Audio & Speech Technology

[1] Background

Sound/Audio/Voice/Speech ?

Sound

Mechanical wave that is an oscillation of pressure transmitted through a solid, liquid, or gas, composed of frequencies within the range of hearing.

Audio

Audible sound coming from a recording, transmission or electronic device.

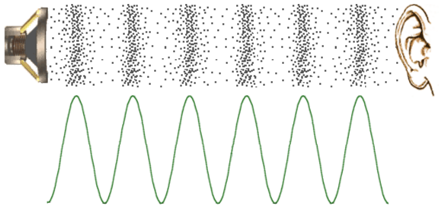
Voice

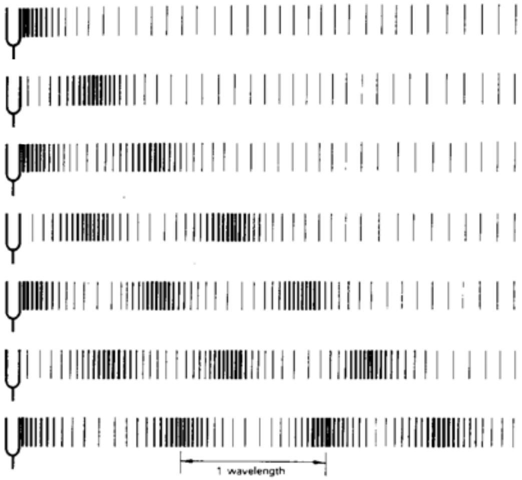
The sound produced by the vocal organs of a vertebrate, especially a human.

Speech

Vocal communication which is an expression thoughts, feelings, and ideas orally.

Physics of Sound





- Waves of pressure
- Longitudinal wave

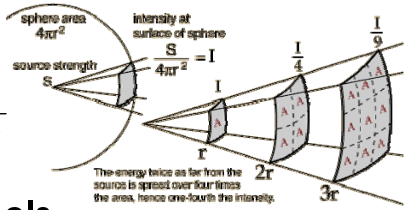
http://www.mediacollege.com/audio/images/loudspeaker-waveform.gif

http://personal.cityu.edu.hk/~bsapplec/Fire/Image222.gif

Sound Intensity

Sound Intensity

$$I = \frac{Sound\ Power}{Area} = \frac{Sound\ Power}{4\pi r^2} \propto \frac{1}{r^2}$$



Sound Intensity Level in Decibels

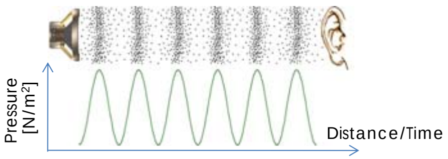
$$I_{dB} = 10 \log_{10}\left[\frac{I}{I_0}\right]$$

Standard threshold of hearing intensity

$$I_0 = 10^{-12}\ watt / m^2$$

http://toonz.ca/bose/wiki/images/1/1e/IntensitySurfaceSphere.gif

Sound Pressure



Sound Pressure

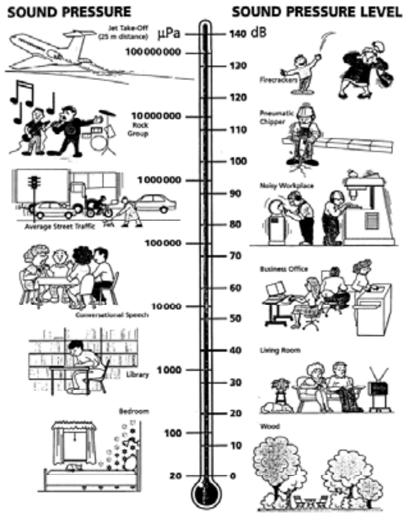
$$p = \text{rms of pressure wave} \propto \frac{1}{r}$$

