Concise Fundamentals of Audio

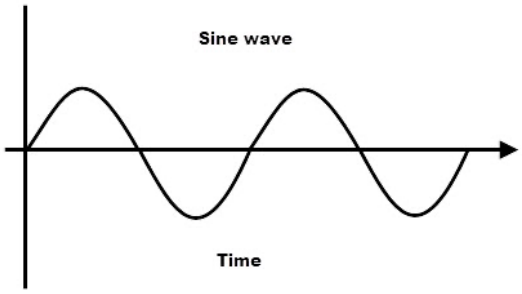
Joel A. Jaffe

INTRODUCTION

This document is intended as a brief handbook for understanding the basic physics of sound and music. I found the impetus for writing such a guide based on my experience as an undergraduate music student in a university department rooted firmly in the conservatory tradition. As someone who plays primarily electric instruments, I was disappointed by the lack of course offerings covering sound physics, signal processing, and music technology. Building my knowledge of these areas in my extracurricular pursuits, I've developed a desire to advocate for pedagogical reform in music departments to position an understanding of sound as a necessary step in understanding music.

SOUND

What humans perceive as sound is the vibration of air molecules as sensed by our ears. Disturbances to static air pressure tend to have **waveforms**, surpluses in air pressure of a certain duration and magnitude that are followed by deficits of equal duration and magnitude. When a sound has a waveform that repeats regularly, it is said to have **periodicity**, meaning that it will be perceived as **pitch**. Healthy human beings are capable of perceiving pitches as slow as 20 periods per second and as fast as 20,000 periods per second. The unit **Hertz** (Hz) is the standard unit used to quantify the cycles (periods) per second of a wave and is augmentable via metric prefixes. 1 Hz is equivalent to one cycle per second, so the range of human hearing could be written as 20Hz-20kHz.



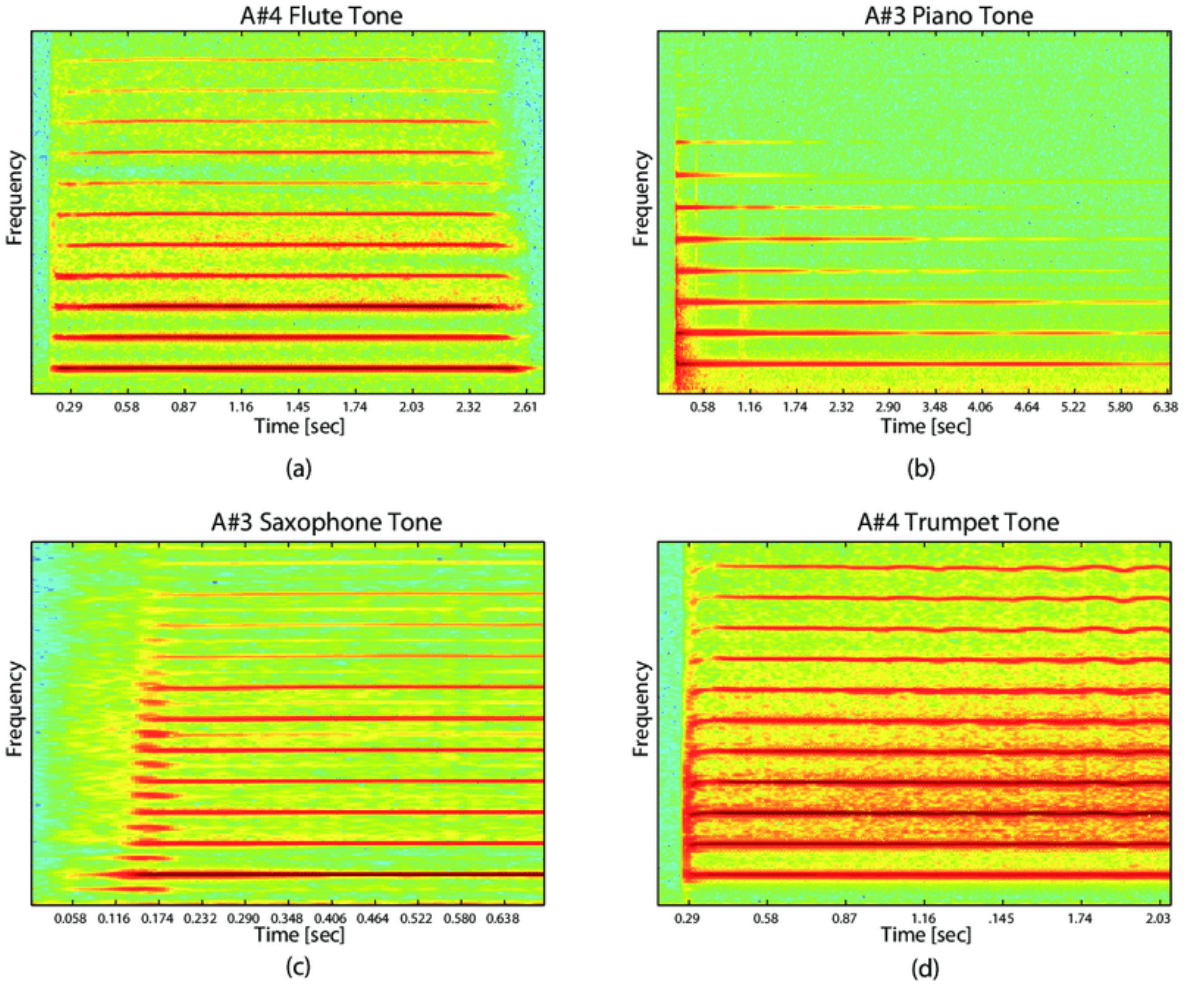
Pictured above is a **sinusoidal wave**, a graphic representation of a steady oscillation between magnitude (y-axis) occurring over time (x-axis). While the sinusoidal waveform is the essential building block of all waveforms, no acoustic sounds occur as pure sinusoids. Due to the nature of the physical materials the world is made of, sound as we know it exists as **complex waves**, which look, sound, and behave very differently than pure sine waves. The French mathematician Joseph Fourier discovered in the 18th century that all complex waves are merely the result of simultaneously occurring sinusoids interacting with each other, and developed a method of analysis that allows for the deconstruction of complex waves into component sinusoids, known as **Fourier analysis**.

