Why does it sound that way?

There are two competing definitions for what a sound is: A sound is a propagation of a pressure wave through a medium—often, air, or it is the actual psychological perception of that wave that we hear. This paper aims to link the two definitions where possible.

A sound (wave) is typically generated by something that vibrates—that oscillates about an equilibrium point. As a vibrating object travels towards or away from its equilibrium point, it will move particles next to it, which will in turn move particles next to it. This chain of movement occurs until it reaches another object and is interrupted. If the object is your ear drum, these particles impact and displace 3 bones, which through a complex system, process these impacts into a sound in the psychological sense.

How we interpret sound is essentially made up by our brains. There is no one to one relation between our perception of sound and how the waves that make up a sound actually are. But there are a few discernable aspects of how we perceive sound that we can measure and quantify.

Pitch

A simple note or pitch, discounting any aspects of sound quality, can be represented by a sine function. As the frequency of the function increases, its perceived pitch increases. Pitches that are an octave apart have a doubling of frequency. That is, each time you double the frequency of a note, you get the same note an octave higher. Most people perceive this interval (or distance between frequencies) as a "higher version of the same note." This phenomenon is called Octave pitch affinity (Laurent Demany, 2021).

The particular details of how notes within an octave are named and obtained, called temperament, is outside of the scope of this paper. But in short summary, historically, musical instruments were tuned based on the interval of the "perfect fifth", or 3:2, because this interval is recognizable by the human ear. By systematically iterating this interval (and the octave), one can create a set of 12 notes (including the octave) that generally sound pleasant together, at least in comparison to just mindlessly adjusting pitch.

But it is important to note that this process, of iterating a fixed interval (called regular temperament), can be used to create an infinite number of scales. There are many examples of musical scales that do not have 12 notes. The system just discussed is known as Pythagorean tuning.

Regardless, the 12-note convention, whether due to merit or just circumstance has become mostly universal today, however, instead of using the perfect fifth to create notes of our scale, the octave is simply broken into 12 equal parts – so that each subsequent note is given by multiplying the frequency by $\sqrt[12]{2}$. This system is called (12 tone) equal temperament.

Loudness

Loudness is one of the more subjective aspects of human sound perception, and there are a variety of things that influence it. One of the greatest influences is sound intensity. Calculating sound intensity is outside of the scope of this paper but is related to amplitude. It can be observed that humans can hear sound intensities from 10^{-12} (watts per meter squared) as the limit of human hearing, and 10^5 at the level that will cause damage immediately (Paul Peter Urone, 2020). However, a simple linear change in the intensity of the sound is not reflected in its perceived loudness.

The most common system, called decibels, tries to measure the loudness of sounds based on intensity. It is a logarithmic scale, which relates the intensity of a sound to a reference (the lower limit of hearing).

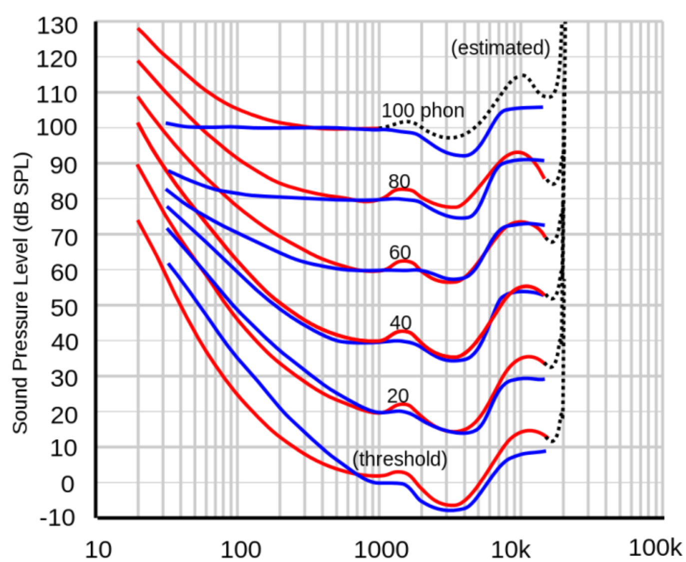
A sound's decibel value β is given by:

