Music and Math Connections

#seminar #music

Title

Mathematics of Music: Connections from harmonics to tuning

Abstract

Mathematics and music have a close relationship with each other throughout history. We will go over some of the connections that can be made between the two. The focus will be on how the PDE solution to vibrating strings and wind instruments informs how western music constructs musical chords and scales. Along the way, we'll see some surprising historical connections like how music informed the naming of our trig functions, and how math has been used to construct and tune musical instruments.

Overview Detail

- 1D vibrating string (15 min)
    - PDE Solution
    - Same solution for most wind instruments
        - half open-close pipes are different!
    - Harmonics
- Aside on naming of math functions (5 min)
    - Harmonics (Harmony)
    - Chords (as a string, line segment, and collection of monochords)
    - Trig function (sine, co-sine, tangent)

- Sum of sine waves (10 min)
    - Close frequencies → beating
    - Integer ratios → pleasant periodic waves
- Western Tuning Systems (20 min)
    - Pythagorean (3-Limit)
        - built from 5ths
        - But three note chords have beating
    - Just intonation (5-limit), (if time 7-limit)
        - built from 4:5:6 major chords
        - But some keys sound bad (wolf fifth)
    - (if time) 1/4-meantone
        - keep major 3rds good, and make other keys ok
        - fifths are now bad
    - equal temperament
        - make every key equally ok
- My current research (10 min)
    - source separation problem
    - my harmonic clustering approach

Overview simple

- 1D Vibrating Strings and Pipes (Aerophones and Chordophones)
- Etymology of chord and its math and music connections
- Sum of pure sine waves
- Western Tuning systems
- Connection to my research

1D Wave Equation

Imagine you are in ancient Greece and you hear some metal workers hit an anvil. You notice some anvils sound nice together and some not as nice. Why is that? You try come up with an instrument that can be tuned to change pitch on the fly. So you use cat or another animals gut and pull it into a cord to string it between two fixed points.

What equation models this string? The Greeks did not have the language of PDEs, but let's use it to gain some intuition behind their choices of string length. The 1D wave equation of course. Let $s(x,t)$ model the displacement of the string from equilibrium along its length. It is modelled by the PDE

$$\frac{\partial^2 s(x,t)}{\partial t^2} = c^2 \frac{\partial^2 s(x,t)}{\partial x^2}$$

such that the ends are fixed

$$s(0,t) = s(L,t) = 0.$$

We'll use separation of variables to solve this equation with the assumption that $s(x,t) = X(x)T(t)$ for some single variable functions X and T.

$$\begin{aligned}
\partial_{tt}(X(x)T(x)) &= c^2 \partial_{xx}(X(x)T(t)) \\
XT'' &= -c^2 X''(x)T(t) \\
\frac{T''}{T} &= c^2 \frac{X''}{X} = -\lambda.
\end{aligned}$$

Since this must hold at every x and t, they must be equal to some constant which we'll call $-\lambda$. At this point, $\lambda \in \mathbb{R}$ so there is no loss of generality, but we'll see later that $\lambda > 0$ which explains the choice of sign.

We will solve the X coordinate first.

$$\begin{aligned}
c^2 \frac{X''}{X} &= -\lambda \\
X'' &= -\frac{\lambda}{c^2} X \\
X(x) &= A \sin\left(\frac{\sqrt{\lambda}}{c} x\right) + B \cos\left(\frac{\sqrt{\lambda}}{c} x\right)
\end{aligned}$$

You can check that this is indeed the general solution to the second order ODE for some constants A, B. Note that λ could still be negative at this point, which would turn the sine and cosine into their hyperbolic versions.

Using our boundary conditions, we get the following.

$$\begin{aligned}
0 &= X(0) \\
&= A\sin\left(\frac{\sqrt{\lambda}}{c}0\right) + B\cos\left(\frac{\sqrt{\lambda}}{c}0\right) \\
&= A(0) + B(1) \\
&= B.
\end{aligned}$$

So $B = 0$. For the other boundary condition, we have the following.

$$\begin{aligned}
0 &= X(L) \\
0 &= A\sin\left(\frac{\sqrt{\lambda}}{c}L\right) + (0)\cos\left(\frac{\sqrt{\lambda}}{c}L\right) \\
0 &= \sin\left(\frac{\sqrt{\lambda}}{c}L\right) \\
n\pi &= \frac{\sqrt{\lambda}}{c}L \quad \text{for any } n \in \mathbb{Z} \\
\frac{n\pi c}{L} &= \sqrt{\lambda} \quad \text{for any } n \in \mathbb{Z}
\end{aligned}$$

Since the ODE is linear, we can have any linear combination of solution for a given n

$$X_n(x) = A_n \sin\left(\frac{n\pi x}{L}\right).$$

Now for T. We solve the ODE similarly, knowing that $\sqrt{\lambda} = \frac{n\pi c}{L}$.

$$\begin{aligned}
T''(t) &= -\lambda T(t) \\
T_n(t) &= C_n \sin\left(\frac{n\pi ct}{L}\right) + D_n \cos\left(\frac{n\pi ct}{L}\right).
\end{aligned}$$

Putting it all together, we have

$$
\begin{aligned}
s(x,t) &= \sum_n X_n(x)T_n(t) \\
&= \sum_n A_n \sin\left(\frac{n\pi x}{L}\right)\left(C_n \sin\left(\frac{n\pi ct}{L}\right) + D_n \cos\left(\frac{n\pi ct}{L}\right)\right) \\
&= \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right)\left(a_n \sin\left(\frac{n\pi ct}{L}\right) + b_n \cos\left(\frac{n\pi ct}{L}\right)\right).
\end{aligned}
$$

Note we have not lost any generality since negative n yields a sign flip that can be absorbed by the coefficients a_n and b_n, and $n=0$ gives us a zero term in the sum.