In the deconstruction of a complex waveform, the most important component sinusoidal wave is the one of the lowest frequency, known as the **fundamental**. The other component sinusoids, known as **overtones**, tend to be mathematically related to the fundamental. Overtones that are integer multiples of the fundamental are called **harmonics**, and they exist in a consistent additive series known as the **harmonic series**, which can be expressed as *fundamental, 2 times the fundamental, 3 times the fundamental…* extending to infinity.

THE NOTE

When a musical instrument sounds a **note**, it is producing a complex waveform whose **spectral content** spans multiple thousand Hertz. The fundamental is the most important piece of spectral content, as the rest of the **harmonic content** is derived from it. When a string of a fixed length and tension vibrates at 55Hz (A1), it also is vibrating at half its length, producing the 1st harmonic at 110Hz (A2). The series continues upward in pitch, with each successive harmonic generally sounding with lesser volume.

Displayed below are a quartet of **spectrograms**, a tool for graphically displaying the spectral content of a sound. Pitch is displayed vertically (y-axis), time horizontally (x-axis), and the volume (amplitude/pressure/magnitude) is displayed as color. Notice that successively higher harmonics do not cleanly diminish in volume relative to one another. These discrepancies, unique to each instrument based on the physical materials it is made of, give each instrument its unique tone, known as its **timbre**.



https://www.researchgate.net/publication/277010607_Pitch-informed_solo_and_accompaniment_separation

TUNING AND TEMPERMENT

The relationship between a note's fundamental and first harmonic constitutes the musical **interval** known to Western music as an **octave**. Similar to how the harmonic series' content is derived from the fundamental, the 12 notes of western music are derived from the harmonic series. The second harmonic (*fundamental times 3*) is approximately equivalent to a perfect 5th (P5), and the fourth harmonic (*fundamental times 5)* is approximately equivalent to a major third (M3). Why only approximately? Because western music is intentionally "out of tune" with the harmonic series. In a **temperament** system that follows the harmonic series exactly, known as **just intonation**, the intervallic distance between C and D is different than the distance between D and E. Consequently, just-tempered keyboard instruments must be retuned to play in specific keys, and pieces that change key are unplayable. **Twelve-tone equal temperament (12-ET)**, standard practice in western music since the mass adoption of the pianoforte, allows for playing in all twelve keys without retuning, albeit with the accepted sacrifice that no instrument is "truly" in tune.

