all* distinct notes we could possibly come up with *must* exist somewhere between 440 Hz and 880 Hz. Any other pitch could be doubled or halved until it lies in that range, and thus would produce a note in that range.

Such a range is called an *octave*, for reasons we’ll see in a moment. Every note exists exactly once within any given octave, no matter how you define it. As a special case, the lowest pitch f is the same note as the highest pitch $2f$, so $2f$ is considered to be part of the next octave.

This is good news! It means we can choose some pitches within a small range — *any* small range, so long as it’s of the form f to $2f$ — and by doubling and halving those pitches, we’ll have a standard set of notes spanning the entire range of human hearing.

How, then, do we choose those pitches? You might say, well! Since we’re going from f to $2f$ anyway, let’s just choose pitches at f , $1.1f$, $1.2f$, $1.3f$, $1.4f$, and so on. Space them equally, so they’re as distinct as possible.

Good plan! Unfortunately, that won’t quite work — if you try it, you’ll find that the difference between f and $1.1f$ is almost twice the difference between $1.9f$ and $2f$.

The human ear distinguishes pitch based on *ratios*, which is why the halving/doubling effect exists. f to $1.1f$ is an increase of 10%; $1.9f$ to $2f$ is an increase of about 5%.

We need a set of pitches that have the same *ratio* rather than the same difference. If we want n pitches, then we need a number that can be multiplied n times to get from f to $2f$.

$$f \times x \times x \times x \times \dots \times x = 2f$$

$$fx^n = 2f \quad x^n = 2 \quad x = \sqrt[n]{2}$$

Ah. We need the n -th root of two. That's a bit weird and awkward, since it's guaranteed to be irrational for any $n > 1$.

Intervals in Western music

Western music has twelve distinct pitches. This is *somewhat* arbitrary — twelve has a few nice mathematical properties, but it's not absolutely necessary. You could create your own set of notes with eleven pitches, or seventeen, or a hundred, or five. There are forms of music elsewhere in the world that do just that.

The ratio between successive pitches in Western music is thus the twelfth root of two, $\sqrt[12]{2} \approx 1.0594631$. Starting at 440 Hz and repeatedly multiplying by this value produces twelve pitches before hitting 880 Hz.

1	0	440 Hz
2	1	466.16 Hz
3	2	493.88 Hz
4	3	523.25 Hz
5	4	554.36 Hz
6	5	587.33 Hz
7	6	622.25 Hz
8	7	659.26 Hz
9	8	698.46 Hz
10	9	739.99 Hz
11	10	783.99 Hz
12	11	830.61 Hz
13	12	880 Hz

No one wants to actually work with these numbers — and when this system was invented, no one knew what these numbers were. Instead, music is effectively defined in terms of ratios.

The ratio between two pitches is called an *interval*, and an interval of one twelfth root of two is a *semitone*. This way, all the horrible irrational numbers go out the window, and we can mostly talk in whole numbers.

(What we're really doing here is working on a logarithmic scale. I know "logarithm" strikes fear into the hearts of many, but all it means is that we say "add" to mean "multiply".)

Now, remember, the human ear loves ratios. It particularly loves ratios of small whole numbers — that's why doubling the pitch sounds "similar", because it creates a ratio of 2:1, the smallest and whole-number-est you can get.

The twelfth root of two may be irrational, but it turns out to *almost* create several nice ratios. (I don't know why twelve in particular has this effect, or if other roots do as well, but it's probably why Western music settled on twelve.) Here are the pitches of all twelve notes, relative to the first. Some of them are *very* close to simple fractions.

1	0	1.000	= 1:1	(unison)
2	1	1.059		(semitone; minor second)
3	2	1.122	= 1.125 = 9:8	(whole tone; major second)
4	3	1.189		(minor third)
5	4	1.260	= 1.250 = 5:4	(major third)
6	5	1.335	= 1.333 = 4:3	(perfect fourth)
7	6	1.414		

8	7	1.498	= 1.500	= 3:2	(perfect fifth)
9	8	1.587			(minor sixth)
10	9	1.682	= 1.667	= 5:3	(major sixth)
11	10	1.782			(minor seventh)
12	11	1.888	= 1.889	= 17:9	(major seventh)
13	12	2		= 2:1	(octave)

Not counting the octave, there are seven fairly nice fractions here.