These coefficients are set by the initial conditions $s(x,0)=f(x)$ and $\partial_t s(x,0)=g(x)$.

Initial conditions

To find a_n and b_n in terms of $f(x)$ and $g(x)$, we use the Fourier series! Observe that

$$
\begin{aligned}
\left\langle f(x), \sin\left(\frac{m\pi x}{L}\right)\right\rangle &= \left\langle s(x,0), \sin\left(\frac{m\pi x}{L}\right)\right\rangle \\
&= \frac{2}{L}\int_0^L s(x,0)\sin\left(\frac{m\pi x}{L}\right)dx \\
&= \frac{2}{L}\int_0^L \sum_{n=0}^{\infty} \sin\left(\frac{n\pi x}{L}\right)(a_n(0)+b_n(1))\sin\left(\frac{m\pi x}{L}\right)dx \\
&= \sum_{n=0}^{\infty} b_n \frac{2}{L}\int_0^L \sin\left(\frac{n\pi x}{L}\right)\sin\left(\frac{m\pi x}{L}\right)dx \\
&= \sum_{n=0}^{\infty} b_n \delta_{nm} \\
&= b_n
\end{aligned}
$$

Similarly,

$$
\left\langle g(x), \cos\left(\frac{m\pi x}{L}\right)\right\rangle = \frac{m\pi c}{L} a_m.
$$

So, depending on where and how you plug a string, the relative amplitudes of the harmonics will change. This explains why plucked guitar, bowed violin, and piano struck with hammers sound different! Plucking a string corresponds to some non-zero initial condition $f(x)$, and no initial velocity $g(x)$. Whereas a bow or hammer forces a string to have a non-zero velocity $g(x)$ where it is bowed or struck. On top of this, the body of the instrument will amplify and diminish particular frequencies which can also affect the *timbre*; the sound quality or colour that makes each instrument sound unique.

It's all sine waves?

A string vibrates in a superposition of waves with (angular) frequency $\omega_n = n\frac{\pi c}{L}\,\mathrm{rad/s}$ or (ordinary) frequency

$$\nu_n = \frac{\omega_n}{2\pi} = n\frac{c}{2L}\,\mathrm{Hz}.$$

The wavelength is given by

$$\lambda_n = \frac{c}{\nu_n} = \frac{2L}{n}\,\mathrm{m}.$$

Connection: This is where the name "harmonic" in "harmonic series" comes from. We call each wave n, the nth "harmonic". The musical harmony we get by combining harmonics of wavelength $\propto \frac{1}{n}$ gives us the harmonic series. This is also where terms like "harmonic mean" and "harmonic progression" come from.

The speed of sound along a string is given by $c = \sqrt{\frac{T}{\mu}}$ where T is the tension, and μ is the linear density (units of mass / length).

We should be thinking of each note as a superposition of pure tones (sine waves). This is how our ears work. There is a line of hairs in the cochlea, that vibrates in resonance to a particular frequency. The shorter hairs (responsible for higher frequencies) are closer to the air, and the most fragile. So they are the first to be damaged with over exposure to loud sounds, and general hearing loss with age.

For most string and air based instruments, what makes them sound different is the relative amplitudes of the harmonics. This lets us easily synthesize instruments by controlling the amplitude of different harmonics (additive synthesis) and is used by organs, accordions, and synthesizers. Electric guitars can sometimes switch pickups to make some harmonics louder than others to give it a different timbre.

1D air columns

For completeness, aerophones (which includes wind and brass instruments) also behave similarly to string instruments, but with different boundary conditions.

Open-open air columns (e.g. Flutes)

We get the same wave equation as before,

$$\frac{\partial^2 u(x,t)}{\partial t^2} = c^2 \frac{\partial^2 u(x,t)}{\partial x^2}$$

but now u represents the change in air pressure from equilibrium. For open-open air columns, we get the same boundary conditions $u(0,t) = u(L,t) = 0$ as a string with two fixed ends. Examples of this set up include blowing across a pipe such as a flute, or some pipe organs. This means the air pressure matches the ambient air pressure at the open ends. Solving this PDE becomes identical to the vibrating string.

$$u(x,t) = \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right) \left(a_n \sin\left(\frac{n\pi ct}{L}\right) + b_n \cos\left(\frac{n\pi ct}{L}\right)\right)$$

So all the harmonics are the same as before, with possibly different relative amplitudes.

Closed-open air columns (e.g. Clarinets)

Again, the same wave equation can be used to model the change in air pressure from equilibrium $u(x,t)$, but we now have asymmetrical boundary conditions

$$\partial_x u(0,t) = 0 \quad \& \quad u(L,t) = 0.$$

At the mouthpiece, the air is forced to be at an anti-node of pressure since the air is being driven back and forth. This explains the different in boundary condition at $x = 0$. What does this do to the solution?