Sound Pressure Level (SPL)

$$P_{dB} = 20\log_{10}\left[\frac{p}{p_0}\right]$$

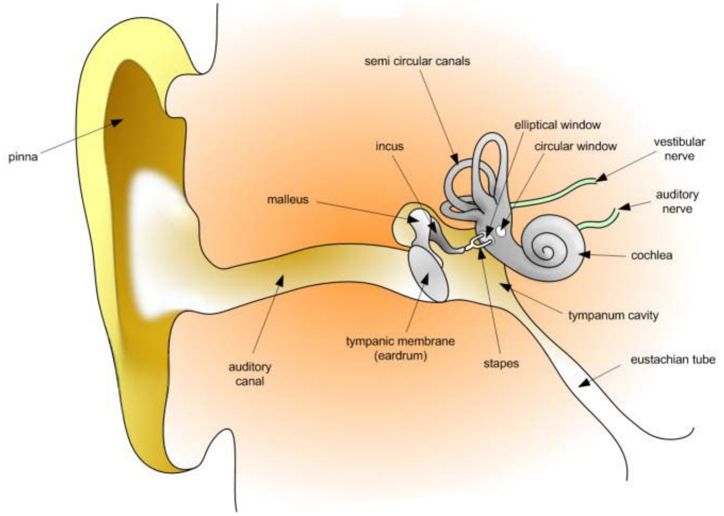
Standard Threshold of Hearing Pressure

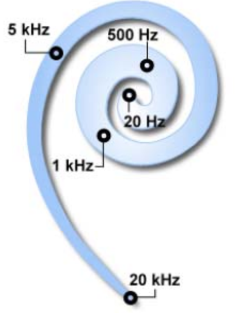
$$p_0 = 2\times10^{-5}\ N/m^2$$



Brüel and Kjær

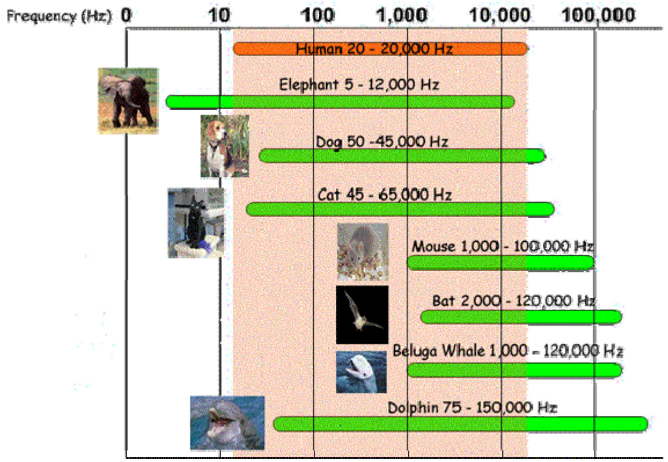
Human Hearing





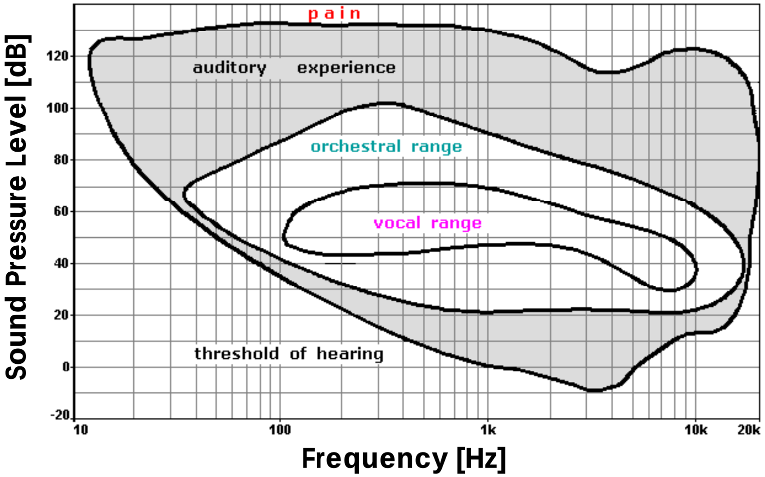
http://hendrix2.uoregon.edu/~dlivelyb/phys152/images/HumanEar.jpg
http://cnx.org/content/m43048/latest/Picture%202.png

Human Hearing



http://www.cyberphysics.co.uk/graphics/diagrams/hearing_range.gif

Human Hearing



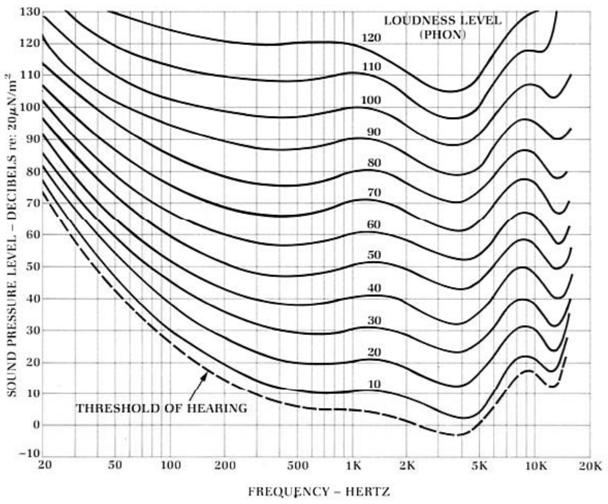
http://sound.westhost.com/articles/fadb.htm

