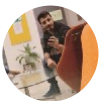




CRUNCHING-AUDIO-DATA

Sound Wave Basics — Every Data Scientist must know before starting analysis on Audio Data

An intuitive overview of sound waves for getting started with audio analysis



Kartik Chaudhary

Dec 16, 2019 · 9 min read

1. Introduction

Humans are born with incredible abilities. Hearing sound is one of the most awesome abilities we have. The sound makes our life easy, It makes us aware of surroundings, possible dangers. Blind individuals use their ears to see the world. The sound makes communication easy for us. Listening to music keeps us entertained.

This amazing ability of ours leaves us with a bunch of questions about the sound waves

1. *What is sound? How does it travel in the air?*
2. *Does it travel underwater?*
3. *How does microphone record songs?*
4. *How do we store sound in the memory?*
5. *How does Loudspeaker work?*

Are you curious to know the answers ?

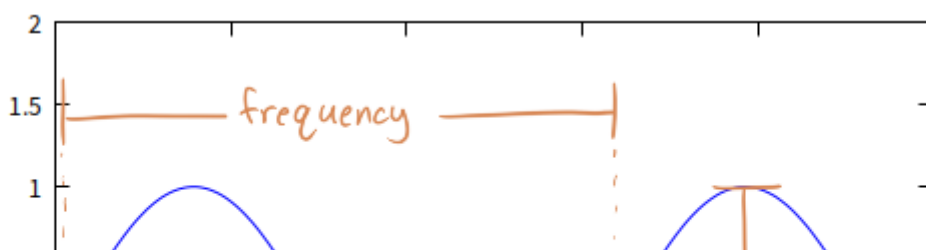
We will try to answer all these questions in this post. Just to give everyone an idea about the content, we will go through the following topics in the same order —

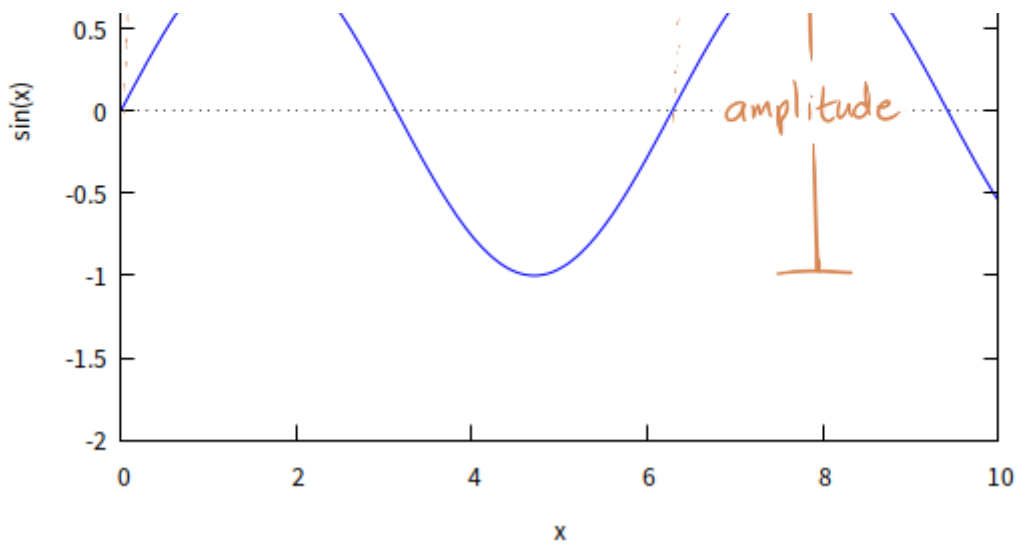
1. Waves
2. Sound Waves
3. Recording Sound Waves
4. Playing Recorded Sound
5. Storing Sound efficiently

. . .

2. Waves

A disturbance traveling through a medium is called a wave. Disturbance can be understood as something which can change the orientations and positions of particles (initially at equilibrium) of a particular medium. Medium is something that carries a wave (disturbance from one place to another). The medium could be any substance or material like — Water, Air, Steel....etc. Remember, the **medium is not responsible for creating waves**, it just helps the waves to transfer energy from one place to another.