As before, we get the general solution for the X coordinate

$$X(x) = A\sin\left(\frac{\sqrt{\lambda}}{c}x\right) + B\cos\left(\frac{\sqrt{\lambda}}{c}x\right).$$

Now let's use the boundary conditions. For the first condition we have

$$\begin{aligned} 0 &= X'(0) \\ &= A\frac{\sqrt{\lambda}}{c}\cos\left(\frac{\sqrt{\lambda}}{c}(0)\right) - B\frac{\sqrt{\lambda}}{c}\sin\left(\frac{\sqrt{\lambda}}{c}(0)\right) \\ &= A\frac{\sqrt{\lambda}}{c}(1) - 0 \\ &= A. \end{aligned}$$

For the second, we have

$$\begin{aligned} 0 &= X(L) \\ 0 &= B\cos\left(\frac{\sqrt{\lambda}}{c}L\right) \\ n\pi - \frac{\pi}{2} &= \frac{\sqrt{\lambda}}{c}L \\ \frac{(2n-1)\pi c}{2L} &= \sqrt{\lambda}. \end{aligned}$$

So we get almost the same solution as before, but with three differences: 1) we have a cosine instead of sine in our X dimension, 2) the open-closed pipe has a fundamental wavelength twice as long (half the frequency) as a similarly sized open-open pipe, and 3) we only get odd harmonics!

$$u(x,t) = \sum_{n=1}^{\infty} \cos\left(\frac{(2n-1)\pi x}{2L}\right)\left(a_n \sin\left(\frac{(2n-1)\pi ct}{2L}\right) + b_n \cos\left(\frac{(2n-1)\pi ct}{2L}\right)\right)$$

A note on brass instruments

The closed-open air column model works well for clarinets. It *should* work for most brass instruments since they are also closed-open air columns, but this turns out to be a bit naïve. We usually like instruments to have the full suite of harmonics. So the mouthpiece in brass instruments is designed as a cone, and the end of the instrument flairs outwards like an amplifier. This causes higher harmonics to shift downwards in frequency, and the lowest harmonics to shift upwards in a nonlinear way so that they approximate the usual harmonic series.

Understanding musical instrument's note

Given some audio, we can find out the harmonics by taking the Fourier transform to extract the relative amplitudes of any pure tones present.

See [flute and guitar spectrums demo in audacity].

More than one note

Congrats, you understand one note! But maybe you want multiple notes at a time. You want to sing together, or find interesting harmonies. So you string up not just one cord (a monochord), but multiple *chords* and play them together.

Aside on the name of chord

This is where the mathematical word for a line between two points come from! A chord (in the circle) is a cord between two points.

This is also where the word for sine comes from! The chord of an angle AOB is the length of the segment AB where OA and OB are unit radius of a circle. This was translated (same meaning) into the Sanskrit word *jya*, then transcribed (same sounding) into Arabic *jib*, and then (mis-)translated from the heteronym *jayb* into Latin *sine* meaning pocket, fold, or cavity. This is the same root word that gives us the word *sinus*, the cavity in our nose.

Sum of two pure tones

Say we have two strings playing at frequencies a and b and we listen to them at a point in space. They are possibly phase shifted by α and β. What is their sum?

$$
\begin{aligned}
&\sin(2\pi at + \alpha) + \sin(2\pi bt + \beta) \\
&= 2\cos\left(\frac{2\pi at + \alpha - 2\pi bt - \beta}{2}\right)\sin\left(\frac{2\pi at + \alpha + 2\pi bt + \beta}{2}\right) \\
&= 2\cos\left(2\pi\frac{a-b}{2}t + \frac{\alpha-\beta}{2}\right)\sin\left(2\pi\frac{a+b}{2}t + \frac{\alpha+\beta}{2}\right) \\
&= 2\cos\left(2\pi\frac{a-b}{2}t + \theta\right)\sin\left(2\pi\frac{a+b}{2}t + \phi\right)
\end{aligned}
$$

Now we don't really notice absolute phase differences (we can notice differences between each ear). So when the frequencies a and b are close, we hear a tone with the average frequency (same pitch), but slowly changing amplitude called "beats". The beat frequency is half the difference between the frequencies $\frac{a-b}{2}$.

See [this desmos demo].

So what notes sound good together?

Depending on your taste, you may like this beating or shimmering quality. This is used in Gamelan music where you have two instruments slight tuned apart so that each note can shimmer with its pair at a consistent beat.

Western music is instead built around integer ratios between frequencies which minimize this beating.

If we have two notes in a $1:2$ frequency ratio (the same as the lengths having a $2:1$ ratio), we don't have any beating, and the period of the wave is the same! It just has a slightly different timbre. Since they are so similar, we identify them as the same note.

Notes in a $1:3$ frequency ratio also sound nice together. It also periodic with the same period as the lower frequency tone. Since this the higher frequency note is "outside" the octave, we will identify it with a note with half its frequency. This gives us two notes with the ratio $1:\frac{3}{2}$ or $2:3$. We'll call this a perfect fifth (why we use fifth will become apparent later).

We'll also remark that these notes in ratio $1:2:3$ are the first three harmonics in a single note played by a string or wind instrument. So not only are the fundamental frequencies in small integer ratios, so are all their harmonics! The higher note reenforces (amplifies) every other harmonic of the lower note.

Let's make some music! (Building a scale)

So you have these three strings playing the tonic, its octave, and a fifth in between. That sounds nice together, but it is starting to get boring. Let's try adding some notes!

What if we add a note that is a fifth *below* our fundamental? We would need a string that is $\frac{3}{2}$ times as long. So if we want this is compact and portable, we'll identify the note with its octave. This gives us a frequency in a ratio of

$1 : \frac{2}{3} \times 2$ or $1 : \frac{4}{3}$. We'll call this a perfect fourth (again for reasons that will become clearer later).