Sound Loudness

- Subjective Term Based on Human Perception
- Same Intensity

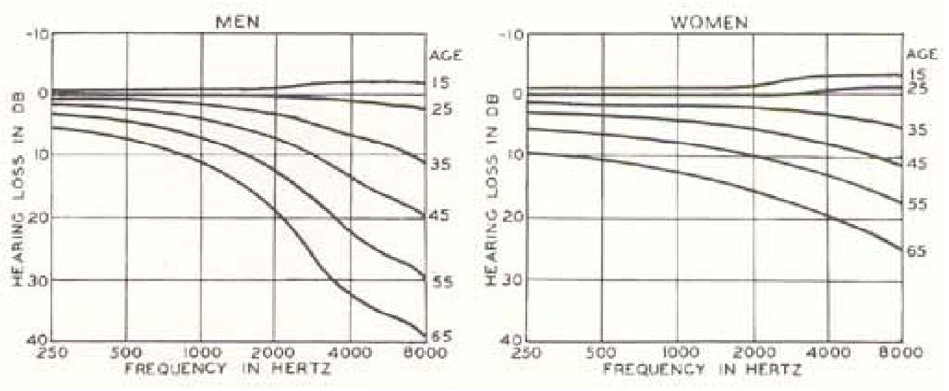
  ≠

  Same Loudness



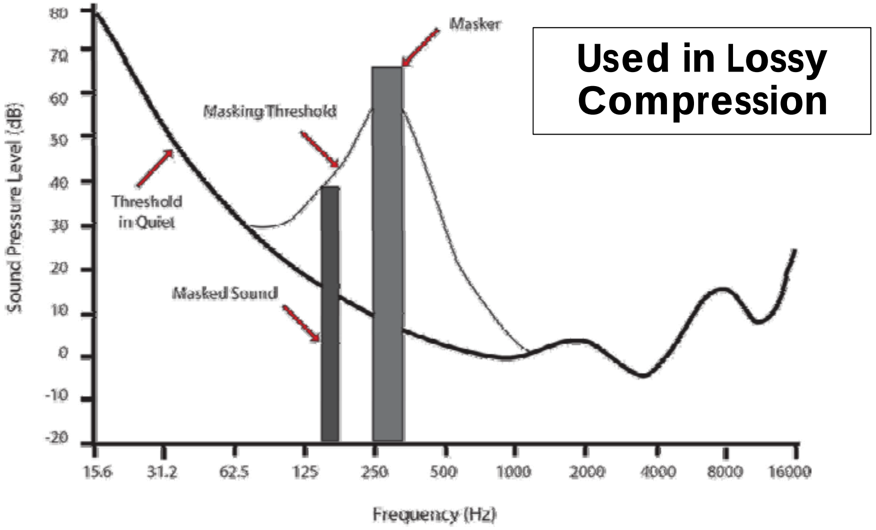
http://jac.michaeldrolet.net/physics/fig07.jpg