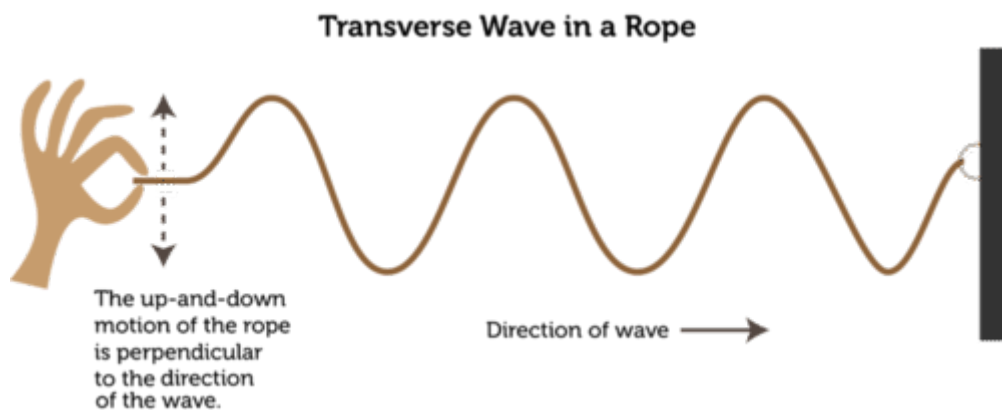


There are many different types of waves like — Mechanical waves, electromagnetic waves, matter waves ... etc. Electromagnetic waves are a special type of waves that do not require any medium to travel, although they can travel through different types of mediums also. For example — light waves are electromagnetic waves that can pass through air, water as well as vacuum.

Looking at their way of propagation in the medium, waves can be classified into three major categories —

1. Transverse Waves
2. Longitudinal Waves
3. Surface Waves

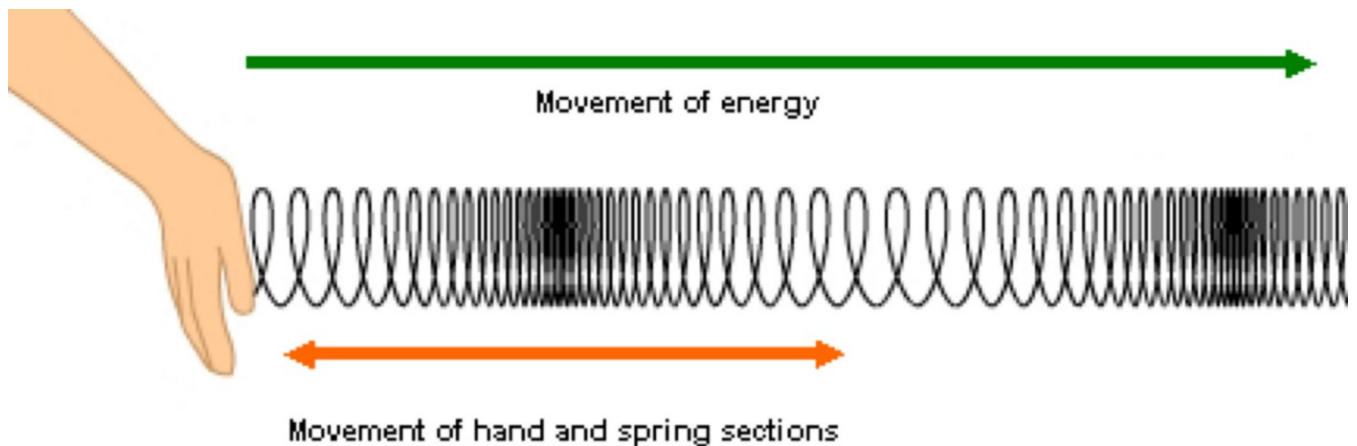
Transverse Waves



If the disturbance caused in the medium is perpendicular to the direction of wave propagation, the wave is called a transverse wave. Medium particles oscillate at 90

degrees to the direction of the wave. Examples — light waves, radio waves.