$$\beta = 10\log_{10}\frac{I}{I_0}$$

Where I is the sound's intensity, and I_0 is the reference intensity. Essentially, for every 10 times increase in sound intensity, the decibel will increase by 10. So, a sound that has an intensity 100 times greater than the lower limit of hearing will have a decibel value of 20. The decibel scale is advantageous because it compresses the wide range of audio intensities into a scale that more accurately reflects how we hear it.

Oftentimes if you see a dB meter used in audio production it will not have this form. The Deci-bel system (10 * bel) can be used in a number of ways, this is just one way to do it for loudness. In the graph below, a system which measures relative sound *pressure*, not intensity, is used. The dB system can also be used to measure many other things, broadly, those which have some logarithmic behavior.

Unfortunately, there is another complicating factor in loudness. Due to the physical properties of the ear, certain frequencies may be heard more loudly than others. A scale, Phons, which tries to alleviate this was initially developed by researchers Harvey Fletcher and Wilden Munson (H., 2021). By asking the listener to report the perceived loudness of a variety of pure tones, they were able to identify "equal-loudness contours", or bands of frequency and pressure which have the same perceived loudness. Today, more accurate contours are known and used as a standard.



Equal-loudness contours (red) (from ISO 226:2003 revision)
Fletcher–Munson curves shown (blue) for comparison

https://en.wikipedia.org/wiki/Equal-loudness_contour#/media/File:Lindos4.svg

Timbre

When we examine (record) a soundwave emitted from any physical instrument, it's not just a sine wave. It's usually complex, without a clear oscillating pattern. One important tool in engineering and mathematics, Fourier analysis, can help us to process and better understand aspects of timbre, or the quality of a sound.

Fourier's theorem states any periodic function can be decomposed into an (infinite) sum of sines and cosines called a Fourier series, each having a frequency which is an integer multiple of the frequency of the original function (Nave, 2017). While this is not exactly true as stated—the Fourier series will not converge to the original function unless the function is "smooth" (or infinitely differentiable) --for our purposes we may consider it true.

The equation for the Fourier series s_n of a periodic function f(x) with period P is:

$$s_n(x) = A_0 + \sum_{n=1}^{\infty}\left(A_n \cos\frac{2\pi n}{P}x + B_n \sin\frac{2\pi n}{P}x\right)$$