Hearing Age



http://www.roger-russell.com/hearing/hearing4.jpg

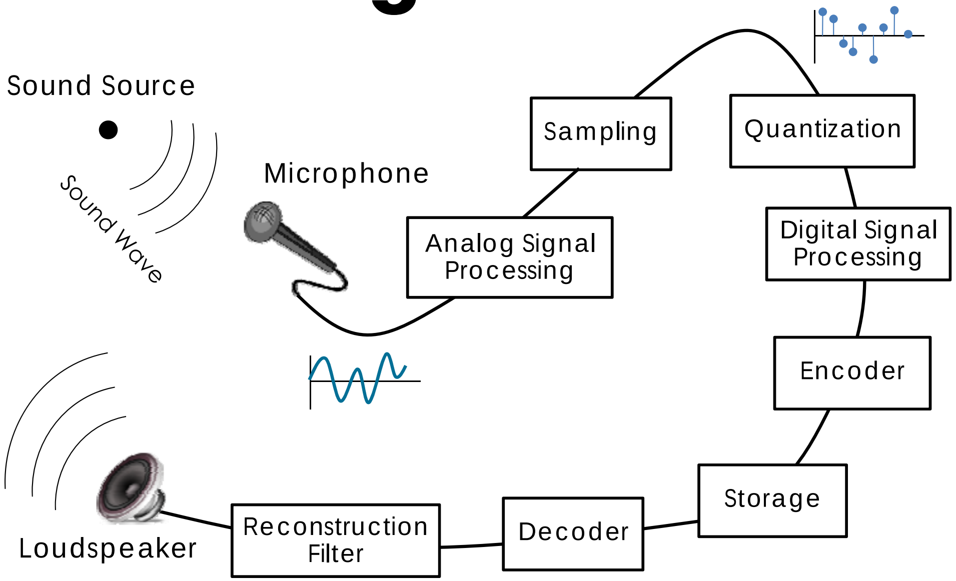
Auditory Masking



Used in Lossy Compression

http://en.wikipedia.org/wiki/File:Audio_Mask_Graph.png

Audio Signal I/O



Sound Source

Sound Wave

Microphone

Analog Signal Processing

Sampling

Quantization

Digital Signal Processing

Encoder

Storage

Decoder

Reconstruction Filter

Loudspeaker

Microphone

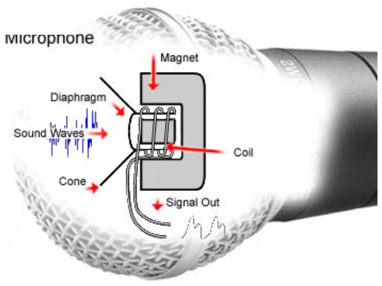
Transducer : Sound ⇒ Electrical signal.

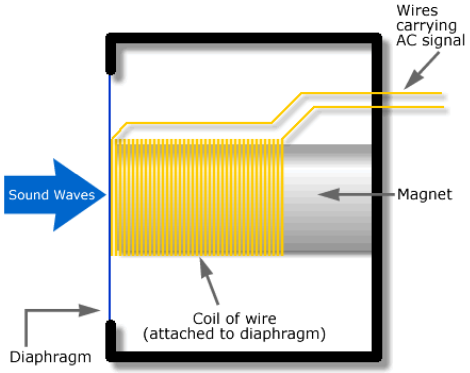
- Condenser microphone
- Dynamic microphone
- Piezoelectric microphone
- Fiber optic microphone
- Laser microphone
- MEMS microphone

(MicroElectrical-Mechanical System)

Dynamic Microphone

- Electromagnetic Induction
- Inexpensive
- Robust & Resistant to Moisture
- High Gain
- Attenuate in High Frequencies

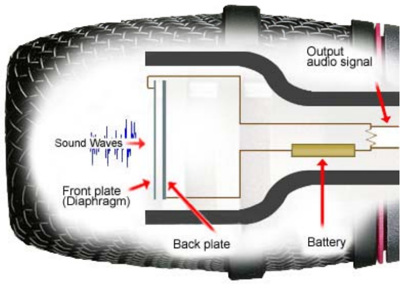


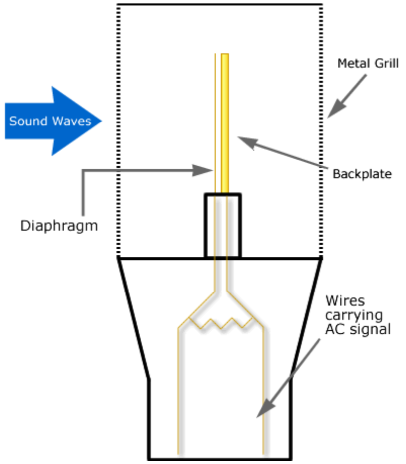


https://microphones.audiolinks.com/Microphones/micdiagram2922.jpg
http://www.burninggrooves.com/articles/introduction-to-dynamic-and-condenser-microphones

Condenser Microphone

- Capacitive Microphone
- Wider & Flatter Frequency Response
- More Expensive
- Fragile
- Need Power