Hmm. *Seven*. What a conspicuous number.

Scales

Surprise! Those nice fractions make up the *major scale*. Starting with C produces the *C major scale* — the “natural” notes. Using # to mean “one semitone up” and b to mean “one semitone down”, we can name all of these notes.

1	0	1.000	= 1:1	C	(unison)
2	1	1.059		C# or Db	(semitone; minor second)
3	2	1.122	≈ 1.125 = 9:8	D	(whole tone; major second)
4	3	1.189		D# or Eb	(minor third)
5	4	1.260	≈ 1.250 = 5:4	E	(major third)
6	5	1.335	≈ 1.333 = 4:3	F	(perfect fourth)
7	6	1.414		F# or Gb	
8	7	1.498	≈ 1.500 = 3:2	G	(perfect fifth)
9	8	1.587		G# or Ab	(minor sixth)
10	9	1.682	≈ 1.667 = 5:3	A	(major sixth)
11	10	1.782		A# or Bb	(minor seventh)
12	11	1.888	≈ 1.889 = 17:9	B	(major seventh)
13	12	2	= 2:1	C	(octave)

I don't know for sure if this is where the modern naming convention came from, but I sure wouldn't be surprised.

You can now see where some of those interval names came from. A *perfect fifth* is the interval between the first and fifth notes of the scale. An *octave* spans eight notes in total. Likewise, the smallest interval is a semitone because most of the notes are two steps apart, so that distance became a whole tone.

The intervals between successive notes can be written as *wwhwwwh*, where *w* is a whole tone and *h* is a semitone or half tone. Because octaves repeat, you can rotate this sequence to produce seven different variations, depending on where you start. The resulting scales are all called *diatonic scales*, and the choice of starting point is called a *mode*. Here are all seven, marked by Roman numerals indicating the start point. I've also chosen different starting notes for each column, so that the resulting notes are all “natural”.

1				I	II	III	IV	V	VI	VII
2	0	1.000	= 1:1	C	D	E	F	G	A	B
3	1	1.059				F				C
4	2	1.122	= 9:8	D	E		G	A	B	
5	3	1.189			F	G			C	D
6	4	1.260	= 5:4	E			A	B		
7	5	1.335	= 4:3	F	G	A		C	D	E
8	6	1.414					B			F
9	7	1.498	= 3:2	G	A	B	C	D	E	
10	8	1.587				C			F	G
11	9	1.682	= 5:3	A	B		D	E		
12	10	1.782			C	D		F	G	A

$$\begin{array}{rcll} 13 & 11 & 1.888 & \approx 17:9 \quad |B| \quad \quad \quad E \quad \quad | \quad | \\ 14 & 12 & 2 & = 2:1 \quad |C| \quad D \quad E \quad F \quad G \quad |A| \quad B \end{array}$$

I've, er, highlighted two columns. Column I produces the *major scales*, and column VI produces the *natural minor scales*. This explains the rest of the interval names: a *minor third* is the span between the first and third notes in a minor scale, whereas a *major third* is the span between the first and third notes in a major scale. The fourth and fifth notes are the same. (The second notes are the same, too, so I don't know where "minor second" came from.)

You can start anywhere you want to produce a major or minor scale, as long as you follow the same patterns of intervals. With twelve notes, there are twenty-four major and minor scales altogether, which would make for a big boring diagram. Here are a few of the major scales.

1 A major:	A	B	C#	D	E	F#	G#	A
2 A# major:	A#	C	D	D#	F	G	A	A#
3 B major:	B	C#	D#	E	F#	G#	A#	B
4 C major:	C	D	E	F	G	A	B	C
5 C# major:	C#	D#	F	F#	G#	A#	C	C#
6 D major:	D	E	F#	G	A	B	C#	D
7 D# major:	D#	F	G	G#	A#	C	D	D#

If you rotate those major scales to start from C, they look like this.

1 A major:		C#	D	E	F#	G#	A	B	C#
2 A# major:	C	D	D#	F	G	A	A#	C	
3 B major:		C#	D#	E	F#	G#	A#	B	C#
4 C major:	C	D	E	F	G	A	B	C	
5 C# major:		C#	D#	F	F#	G#	A#	C	C#
6 D major:		C#	D	E	F#	G	A	B	C#
7 D# major:	C	D	D#	F	G	G#	A#	C	

Here are some minor scales, written the same way.