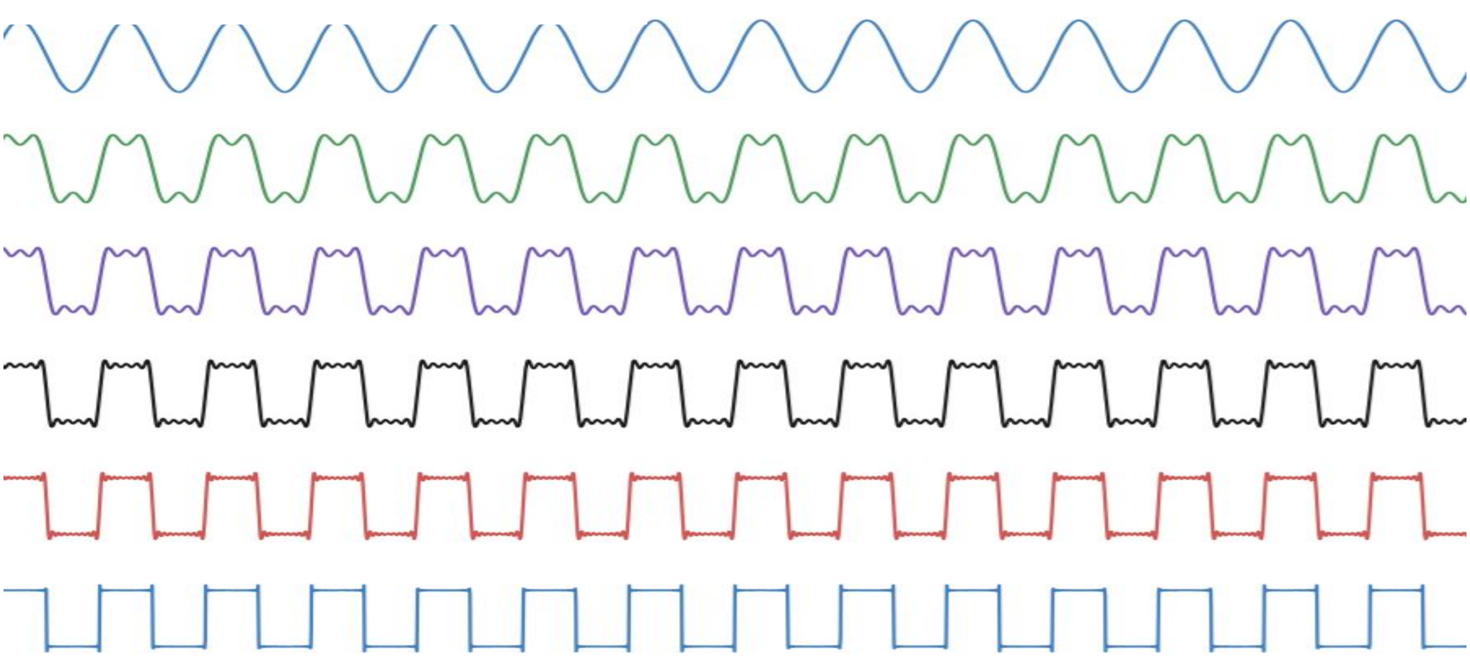
with coefficients

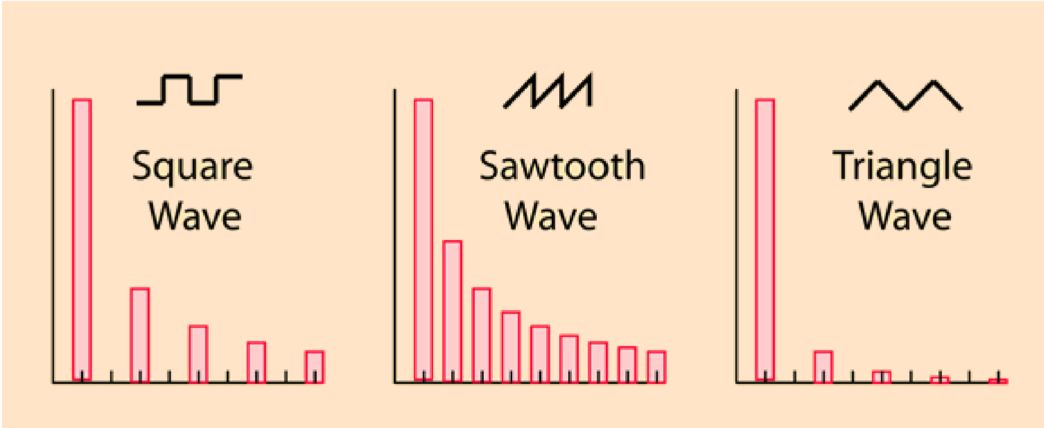
$$A_0 = \frac{1}{P}\int_0^P f(x), \quad A_n = \frac{2}{P}\int_0^P f(x)\, cos\frac{2\pi n}{P}x\, dx, \quad B_n = \frac{2}{P}\int_0^P f(x)\, sin\frac{2\pi n}{P}x\, dx,$$

