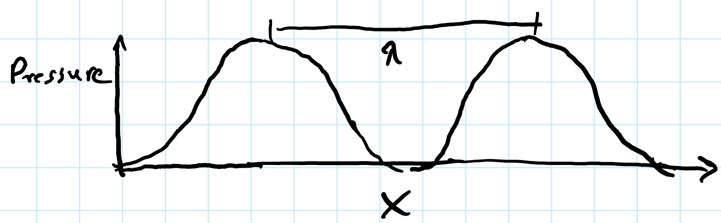
1) Basics of Audio

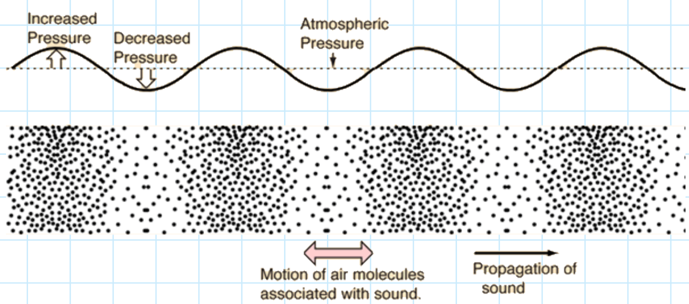
Sunday, June 18, 2017 4:19 PM

Working with Audio Data

What is sound?

Sound waves are travelling patterns of compressions and rarefractions of particles in a medium (e.g. air). This corresponds to regions of high and low pressure, respectively





λ: wavelength of a sound wave is the distance between high-compression regions: meters
- more generally, it is the distance between identical points on consecutive wave fonts.

f: how many "peaks" pass through a given point every second: 1/s (Hz)

In general, these waves propagate through the medium, thus the pressure wave is a function of x and t:

$$f(kx - \omega t)$$

Any function of this form will propagate, in the $+x$ direction, without changing shape at a speed ω/k. Pressure waves travel at a nearly constant speed in air. $340\,\frac{m}{s}$ (761 mph)

A single note is produced by a pressure wave that varies as a propagating sinusoidal wave:

$$p(x,t) = A\sin(kx - \omega t)$$

A: amplitude

$p(x+\lambda, t) \equiv p(x,t)$

$k(x+\lambda) = kx + 2\pi$

wavelength: $\lambda = 2\pi$

$p(x, t+T) \equiv p(x,t)$

$\omega(t+T) = \omega t + 2\pi$

$T = 2\pi$ ... $f = \omega$

$p(x+\lambda, T) \equiv p(x,T)$ $\quad\quad$ $p(x, t+T) = p(x,t)$

$k(x+\lambda) = kx + 2\pi$ $\quad\quad$ $\omega(t+T) = \omega t + 2\pi$

wavelength: $\lambda = \frac{2\pi}{k}$ $\quad\quad$ $T = \frac{2\pi}{\omega} \rightsquigarrow f = \frac{\omega}{2\pi}$

$$\text{Speed of sound: } v = \lambda f \quad (340\ \tfrac{m}{s}\ [761\ mph])$$
$$= \omega / k$$

"Pure" soundwave, travelling in $+x$ direction w/ speed $v = \lambda f$

$$p(x,t) = A \sin\left(\frac{2\pi f}{v} x - 2\pi f t\right)$$

v is constant, thus there are only two parameters that control a "pure" sound wave:

- f: the frequency (i.e. the note)
- A: the amplitude (i.e. the loudness)

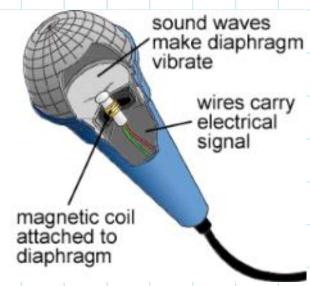
speed of sound

$v = \lambda \cdot f$

λ wave lenghtl

How do we record sound?

A microphone is designed to convert the vibrations created by an impinging soundwave into an electrical voltage via a 1-to-1 mapping. That is, the voltage can be "decoded" so that it can drive a soundwave, via a speaker, that is the same as the soundwave it recorded.



Because the microphone is stationary, the varying voltage is only a function of time:

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

$$P(t) = A \sin\Big(\underbrace{\tfrac{2\pi f}{v} x_0}_{\text{constant}} - 2\pi f t\Big) \rightarrow V(t) = V_{max} \sin(2\pi f t + \underbrace{\phi}_{\text{constant}})$$

microphone is located at x_0, thus the pressure wave is only a function of time

Pressure wave drives oscillating voltage in mic's coil. 1-to-1 correspondence w/ pressure wave

The voltage signal needs to be encoded into binary memory. This is done by sampling the signal (i.e. record the voltage value) at regular intervals.

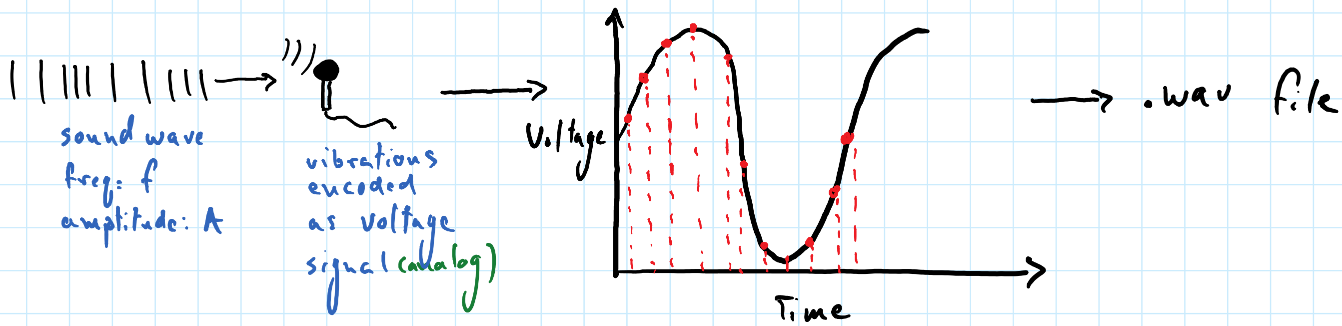
The defacto industry standard is to sample the signal 44,100 times every second. That is, us a sampling rate of 44.1 kHz (kilo-Hertz).

For each sample, the voltage value is encoded as a 16-bit value, meaning that it is an integer $0 - 2^{15}$, and has a $+$ or $-$ sign (2^1)

$2^{15} \cdot 2^1 \rightarrow 2^{16}$ 16 bits!

There are additional details regarding how to map a voltage to its corresponding binary representation, but that is beyond our scope.

Pulse-code modulation (PCM) is the standard method for converting the voltage (an analog signal) to a digital representation.



The signal is sampled at a constant rate (44.1 kHz), and the voltages are binned into a 16-bit encoding (digital)

Ultimately, the wav file is a header that contains the sampling rate, and is a sequence of 16-bit numbers - the recorded voltages.

Audio signal notebook