1 F# minor:	C#	D	E	F#	G#	A	B	C#
2 G minor:	C	D	D#	F	G	A	A#	C
3 G# minor:	C#	D#	E	F#	G#	A#	B	C#
4 A minor:	C	D	E	F	G	A	B	C
5 A# minor:	C#	D#	F	F#	G#	A#	C	C#
6 B minor:	C#	D	E	F#	G	A	B	C#
7 C minor:	C	D	D#	F	G	G#	A#	C

Hmmmm. Yes, every major scale is equivalent to the minor scale starting from its second-to-last note, due to the way octaves wrap around. They're called each other's *relative major* and *relative minor*.

Also, this notation has a *slight* problem. That problem is that sheet music is terrible.

Sheet music and key signatures

If you know anything about sheet music, you may have noticed that there are no spaces for writing flat or sharp notes.

If you don't know anything about sheet music, well, there are no spaces for writing flat or sharp notes.

If you want to put any of the other notes in, you keep them on the *same line*, but with a # or b next to them. So the notes in D major, which includes F# and C#, are written as though they were F and C, but with some extra #s scattered around. That's not very convenient, so instead, you can put a *key signature* at the very beginning — it takes the form of several #s or bs written in specific places to indicate which notes are sharp or flat. Then any such unadorned note is considered sharp or flat.



(The notes may not actually be arranged this way, depending on the particular squiggle on the left and its vertical position.)

I... guess... that's convenient? If your music *mostly* relies on the seven notes from a particular scale, then it's more compact to only have room for seven notes in your sheet music, and adjust the meaning of those notes when necessary... right?

It completely obscures the relationship between the pitches, though. You can't even easily tell which scale some sheet music is in, short of memorization. In the example above, there are #s for C and F; what about that is supposed to tell you "D"?

Anyway, I really only brought this up to make a point about notation. Look at C# major again.

1 C# major: C# D# F F# G# A# C C#

Two pairs of notes are using the same letter — C and C#, F and F# — so they'd occupy the same position in sheet music. That won't work with the scheme I just described.

To fix this, several scales fudge it a bit. C is one semitone above B, so you could also write it as B#. F is one semitone above E, so you could also write it as E#. Here's how C# is actually written.

1 C# major: C# D# (E#) F# G# A# (B#) C#

Now all seven letters are used exactly once.

I don't think I entirely understand this, because it still seems so convoluted to me. You have to mentally translate that C to a C#, and *then* translate the C# to however that particular note is actually played on your instrument. What does this accomplish? It keeps sheet music more compact — seven notes to express per octave rather than twelve — but I can't think of any better reason.

I suppose it's possible to change the sound of an entire piece of music just by changing the key signature, sometimes without otherwise altering the sheet music at all. How would that work for music that also uses notes outside the scale, I wonder? These seem more like questions of composition, which I definitely don't know anything about.

[Something, something, chords](#)

I'm getting out in the weeds a bit here.

As I said, major and minor scales come in pairs; every major scale has a relative minor scale with exactly the same set of notes, and vice versa. So C major is identical to A minor. Why do we need both? More importantly, since they both use the same key signature, how can you even say that a piece of music is definitively one or the other?

A lot of people have tried to explain this to me as being about mood and different sound and whatnot, but that moves the question rather than answering it. From what I can gather, the real answer is twofold.

One: music is written against a *key*, which includes both the scale and common *chords* and maybe some other stuff. A chord is multiple notes played together, or almost together. You can construct plenty of different chords, but some really big players are the *major chords* and *minor chords*, which are the first, third, and fifth notes in a scale. The C major chord (confusingly written “C”) is thus comprised of C, E, and G, whereas the A minor chord (written “Am”) is A, C, and E.

Major chords consist of some *root* note, the note 4 semitones up from the root, and the note 7 semitones up from the root — less clumsily, {0, 4, 7}. Minor chords are {0, 3, 7}. The first and last notes in both chords are seven semitones apart, which is a perfect fifth, that nice 3:2 (ish) ratio. A major and minor chord with the same root note sound somewhat similar, but because the middle note is slightly lower in a minor chord, it often sounds a little more dramatic or moody.

Speaking of which, something slightly interesting happens when you compare the major and minor scales starting with the same note. They’re very similar, except that three notes are a semitone higher in the major scale.