Here is the current state of the frequencies of our 4 notes so far.

$$1 : \frac{4}{3} : \frac{3}{2} : 2$$

Let's keep playing the game of jumping up and down fifths and octave to fill out more and more notes between the tonic and its octave.

$$\begin{aligned}
1 : \frac{4}{3} : \frac{3}{2} : 2 \\
\text{tonic, forth, fifth, octave} \\
1 : \frac{9}{8} : \frac{4}{3} : \frac{3}{2} : 2 \\
\text{fifth above fifth, octave below, call this 2nd} \\
1 : \frac{9}{8} : \frac{4}{3} : \frac{3}{2} : \frac{27}{16} : 2 \\
\text{fifth above 2nd, call this 6th} \\
1 : \frac{9}{8} : \frac{81}{64} : \frac{4}{3} : \frac{3}{2} : \frac{27}{16} : 2 \\
\text{fifth above 6th, octave below, call this 3rd} \\
1 : \frac{9}{8} : \frac{81}{64} : \frac{4}{3} : \frac{3}{2} : \frac{27}{16} : \frac{243}{128} : 2 \\
\text{fifth above 3rd, call this 7th}
\end{aligned}$$

Now we have 8 notes between (and including) the tonic and octave. We'll call this the Pythagorean major scale and is where the Greeks stopped at first. This should now explain where the numbering comes from. Even the word "octave" has the prefix *oct-* meaning eight (think of an octopus with eight tentacles).

There is also a nice pattern if we look at the ratio between neighbouring notes. We have big steps that are in a ratio of $\frac{9}{8}$, and small steps that are in a ratio $\frac{256}{243}$. The pattern is big-big-small-big-big-big-small.

We could also have different modes of the scale by treating different notes as the tonic. For example, if we treat the 6th note as the tonic, we get the Pythagorean minor scale. Let's move the top two notes down an octave to do this!

$$\begin{array}{ccccccccccccccc} \frac{27}{16}\frac{1}{2} & : & \frac{243}{128}\frac{1}{2} & : & 1 & : & \frac{9}{8} & : & \frac{81}{64} & : & \frac{4}{3} & : & \frac{3}{2} & : & \frac{27}{16} \\ 1 & : & \frac{9}{8} & : & \frac{32}{27} & : & \frac{4}{3} & : & \frac{3}{2} & : & \frac{128}{81} & : & \frac{16}{9} & : & 2 \\ A & : & B & : & C & : & D & : & E & : & F & : & G & : & A' \end{array}$$

While we are at it, we better name all these notes with a letter for convenience.

Since we may want to play major and minor scales, we will keep notes between A and C'.

$$\begin{array}{ccccccccccccccccccc} \frac{27}{16}\frac{1}{2} & : & \frac{243}{128}\frac{1}{2} & : & 1 & : & \frac{9}{8} & : & \frac{81}{64} & : & \frac{4}{3} & : & \frac{3}{2} & : & \frac{27}{16} & : & \frac{243}{128} & : & 2 \\ A & : & B & : & C & : & D & : & E & : & F & : & G & : & A' & : & B' & : & C' \end{array}$$

Sharps and flats

If we wanted to play a minor scale starting on C, we would have to add extra notes with ratio $\frac{32}{27}$, $\frac{128}{81}$, and $\frac{16}{9}$. These are close to our notes E, A',and B', but slightly lower. In European history, they may have used lower case letters e, a, and b to distinguish them. Notes e and a were not used much at first, but distinction between the softer b and the natural B' became especially used.

The distinction was necessary when looking at the hexachords (six-notes) built off of the three most pleasant sounding notes C, F, and G. The hexachord build off G would be G-A-B-C-D-E, but playing F-G-A-B-C-D would give the wrong fourth note that sounded too harsh and sharp. The softer b would be needed to make the proper hexachord F-G-A-b-C-D. This inspired the notation $\flat$ for a soft or *flattened* note B. The sharper note was drawn with straight edges and became the foundation for both the natural $\natural$ and sharp $\sharp$ symbols.

More notes at once!

The above system prevailed for a long time, with the addition of some more notes by jumping up and down fifths.

But mostly perfect fourths, fifths, and octaves really sounded nice together. It was in the Renaissance that more harmony and notes played together was being used, rather than mostly overlapping melodies. It became important that multiple notes being played at the same time sounded consonant together.

The three triads built on the three perfect notes C, F, and G are the major triads C-E-G, F-A-C, and G-B-D. They would outline the movement from the home note C, to a pre- or sub-dominant F, and then the dominant G before returning home.

The fifth were already well tuned, but a different ratio for the thirds were picked build off dropping the 5th harmonic down two octaves to be within 1 and 2. This gives us the ratio $1 : \frac{5}{4}$ for the ratio between the tonic (home) and the third. The triad becomes the nice simple ratio $1 : \frac{5}{4} : \frac{3}{2}$ or $4 : 5 : 6$. Constructing our major scale in this way, we replace those trouble notes E, A, B with their nicer ratio.

$$\begin{array}{ccccccccccccccccccccc} \frac{5}{3}\frac{1}{2} & : & \frac{15}{8}\frac{1}{2} & : & 1 & : & \frac{9}{8} & : & \frac{5}{4} & : & \frac{4}{3} & : & \frac{3}{2} & : & \frac{5}{3} & : & \frac{15}{8} & : & 2 \\ A & : & B & : & C & : & D & : & E & : & F & : & G & : & A' & : & B' & : & C' \end{array}$$

The numbers look much less scary now! This is also reflected in the sound. The major third in each triad reinforces every third harmonic its root note. The numbers are all formed by ratios of primes up to 5 (so 2, 3, and 5.) We call this kind of tuning 5-limit just intonation.