Longitudinal Waves



If the disturbance of the medium is in the **parallel (same or opposite)** direction to the direction of wave propagation, the wave is said to be a longitudinal wave. Medium particles oscillate in the same or opposite direction to the direction of the wave. Examples — sound waves, ultrasound waves.

Surface Waves



Surface waves travel along with the **interface** between two different mediums. They often travel in a circular manner over on the interface. Examples — ocean waves

generated by gravity on the surface of the water, waves generated by throwing a stone in still water.

. . .

2. Sound Waves

Overview

Sound waves are mechanical waves as they require a medium. Sound waves transfer energy from one place to another. The flow of energy is always in the same direction as of the wave. Sound waves can travel through air, liquid and also solid mediums. Sound travels slowest in the air, much faster in the liquids and fastest through the solids as medium particles are much closer (tightly-packed) in liquids and solids compared to air particles.



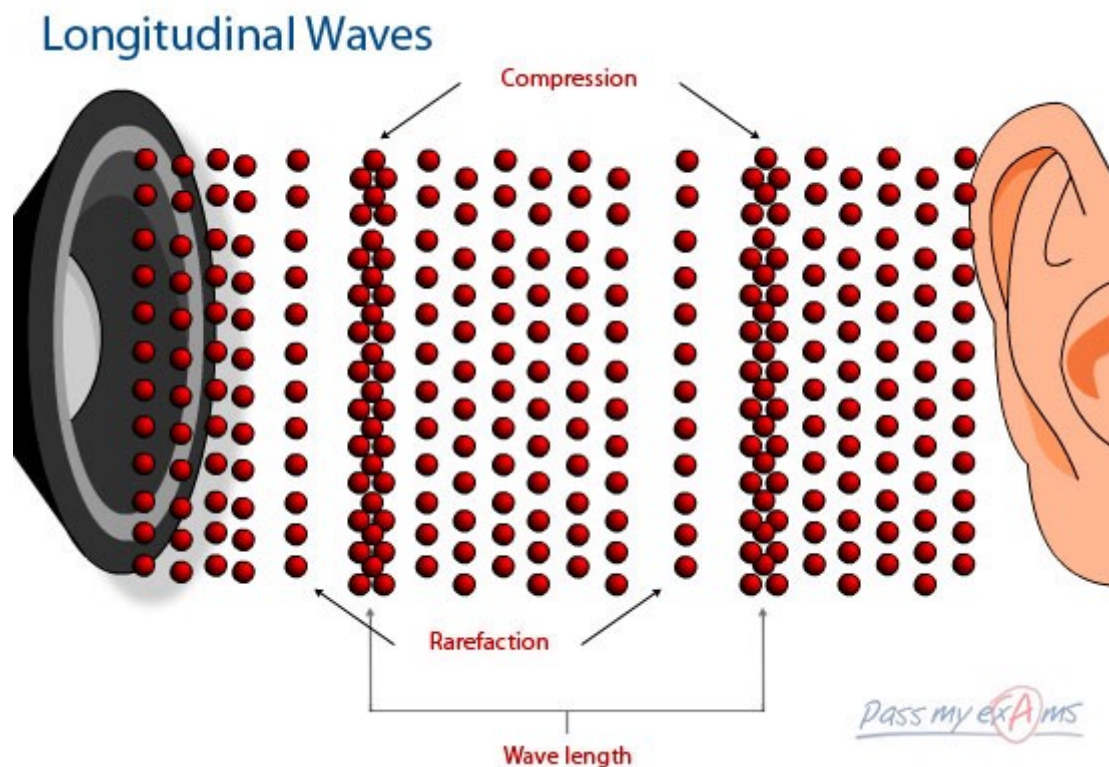
Sound waves travel as longitudinal waves in the 'air and liquid' mediums. While in **solid mediums**, it can travel as a **longitudinal wave as well as a transverse wave**. Moving forward, we will only talk about the sound waves passing through the air.