We can make some headway in understanding Fourier analysis and the timbre of sounds by decomposing some basic geometric waves into their Fourier series and analyzing the amplitudes of each frequency. This is simpler to do because the geometric properties of these waves allow for easier evaluation of the integrals—for example, since the square wave is either 1, or 0, the coefficient integral is simply the integral of the sinusoid over the second half of the period (0.5 to P), which is easy to calculate. Ultimately, the square wave's Fourier series (with period p = 1) is just a sum of sins:

$$SQ_n(t) = \sum_{\substack{k=1 \\ odd}}^{n} \frac{sin(k\pi t)}{k}$$

Fourier series of the square wave for 1, 2, 3, 4, 8, and 100 terms



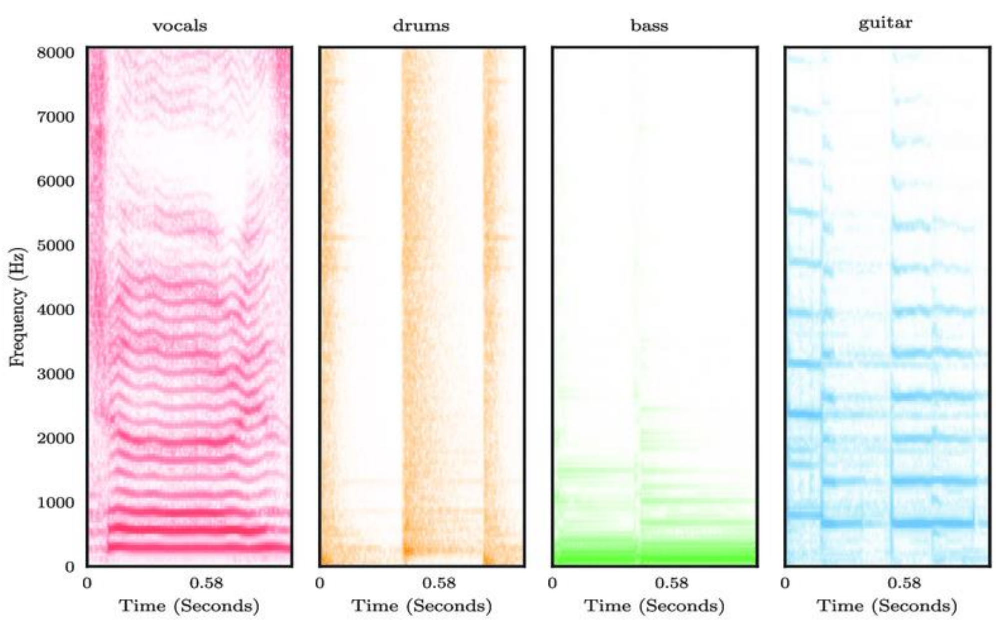


http://hyperphysics.phy-astr.gsu.edu/hbase/Audio/geowv.html#c1

The frequencies of the sinusoid functions are known as "harmonics" (or overtones), and as a whole are referred to as the "harmonic content" of a sound**.** Pictured above is a graph of the harmonic content of three stereotypical soundwaves. The sawtooth can be said to sound "fuller" or "wider" in its sound, and this is perhaps reflected in its higher concentration of harmonics. Both the triangle waves and square waves are notable for having every other (odd) harmonic, but their amplitudes differ, and so does their quality—the square wave is quite harsh in comparison to the triangle wave.

These waves are prototypical in audio engineering because of electrical properties (they are relatively easy to electrically produce), not their actual sound. In fact, they sound quite bad: they are rather unlike the traditional timbres we enjoy from musical instruments and require at least some modification to be tolerable to the average listener. This modification, through adding new harmonics or otherwise modifying the current harmonic content is known as additive synthesis.