Functional harmony is often discussed in terms of consonance vs. dissonance, tension vs. release, stability vs. instability etc. The intervals we label as **consonant** are those derived from harmonics with simple mathematical ratios to the fundamental, and the intervals we label as **dissonant** from harmonics with more complex ratios to the fundamental. Below is a diagram ranking the intervals of Western music by increasing dissonance, based on the work of Hermann L.F. Helmholtz.

Interval Evaluation	Interval Name	Short Name	Interval Ratio
Absolute consonances	Unison	P1	1:1
	Octave	P8	1:2
Perfect consonances	Fifth	P5	2:3
	Fourth	P4	3:4
Medial consonances	Major sixth	M6	3:5
	Major third	M3	4:5
Imperfect consonances	Minor third	m3	5:6
	Minor sixth	m6	5:8
Dissonances	Major second	M2	8:9
	Major seventh	M7	8:15
	Minor seventh	m7	9:16
	Minor second	m2	15:16
	Tritone	TT	32:45

https://comparative-cognition-and-behavior-reviews.org/vol12_toro_crespo-bojorque/

SIGNAL: THE ELECTRIFICATION OF SOUND

In 1876, Alexander Graham Bell invented the **microphone**, a device central to the development of the telephone. A microphone is a type of **transducer** that converts the mechanical energy of sound waves to electrical **signal**, translating the oscillations between high and low air pressure into oscillations between high and low voltage. This invention led the way for **sound recording** and **playback** technology, which in the 20th century would become the dominant way people consume music. Sound recording is the process of capturing and storing sound, and playback is the process of recreating a captured sound event. The degree to which a system of recording and playback is capable of faithfully reproducing a sound event is known as its **fidelity**. While microphones are used to capture the sound of **acoustic instruments**, **electric instruments** typically use magnets called **pickups** to translate the oscillation of metal strings into electrical signal that serves as the primary output of the instrument. **Electronic instruments** do not capture physical oscillation, but instead **synthesize** electrical signal from input electric current, and therefore do not use transducers. The user interface for manipulating electronic instruments is called the **controller**, and keyboards (based on the piano) have historically been the dominant type of controller.