Sound Waves in the Air

When we speak, we change the pressure of the air closest to our mouth. This change in the pressure (disturbance of the medium for this case) travels in the atmosphere and is called a sound wave. Because this change of pressure in the air travels in the same direction as of the sound wave, it makes it a longitudinal wave.

-----Three important properties of Sound Waves-----

1. Mechanical wave
2. Pressure wave
3. Longitudinal wave



The definition of a sound wave would be — 1. Sound is **energy carried by vibrations** in the air. 2. Sound is a longitudinal pressure wave. It is made up of compressions and de-compressions (also called **rarefaction**) which travel in the atmosphere.

Sound Generation

To generate a sound wave, you need to **compress/put-pressure** on the air. Anything which **vibrates** or is capable of changing the air pressure can create a sound wave. For example — when we clap, slap or smash, we disturb the air nearby and this disturbance moves as a sound wave.

Sound Speed in Air

When a sound wave is generated in the air, the pressure disturbance (compressions and decompressions in the air) travels at **~330 to ~340 metres per second** speed. This is called the speed of sound in the air. The speed of sound highly depends upon the atmospheric temperature. It increases with the increase in atmospheric temperature. Because the gas molecules get more freedom and energy at high temperatures.

Sound Frequency

Number of **air compression-decompression pairs** done by the disturbance in **one second** is called the frequency of a given sound wave. Something vibrating at a certain frequency generates a sound wave with equal frequency. But in reality, it's hard to find a wave with just a single frequency because there are multiple unknown factors that cause pressure-change in real life. This kind of additional pressure-change due to unknown (unwanted) factors is termed as **noise**.

Human Speech

When we speak (talk), we generate multiple sound waves with multiple frequencies simultaneously. A collective pressure change caused by all these waves and surrounding noise travels in the atmosphere. When this collective pressure wave reaches our ears, we are able to hear it.

. . .

3. Recording Sound Waves

We know that sound is nothing but a continuous change in the air pressure. If we want to record it, all we need to do is **measure and record** the air pressure in the atmosphere with time. Now there are two challenges in doing that —

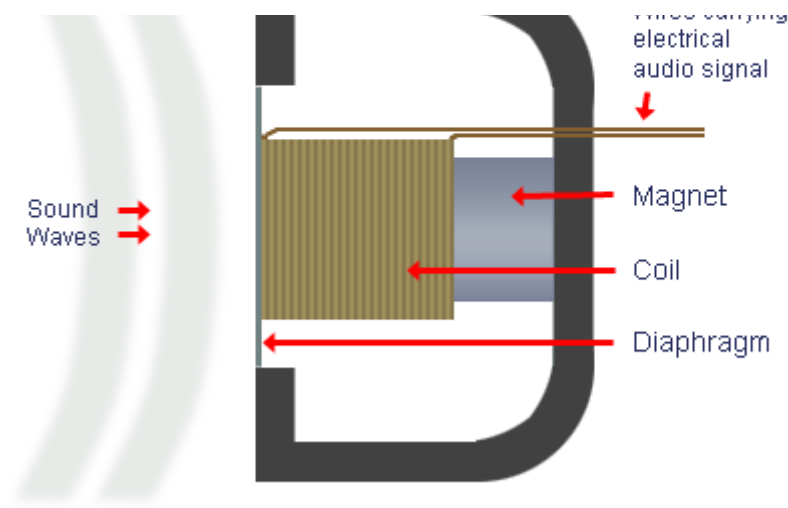
1. How to measure it → **Microphones**
2. How frequently do we need to measure it → **Sampling Rate**