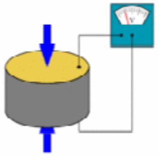


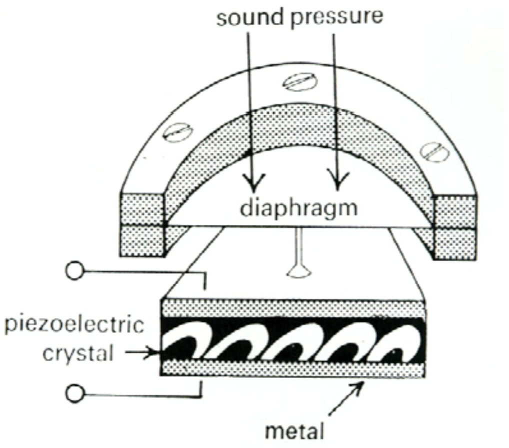


http://www.burninggrooves.com/articles/introduction-to-dynamic-and-condenser-microphones
https://microphones.audiolinks.com/Microphones/micdiagram02569.jpg

Piezoelectric Microphone

- Piezoelectricity : Pressure ⇒ Voltage
- Contact Microphone
- Durable
- High Pressure/Temp Environments

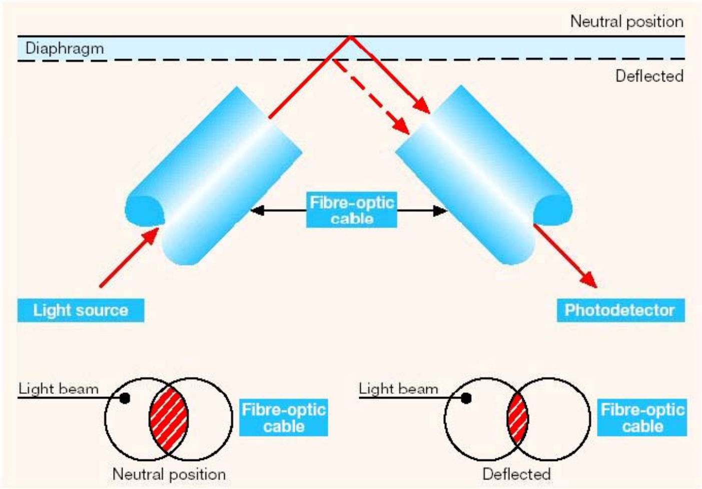




)8/06/piezoelectric-microphones-fourth-in.html
nttp://www.scienceinthenews.org.uk/images/contents/29/en/181.jpg
http://www.d.umn.edu/~mharvey/th1551mpacmikes.html

Fiber Optic Microphone

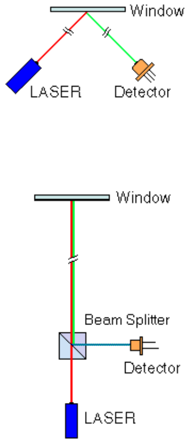
- Light Intensity
- High Dynamic and Frequency Range
- Robust, resistant to changes in heat and moisture, do not influence by any electrical, magnetic or radioactive fields

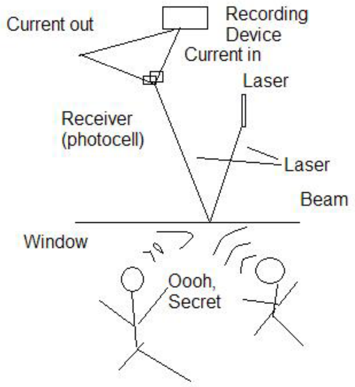


http://www.prosoundweb.com/images/uploads/SennDiagram.jpg

Laser Microphone

- Light Intensity
- Pick up Sound at a Distance

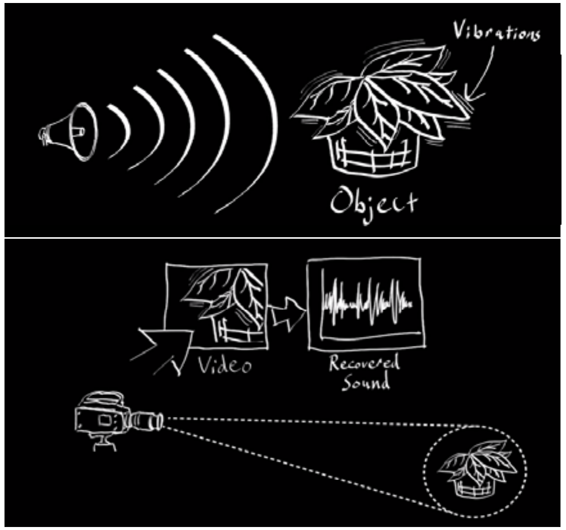




http://williamson-labs.com/laser-mic.htm
http://hackedgadgets.com/2007/08/16/the-laser-listener/