MORE notes please

If I want a four note chord C-E-G-B, I will need a nice ratio for all these notes. Following the pattern, we can pick the ratio $4 : 5 : 6 : 7$ which would give us a *very* flat B. It has a ratio $1:\frac{7}{4}$ to the tonic and is called the harmonic 7th after dropping the 7th harmonic down two octaves. This can still be heard in barbershop music which is built around harmony for four voices, each on their own note.

Changing the tonic

The system for just intonation works great!...if you keep C as the home key. If you wanted a different note to be the tonic, you have to keep adding more notes to make sure the major third stays in a $4:5$ to the root. This is exactly what they did.

(Image Credit Carey Beebe)

But you would have to keep adding more and more keys if you wanted to change the tonic. You never got back to the tonic no matter how many fifths and octaves you jumped around. This is because

$$\left(\frac{3}{2}\right)^m\left(\frac{1}{2}\right)^n=1$$

can only be satisfied exactly for integers m and n when $m=n=0$.
The first time you can come close is when $m=12$ and $n=7$

$$\left(\frac{3}{2}\right)^{12}\left(\frac{1}{2}\right)^7=\frac{531441}{524288}\approx 1.0136.$$

What should we do? We could make every fifth a little smaller. For example build Pythagorean tuning off of a tempered fifth $\frac{3}{2}\alpha$ for some small factor α. If we still wanted 4 perfect fifths to give us our well tuned major third $\frac{5}{4}$, we need

$$\begin{aligned}\left(\frac{3}{2}\alpha\right)^4\frac{1}{4} &= \frac{5}{4}\\ \alpha &= \frac{2}{3}\sqrt[4]{5}.\end{aligned}$$

This makes our fifths $\sqrt[4]{5} \approx 1.495$ rather than the ideal ratio $\frac{3}{2} = 1.5$. This is called $\frac{1}{4}$-comma meantone (pronounced quarter-comma) since we have flattened the fifth by a quarter of a comma $\frac{80}{81}$

$$\frac{3}{2}\left(\frac{80}{81}\right)^{1/4} = \sqrt[4]{5}.$$

This was a real tuning system that was used in Europe!

But it still did not wrap 12 fifths around to a pure octave.

$$\left(\sqrt[4]{5}\right)^{12}\left(\frac{1}{2}\right)^7 = \frac{125}{128} \approx 0.97656.$$

To do that, we need

$$\begin{aligned}x^{12}\frac{1}{2^7} &= 1\\ x &= \left(\sqrt[12]{2}\right)^7.\end{aligned}$$

New plan, equal temperament

"To make every key equally ok (but none of them perfect), let's use 12 notes, each in a ratio of $1 : \sqrt[12]{2}$ to its neighbouring note" - said some European in the 1600s, probably. Actually, this tuning was likely discovered earlier in the 1500s in China.

Regardless, we get the scale that is commonly used today in western harmony.

$$\underbrace{1}_{C}, \underbrace{2^{\frac{1}{12}}}_{D^\flat}, \underbrace{2^{\frac{2}{12}}}_{D}, \underbrace{2^{\frac{3}{12}}}_{E^\flat}, \underbrace{2^{\frac{4}{12}}}_{E}, \underbrace{2^{\frac{5}{12}}}_{F}, \underbrace{2^{\frac{6}{12}}}_{G^\flat}, \underbrace{2^{\frac{7}{12}}}_{G}, \underbrace{2^{\frac{8}{12}}}_{A^\flat}, \underbrace{2^{\frac{9}{12}}}_{A}, \underbrace{2^{\frac{10}{12}}}_{B^\flat}, \underbrace{2^{\frac{11}{12}}}_{B}, \underbrace{2}_{C'}$$

How nice (sarcastic?). But this is a very convenient system for playing in any key without needing to retune your instrument each time.
Fretless string instruments like violins, and voices will still naturally tune to the more just intervals on the fly, and with the help of digital instruments, we can now do the same with keyboards. But we are stuck with every note being equally ok, and I'm ok with that.

Active research

What's the problem?

The fundamental problem is source separation is that it is easy to add signals together, but hard to meaningfully separate them. This is analogous to factorizing prime numbers. Actually, it is more like trying to find numbers that sum to a larger number, where the constituent numbers need to be "meaningful".

My approach

Thinking back to the spectrums of flute and guitar, visualizing each note in this way works well if we have isolated each note. But if we look at the whole file, we get a mess of peaks that is hard to discern.

This is where the spectrogram, or short-time Fourier transform (STFT) comes in! This less us see how the Fourier transform evolves in time, by applying it to small windows of the whole picture.

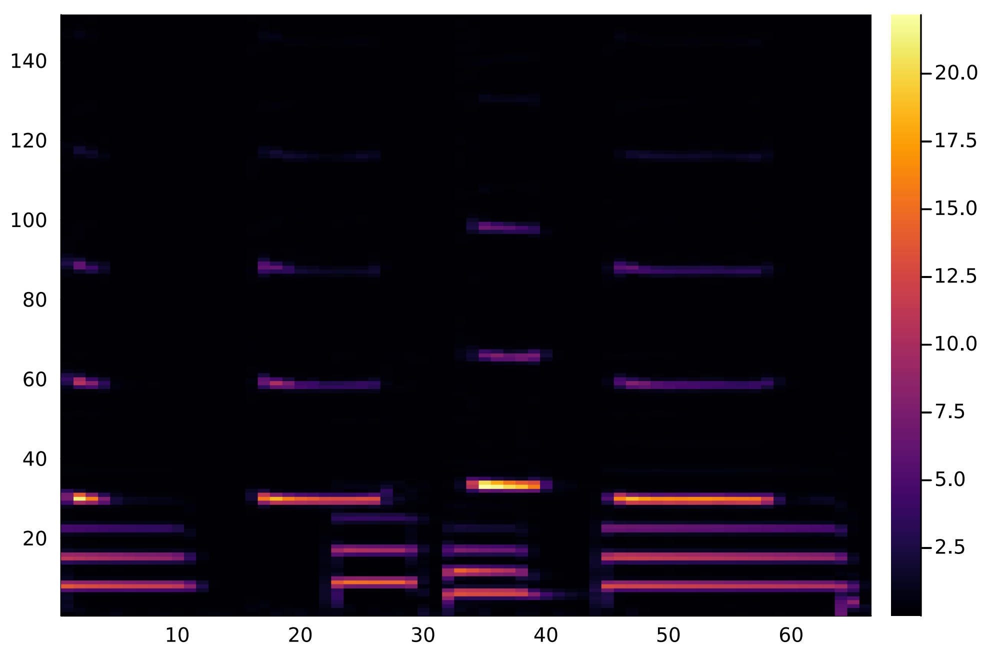
See flute and guitar spectrograms demo in audacity.