Microphones

Cross-Section of Dynamic Microphone



Wires carrying



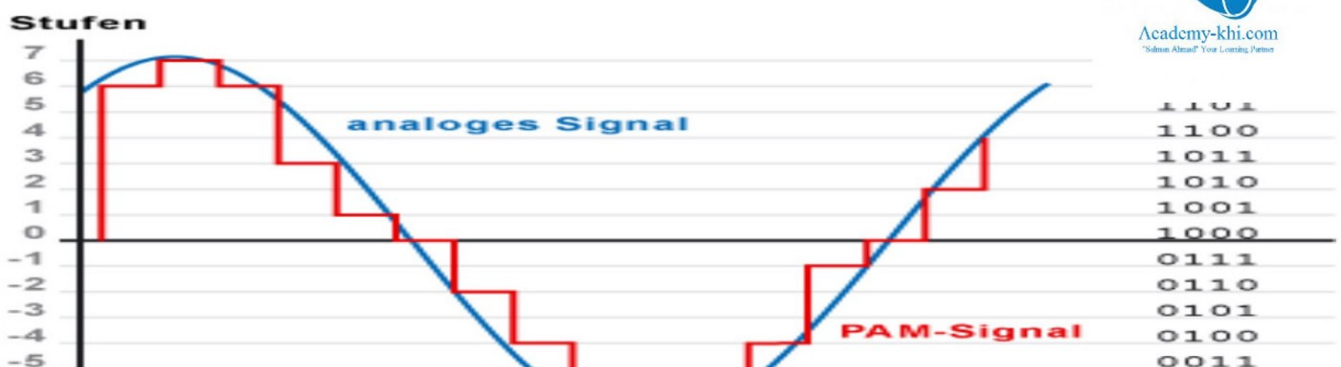
inside view of a microphone

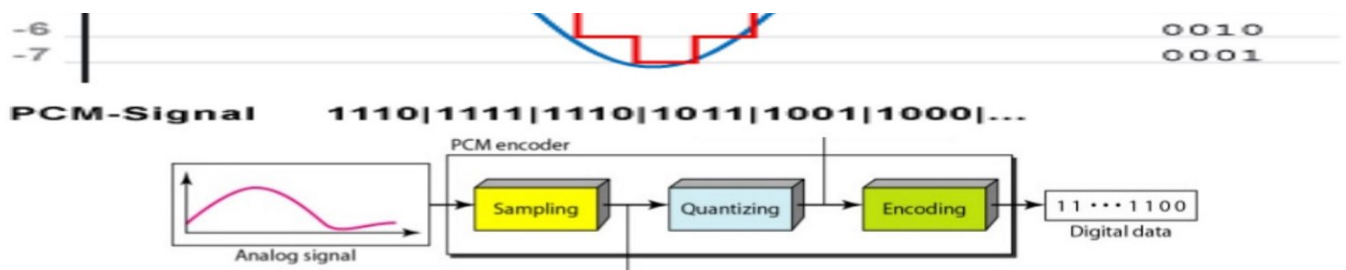
Microphones are the devices capable of converting mechanical energy of sound waves into electric energy. When a pressure wave (sound wave) hits the diaphragm (usually made of thin plastic) inside the microphone, the diaphragm also moves along with the disturbance with the same frequency. A metal coil, which is in the contact of a fixed magnet and also attached to the diaphragm, moves back and forth with it. Because this movement of coil cuts through the magnetic field generated by the fixed magnet, an electric current flows through the coil. Now we can record this electric current, and our job is done.

Pulse Code Modulation

Now that we have a way to record sound waves as electric current using the microphones. Storing this wave is still a problem as this is a continuous wave. In order to store it, we need to convert this continuous signal into a discrete signal. Once we have discrete values of electric voltages at regular time intervals, we can directly write them into a file and save. This way of storing an analog signal as a digital signal is called Pulse Code Modulation (PCM).