1 C major:	C	D	E	F	G	A	B	C
2 C minor:	C	D	D#	F	G	G#	A#	C

Every major and minor scale has seven chords of this form, depending on where you start; the second chord in the C major scale, for instance, is D-F-A, which is D minor. Yes, minor; it's the same arrangement of notes as the first chord you'll find in the D minor scale.

Sometimes you'll see chords written using Roman numerals, with capital letters for major chords and lowercase for minor chords. The chords of a major scale are I, ii, iii, IV, V, vi, and vii; the chords of a minor scale are i, ii, III, iv, v, VI, and VII. “I” just means the chord built from the first note, and so on. This lets you talk about, say, [chord progressions](#) without worrying about any particular key.

Anyway, getting back to why we have both A minor and C major, this brings me to...

Two: it's just custom. Western music tends to be written according to certain conventions, and people versed in those conventions can identify which one was being used. Music written in C major will often begin and/or end with C or even a C major chord; music written in A minor will often begin and/or end in A or an A minor chord. As far as I can tell, the two sets of notes aren't fundamentally different in any way, and there's no hard requirement to follow these conventions.

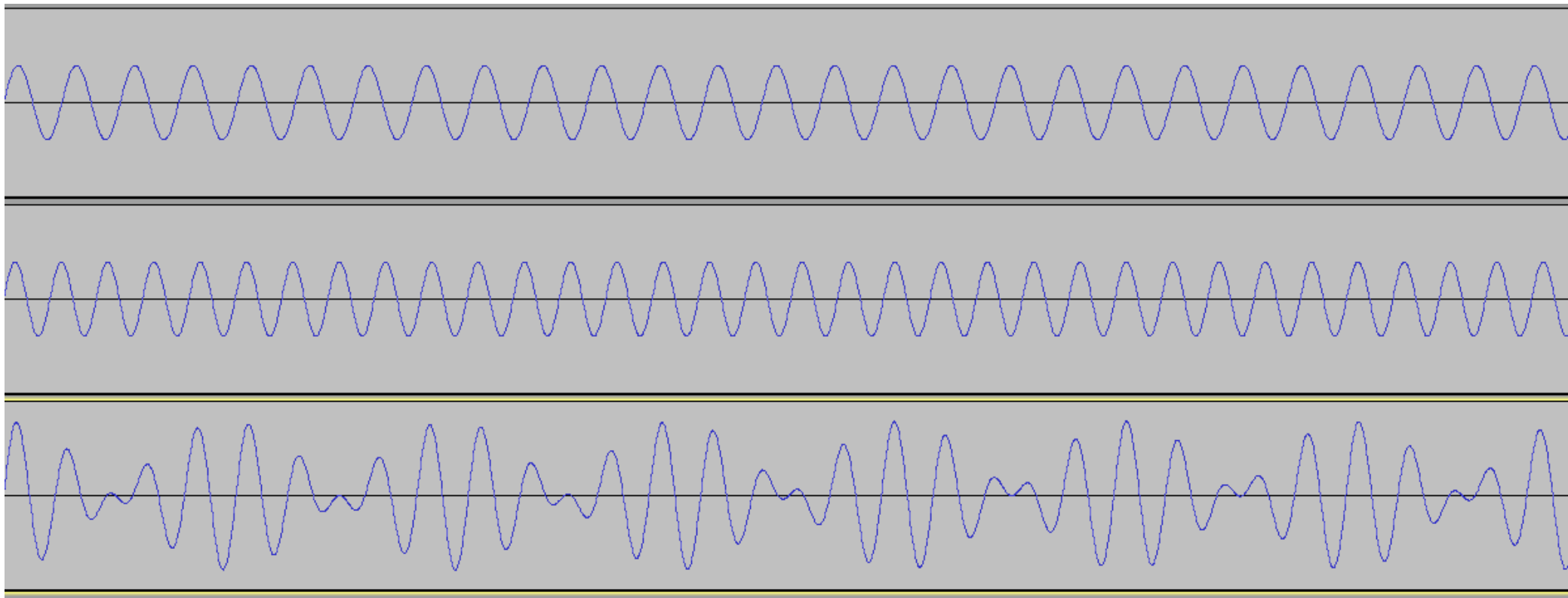
I suppose the advantage is the same as with any convention: your work will be more accessible to others in the same field. Transposing music between keys, for example, only really makes sense if you can confidently say what the original key is. Incidentally, someone linked me an example of [Für Elise being played in A major](#), rather than the A minor it was written in. (And if you played it in C major, it would sound like... uh... wait, I confused myself.)