SIGNAL PROCESSING

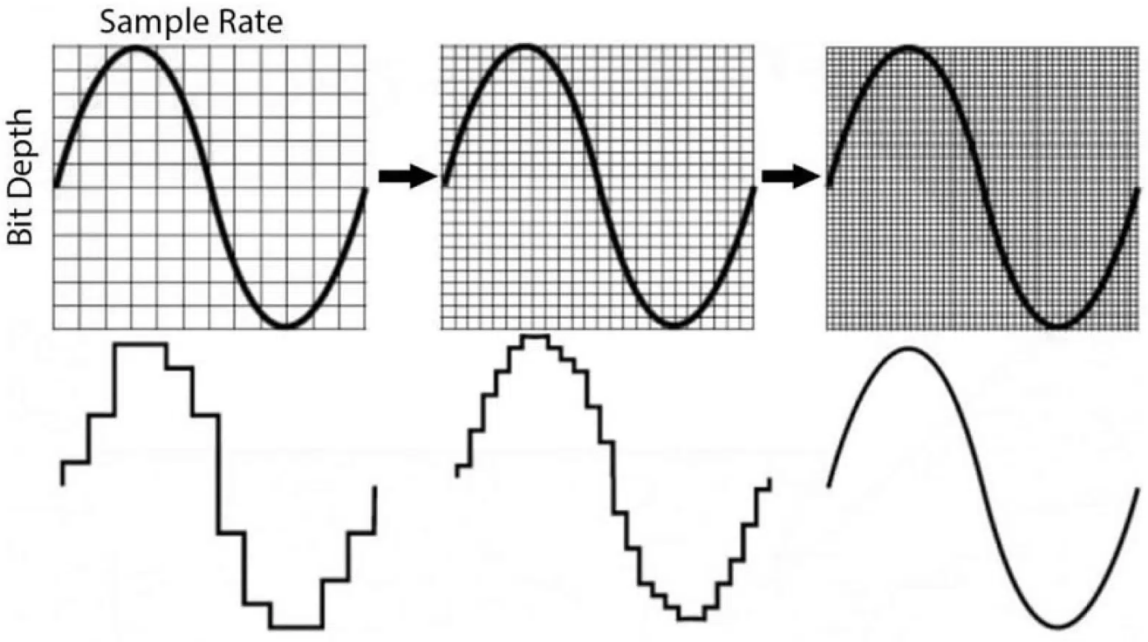
As the so-called **recording industry** developed in the first half of the 20th century, **audio engineers** began to experiment with methods of modifying the electrical signals sound was being represented as. Motivations for this arose from a lack of fidelity in recording devices, which necessitated modification of the recorded audio to appease the human ear. As electric instruments rose to prominence in the second half of the 20th century, fidelity became less important than the pursuit of novel sounds that acoustic instruments were incapable of producing. Common processes (then and now) include **dynamic-range compression**, **equalization (EQ)**, and **time-domain** effects such as **reverb** and **delay**.

DIGITAL: SIGNAL BECOMES DATA

So called "**analog**" technology is a designation born from the fact that the electrical signals created via transducers are **analogous** to the physical vibrations being represented. Since the advent of **digital signal processing (DSP)**, the representation and modification of sound as **binary code**, the term analog has come to denote any signal processor that is not digital. The question at the inception of DSP technology was how to represent a smooth curve (**continuous data**) as binary code, which can only handle stepwise data (**discrete data**). The answer is that methods for encoding signal are not truly smooth curves, but curves chopped into steps so small that they are majorly imperceptible to the human ear. Devices and programs that convert analog signal to digital are called **analog to digital converters** or **ADC**s. Devices or programs that perform the opposite are called **digital to analog converters** or **DAC**s.

The standard **protocol** for the **encoding** of audio signal is the .WAV format, developed by IBM and Microsoft. The .WAV is considered **lossless** because no data compression is being

applied, which at the service of fidelity results in large file sizes. The .WAV encodes audio by breaking the two-dimensional data stream of changing amplitude over time into discrete amplitudes **sampled** at a rate well above the upper limit of frequencies perceived as pitch. The **Nyquist Theorem** proposes that in order to accurately encode and reproduce waveforms, the **sample rate** must be at least twice the frequency of the wave. Consequently, in order to accurately represent the entire spectrum of heathy human hearing (20Hz-20kHz), standard sampling rates have settled at 44.1kHz and 48kHz. The amplitude of each sample is measured as a decimal between 0 and 1, with the rounding place determined by the **sample depth** or **bit depth** of the file. Standard sample depths are 16 and 24 bits. A stereo .WAV file with a bit rate of 48kHz and bit depth of 24 bits will constitute 17.28 **megabytes** of data per minute of audio.



https://unison.audio/sample-rate/

THE FUTURE OF AUDIO SIGNAL PROCESSING

The advent of digital signal processing has dramatically altered the ways in which humans interact with music. The encoding, storage, and transmission of audio as binary has given rise to the **streaming industry**, a model of music distribution where individuals can access nearly all music ever recorded via the **internet** with a subscription service. With the development of **digital audio workstations** (DAWs) and the **digitization** of analog signal processors, affordable software now allows for anyone with a personal computer to create and distribute music. As **machine learning** automates the process of **analog emulation**, research in DSP techniques will be liberated towards the further development of sounds analog technology was never capable of producing. As digital technology continues its pursuit of the **seamless**, digital audio technology will become continually more responsive to natural performance gesture, and the development of the **brain-computer interface** may lead to the ability to record music direct from one's mind.