Pulse Code Modulation



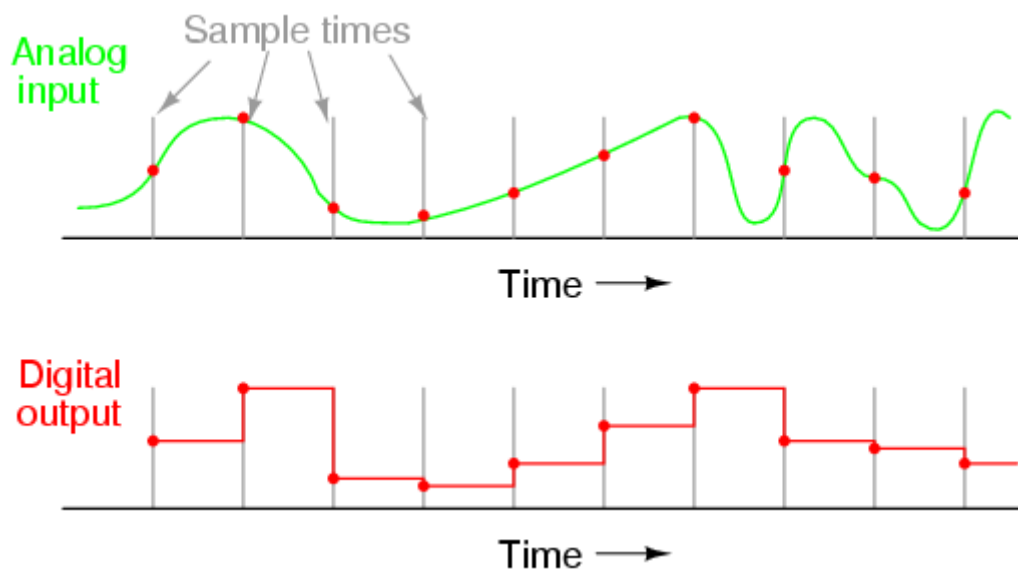


Sampling Rate / Sampling Frequency

The frequency at which we capture these electric voltages (amplitudes), is called the **sampling rate** of the sound file. In other words, Number of electric voltage values noted down in one second is called the sampling rate or sampling frequency of the recorded file. How much sampling rate is good for recording songs? Nyquist's theorem will answer —

Nyquist-sampling theorem

According to Nyquist — Your sampling rate should be at least twice to the maximum frequency you want to capture from the given signal. In other words — If you are given a signal having frequencies ranging from 1 to f Hz, and you don't want to lose any information (frequency), your sampling rate (F) should be ($F \geq 2 * f$).