It should come as no surprise that all of these conventions have myriad variants. A *harmonic minor scale* is a minor scale with its seventh note raised by a semitone. A *melodic minor scale* adjusts several notes, but only when “going up”, not down. There are *augmented* chords (and intervals) whose highest notes are raised by a semitone, and *diminished* chords (and intervals) whose highest notes are lowered by a semitone. And on it goes. All of these things messily overlap and create multiple conflicting names for the same things, because they're attempts to describe human intention rather than an objective waveform.

The [“circle of fifths”](#) is a thing, showing all the major and minor scales in a circle — it turns out that if you name and arrange them the right way, each scale has a different number of sharp or flat notes in it, and no scale has both sharps and flats. The “right way” is to iterate the notes seven semitones at a time (hence “circle of *fifths*”), leading you from C to G to D

and so on. A scale uses sharps or flats depending on which symbol will allow all the notes to be assigned to different letters and thus work nicely on sheet music. I'm sure there's a modular arithmetic explanation for why this all works out as nicely as it does, but I don't know it offhand.

Oh, and integer ratios probably appeal to the human ear because they make the combined waves line up every so often. Here's what a perfect fifth looks like — the top two tones are A4 and E5, using the not-quite-perfect twelfth root of two. Because they're in a nice ratio of 3:2, adding them together creates a repeating set of six waves, which itself resembles a wave.



(A4 is the A note within octave 4. Octave 4 starts at middle C, and both are named based on the layout of a piano. A common reference point for tuning is to set A4 to 440 Hz.)

Finally, notes with the “same name” might not actually be the same note, depending on how the instrument is tuned; various schemes make certain chords have *exact* integer ratios, not just approximations. More “fake” notes exist than E#, too; I hear rumor of such nonsense as G*, “G double sharp”, which I would rather call “A”. I suspect these two trivia are related, but I don't quite know how.

These are all the things that I know. I don't know any more things.

In conclusion

This has got to be some of the worst jargon and notation for anything, ever.

I only looked into this because I want to compose some music, and I feel completely blocked when I just don't understand a subject at all. I'm not sure any of this *helped* in any way, but at least now I'm not left wondering.

It seems that everything not expressible purely as math and waves is pretty much arbitrary. You can pick whatever set of the twelve notes you want and make music with those. Someone pointed out to me that if you just use the black keys of a piano (i.e. the non-natural notes), you get a [pentatonic scale](#) where nothing can possibly sound bad, because no two notes are closer together than a whole tone. You can also use pitches outside these twelve notes, as a lot of jazz and non-Western and other not-classically-inspired music does.

I get the feeling that treating the whole chord/key ecosystem as a set of rules is like studying Renaissance paintings and deciding that's how art *is*. It's not. Do what you want, if it sounds good. I'm gonna go try that. Consensus seems to be that the real heart of music is managing contrast — like every other form of art.

If you aren't quite as ready to abandon the entire Western musical tradition, here's some stuff people linked to me while I was figuring this out in real time on Twitter.

- ["This Week's Finds in Mathematical Physics"](#), some stuff based in group theory
- ["Music"](#), by Toby Fox, best known for creating the soundtrack to Undertale, and also the rest of Undertale; short and mostly lists the considerations
- ["Musimathics: The Mathematical Foundations of Music"](#), a \$33 book that several people recommended but that I have not yet bought
- ["How Music Really Works"](#), a book that brags from the outset that it has no music notation, and offers the first six chapters free; reviews suggest it is particularly helpful for composing, so maybe I should read it sometime
- ["Combinatorial Music Theory"](#), some music theory that appears to have been written by a mathematician who's forgotten how to explain anything without heavy math notation
- ["The Geometry of Musical Rhythm: What Makes a 'Good' Rhythm Good?"](#), a book whose title pretty much gives it away
- [an answer from the music Stack Exchange](#), which touches on attempts to understand *why* music theory is the way it is
- [A Geometry of Music](#), recommended as the best place to start by the above SE answer