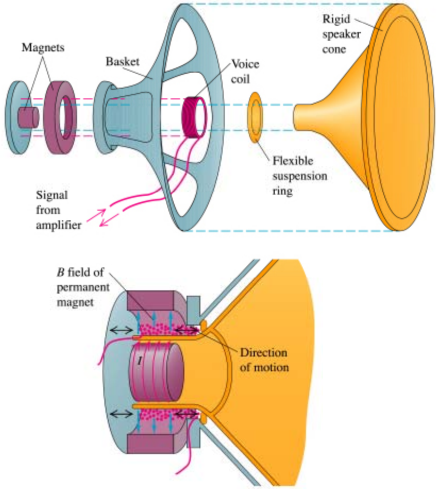
Visual Microphone

- Using High Speed Video
- Image Processing
- Passive Recovery of Sound from Video

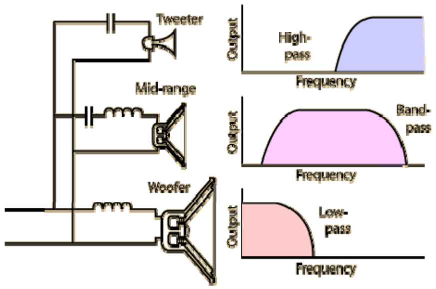


Loudspeaker

Transducer : Electrical signal ⇒ Sound



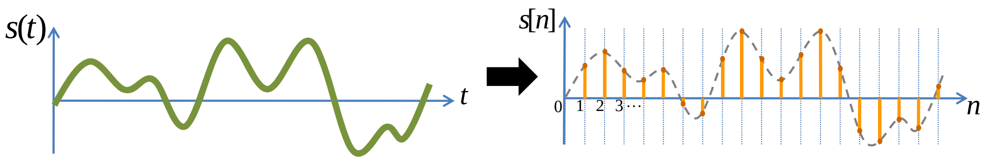
Copyright © Addison Wesley Longman, Inc.



Audio Crossover

http://wiki.vpa.mtu.edu/wiki/index.php/Speakers
http://hyperphysics.phy-astr.gsu.edu/hbase/audio/imgaud/cross6.gif

Sampling



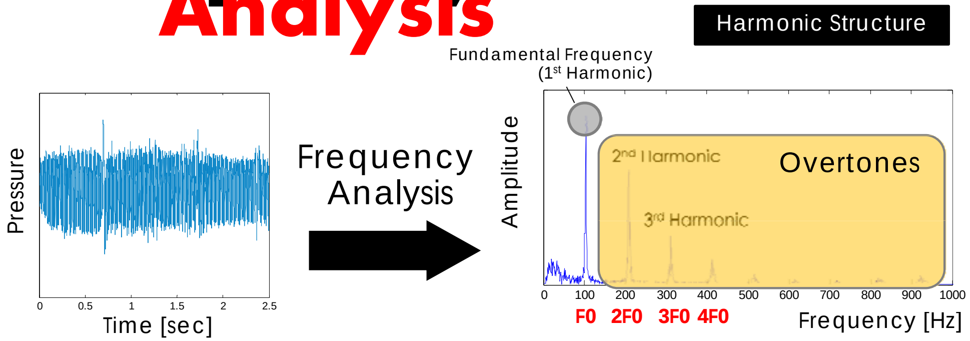
Analog Audio Signal

Discrete Audio Data

To avoid Aliasing
Sampling Frequency (F_S) > 2 × Signal Max. Frequency

human can hear ~ 20 Hz-20 kHz	➡	CD : 44.1 kHz Digital Audio Tape : 48 kHz MP3 : 32 kHz, 44.1 kHz, 48 kHz, etc.
human speech ~ 5 Hz-4 kHz	➡	Telephone-quality Audios : 8 kHz

Frequency Analysis

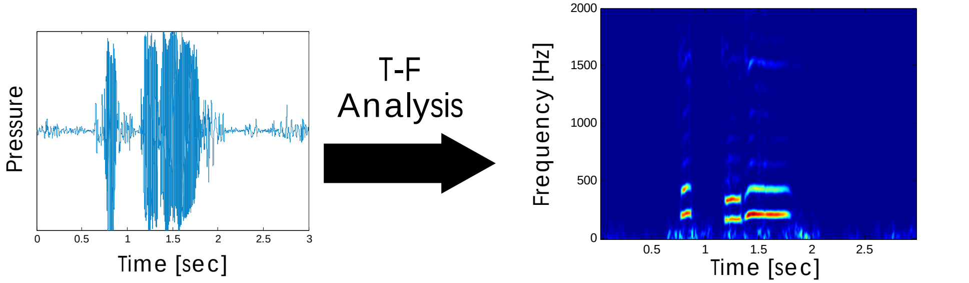


Audio Signal

Spectrum
(Frequency Components)