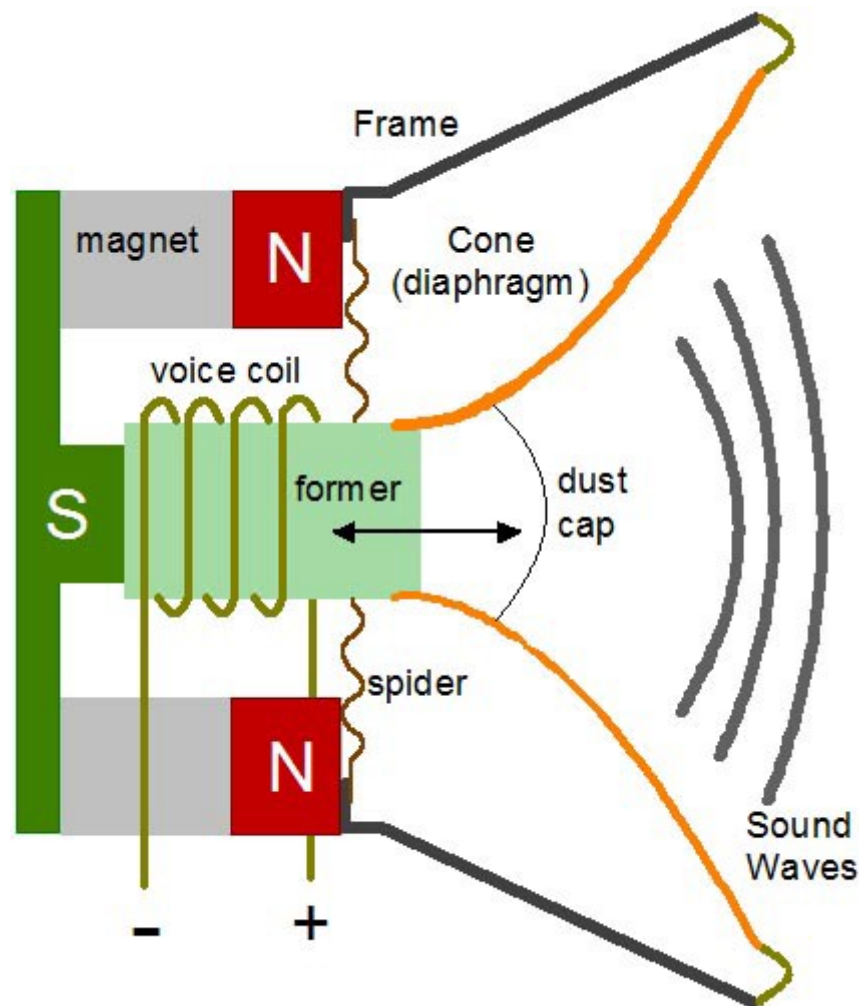


Because the human hearing range is (~ 20 to $20,000$ Hz) there is no point in capturing frequencies greater than 20kHz . So, according to Nyquist, if we sample it at $\geq 40\text{k}$ sampling rate, we won't lose any information which is in our hearing range. This is the reason, **most songs are sampled at 44.1kHz sampling rate.**

5. Playing Recorded Sound

Remember when we recorded the sound, we converted vibrations of the air into electric current, Here we need to do just the opposite. We need to take electric current as input and convert it into the vibrations. A **loudspeaker** (formally called speaker) is the perfect device that does all this.

Speakers



Inside view of a loudspeaker

Speakers are made up of three basic things — a coil, a magnet, and a diaphragm. When changing electric current is passed through the coil of metal wire, it creates a magnetic field around it. This magnetic field comes in contact with the magnetic field created by a fixed magnet. Because the magnet is fixed, the only coil moves back and forth (because of the clash of two magnetic fields). Diaphragm disc connected to the coil also moves with the coil. This movement of the diaphragm pushes the air back and forth and results in a sound wave that we can hear.

. . .

4. Storing Sound Efficiently

Human voices or songs are sampled at a very high frequency. Every second we get few thousands of amplitudes (equal to the sampling rate) to store. As we know each amplitude will take 8 bit of memory if stored in a single byte or 2 bytes if stored as a 16-bit integer. So for a sampling rate of 44k (generally used to record songs), each minute of audio will take ~5MB memory (considering 2 Bytes per amplitude value), which is huge.

Looking at the recorded audio, we see that there is a good percentage of amplitudes having 0 (silence) values. As a 'zero' can be represented in a single bit, why are we wasting the other 15 bits? Idea is to design an algorithm that can store audio data using the minimum number of bits without losing the quality of audio.



Popular audio codecs

There are quite a few algorithms to store audio data efficiently. These algorithms are known as **audio codecs**, for example- MP3, WMA.....etc. These codecs can be classified into two categories —

1. Lossless Codecs

2. Lossy Codecs

Lossless Codecs

An audio codec is said to be lossless if it preserves all the information of the original audio. In other words, when compressed data is decompressed it produces the same quality as it was before compression. Examples —

1. Free Lossless Audio Codec (FLAC)

2. Windows Media Lossless (from Microsoft)

3. Apple Lossless

Lossy Codecs

Lossy codecs discard some information in order to make compressed audio smaller. Human ears possibly won't notice the difference in the quality. These codecs make the file smaller, thus storage, as well as transfer over the internet, becomes really fast. Popular lossy codecs are —

1. MP3 (MPEG — Motion Picture Experts Group)

2. Windows Media Audio (WMA — from Microsoft)

3. Advanced Audio Codec (AAC)

Conclusion

Although there is much more to Sound waves and Signal Processing Techniques. But I hope, this article is good enough for the people who are from Machine Learning (Artificial Intelligence) backgrounds to get started with the Audio Analysis field. As this post is already quite big, I will be covering the next steps (like Reading audio in python, FFT, Spectrogram creation...) in a separate post.





Waves are Awesome!

Thanks for reading, please let me know your thoughts/ feedback.

[Sound](#) [Speech Recognition](#) [Audio](#) [Data Science](#) [Audio Analysis](#)

[About](#) [Help](#) [Legal](#)