We can also look at the harmonic content of real physical instruments by recording them and using a (computer) process called Fast Fourier Transform (or FFT) to process the incoming signal into its harmonic content. Often, this harmonic content varies over time, and we may display this in what is called a spectrogram—where brighter points indicate greater values for coefficients. There is significantly more to the field of Fourier analysis in regard to audio and in regard to many fields outside acoustics, but that is all that will be mentioned here.



https://www.researchgate.net/figure/Magnitude-spectrogram-of-four-example-music-signals-vocals-left-drums-mid-left_fig2_329464103

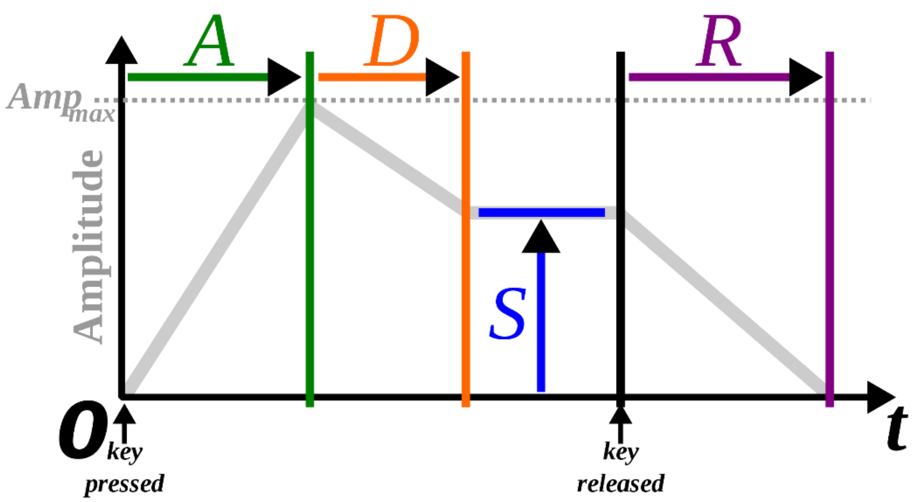
In addition, the envelope, a function outlining the extremes of an oscillator (it's maximum amplitudes over time) are another factor in determining timbre. The ADSR envelope is a common tool used in music production and tries to replicate a general pattern in the volume of a natural sound—a short initial burst of volume, then a decay to a lower sustained volume, before falling silent.

i. Attack: The time it takes for a signal to increase to its maximum amplitude.

ii. Decay: The time is takes for a signal to increase to Its sustained amplitude

iii. Sustain: The sustained amplitude

iv. Release: The time it takes to decrease to 0 amplitude from sustained amplitude



https://en.wikipedia.org/wiki/Envelope_(music)#/media/File:ADSR_parameter.svg

There are other envelopes, featuring additional parameters like "hold" (how long the sustain is held). One may also create a completely custom envelope. Together these two factors (harmonic content and envelope) create timbre. Using synthesizers, we can recreate some timbres that occur in the natural world, but oftentimes the waves produced by a physical instrument are complicated to reproduce—sampling remains a more common and effective tool to recreate instruments virtually.

In conclusion, there is much to say about how we hear sound. There are some relationships between the actual quality of sound waves and how we hear them, like the relation between pitch and frequency, and sound intensity and perceived loudness (and frequency). Most of these relations are well studied and understood, but they are all very complex.

These relationships are crucial in the process of synthesizing sound. Unfortunately, their complexity makes it very difficult to use them directly. Software makes it possible to avoid this complexity, thanks to years of development. I think it is valuable to appreciate the work that goes into making these tools useable without extensive knowledge. Even if they are not obvious the relations are still there, and without a lot of work and math it would not be possible to create or process music in the ways we do today.

Bibliography

H., J. (2021). *Fletcher Munson Curve: The Equal Loudness Contour of Human Hearing*. Retrieved from https://ledgernote.com/columns/mixing-mastering/fletcher-munson-curve/

Laurent Demany, G. M. (2021). The perception of octave pitch affinity and harmonic fusion have a common origin. *Hearing Research*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7614450/

Nave, C. R. (2017). *Sound: Fourier Analysis and Synthesis*. Retrieved from Hyperphysics: http://hyperphysics.phy-astr.gsu.edu/hbase/Audio/fourier.html#c1

Paul Peter Urone, R. H. (2020). Physics. OpenStax. Retrieved from https://openstax.org/books/physics/pages/14-2-sound-intensity-and-sound-level