The approach to separating the mixture is to identify each note, based on the harmonic timbre. This step can be done with a nonnegative matrix factorization, where the rank is given by the number of unique notes present repeated notes by the same instrument do not need to be double counted.

The second step is to group the notes that "sound similar". This means we need to compare the timbre of each note to each other, and cluster them based on similarity. We can do this by checking the cross-correlation (in log-frequency space) between the notes, and grouping notes that have similar looking peaks.

Here's a Demo!

This is the input spectrogram (as a matrix).

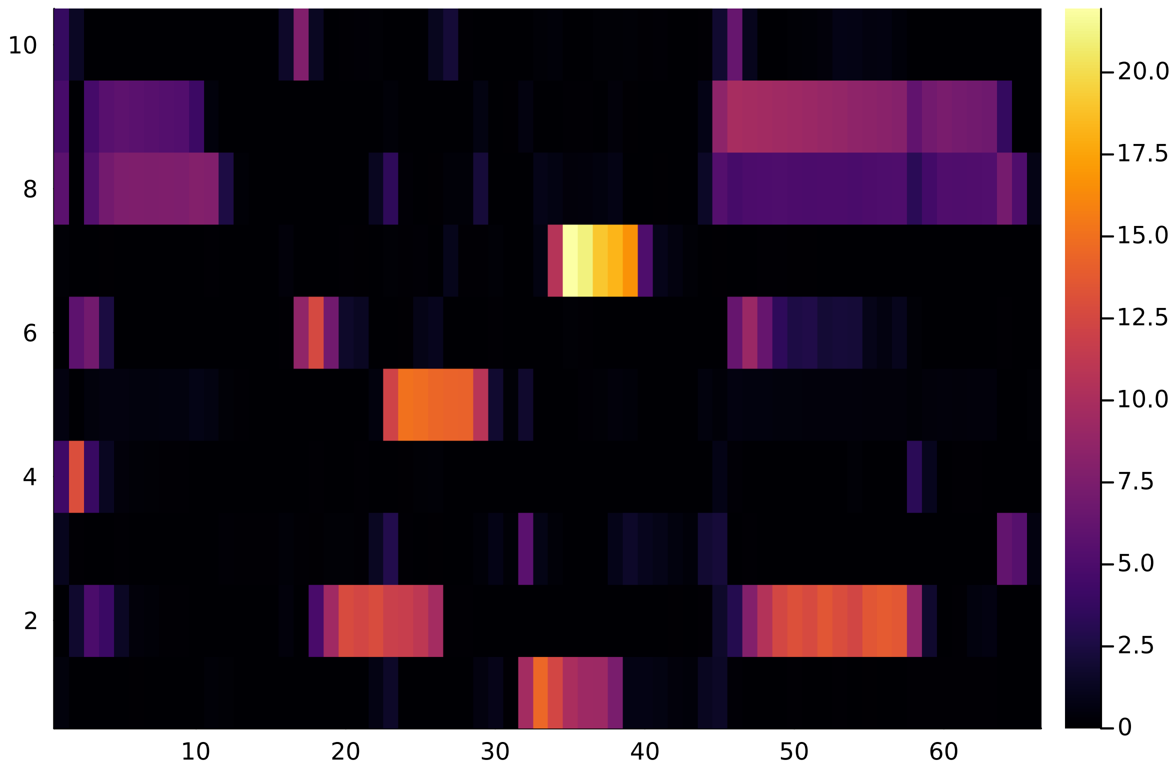


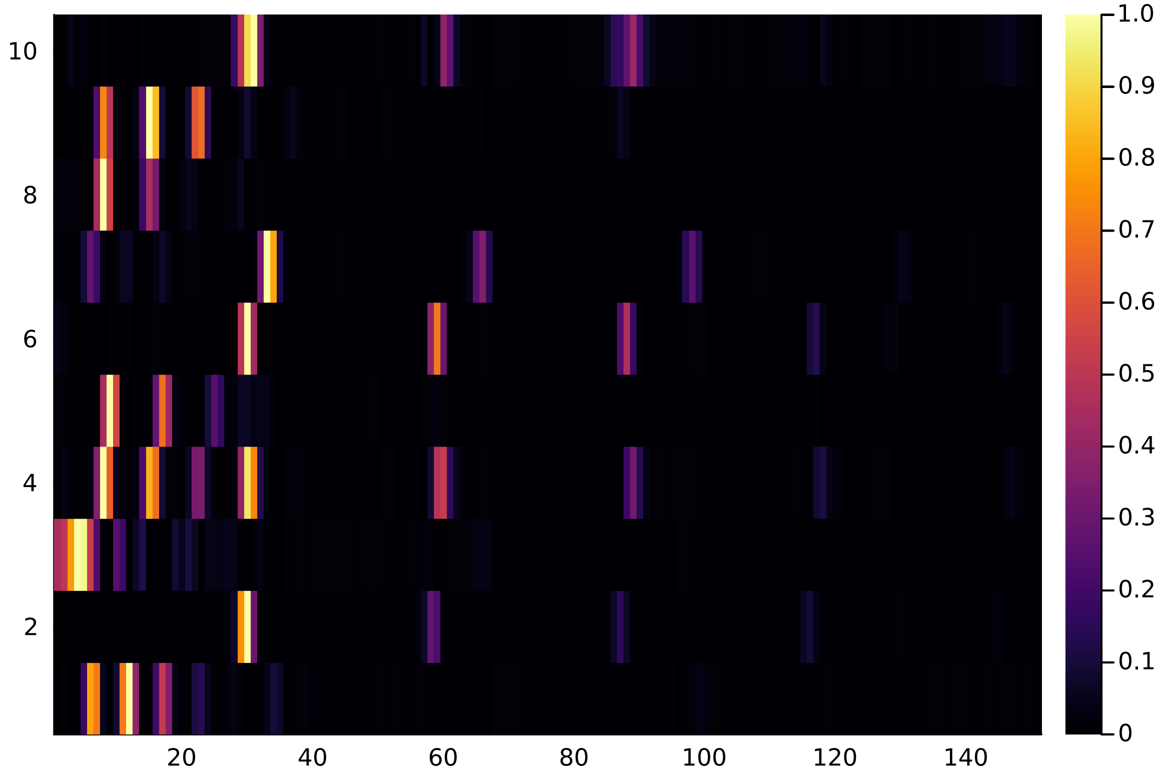
> Experiment Two-step harmonic clustering part 2
>
> #nnmf #music
>
> Trying two-step low rank and harmonic decomposition again with some more hyperparameter tweaks, we get two clusters as expected with only a little bit of mismatch/bleeding.

The main tweek was the increasing the number of "notes" to capture the small changes in pitch throughout a note. And fine tuning the threshold to get the right number of groups (2).