- Continuous System ⇒ Fourier Transform
- Digital System ⇒ Discrete Fourier Transform [DFT]
- Fast Fourier Transform [FFT] = Fast Algo. for DFT

Time-Frequency Analysis

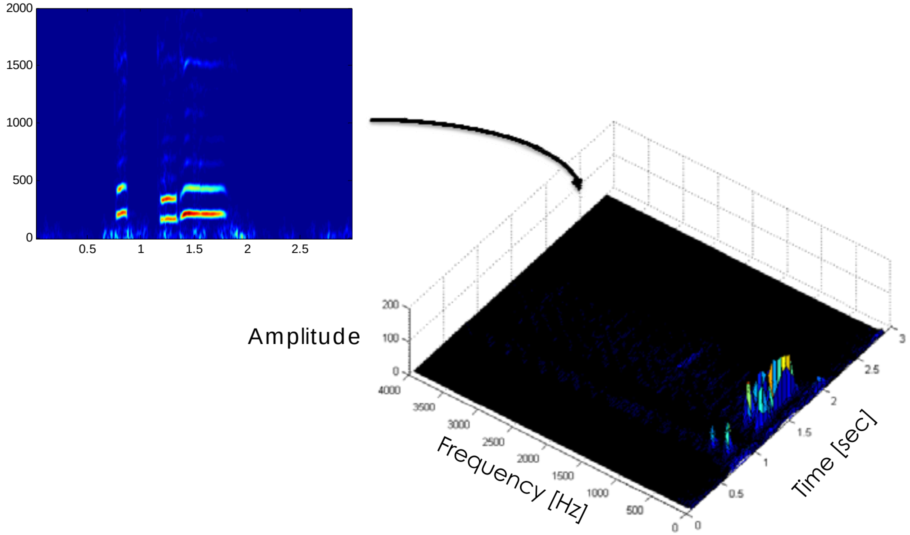


Audio Signal

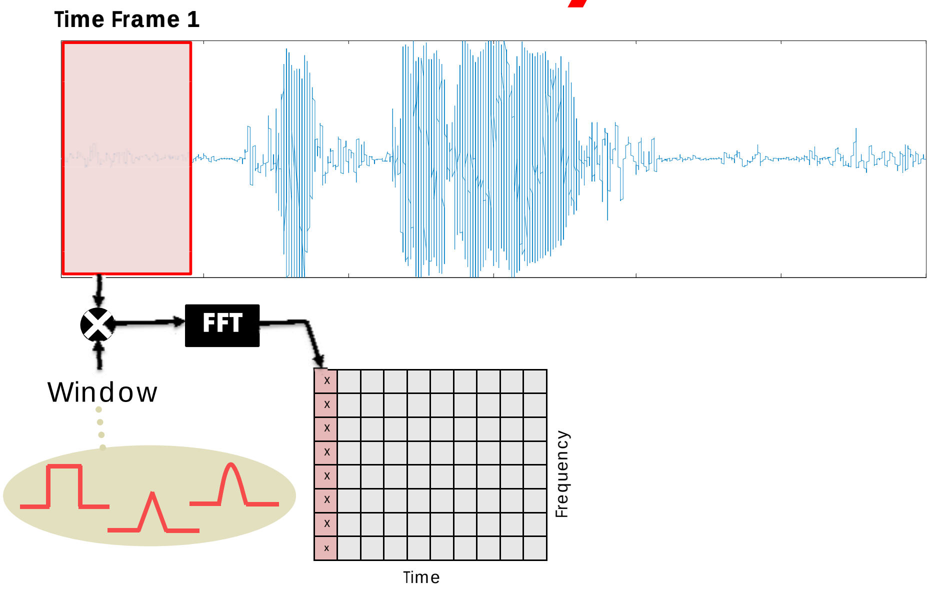
Spectrogram
(Frequency Components over Time)