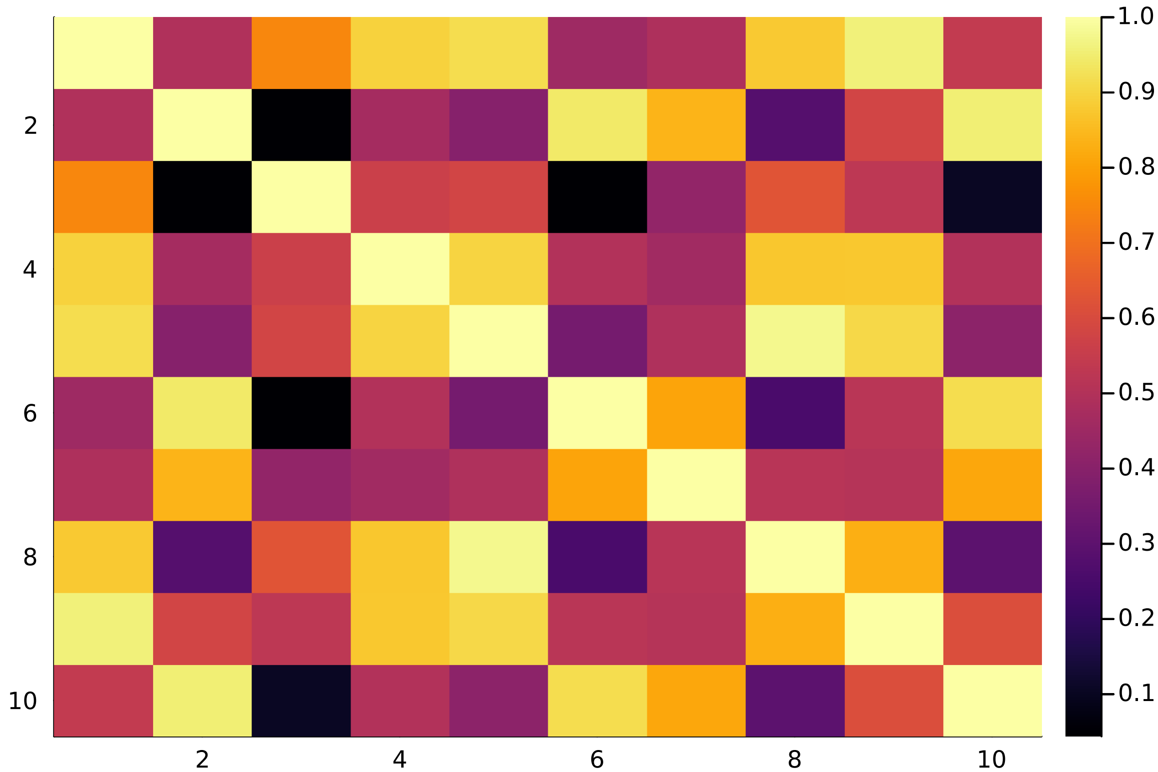
Here are the plots

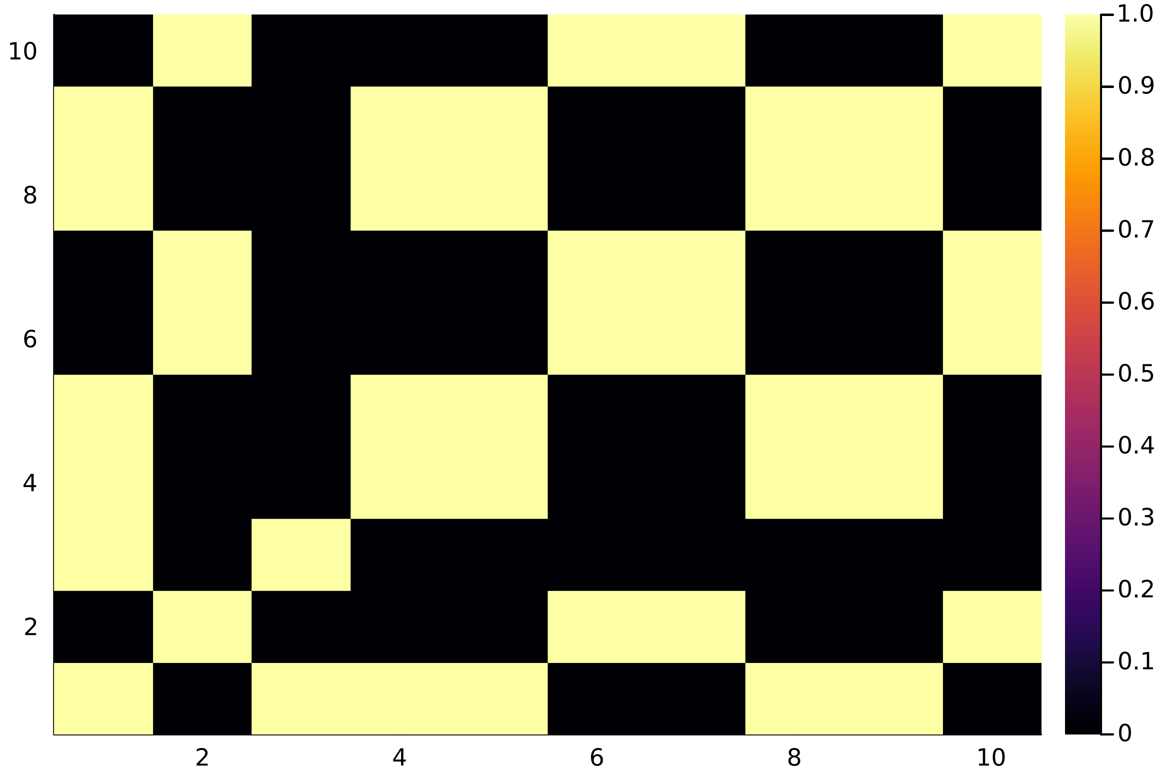
Amplitudes and spectrums of each low rank factor



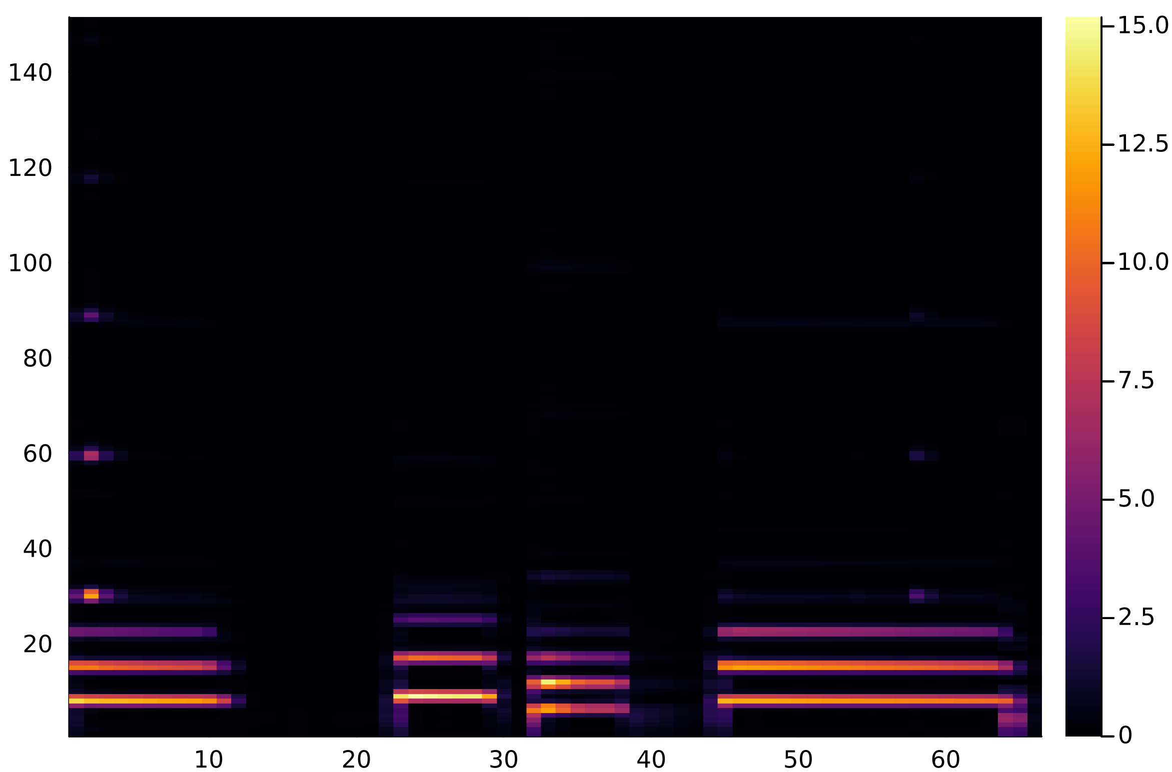


Computed and thresholded cross correlation





Instrument 1



instrument 2