- Short-Time Fourier Transform

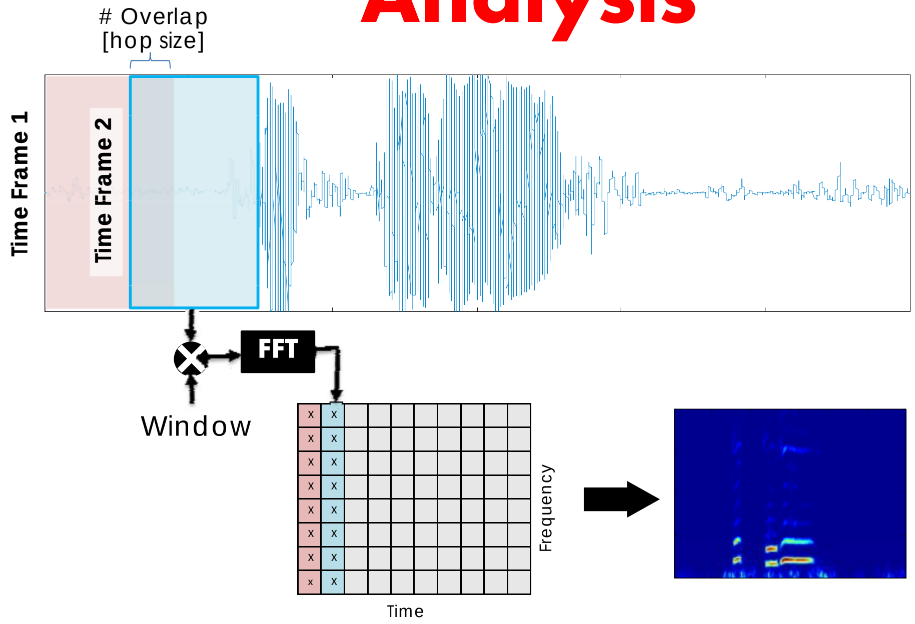
Time-Frequency Analysis



Time-Frequency Analysis

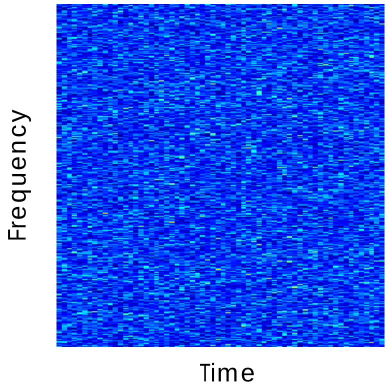


Time-Frequency Analysis

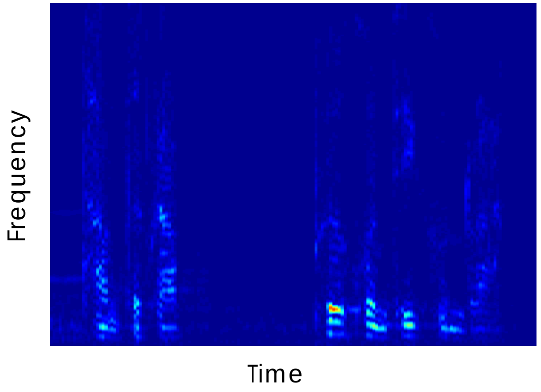


Spectrogram

White Noise

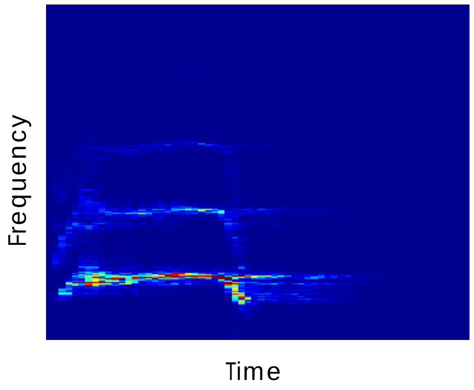


Speech

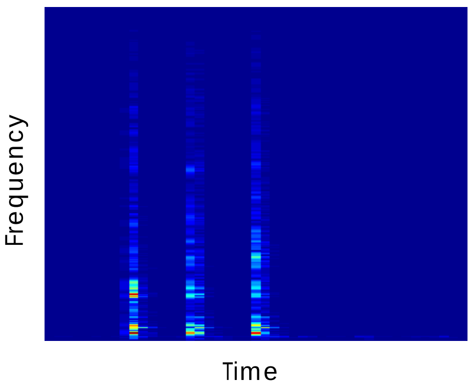


Spectrogram

Pitched Sounds

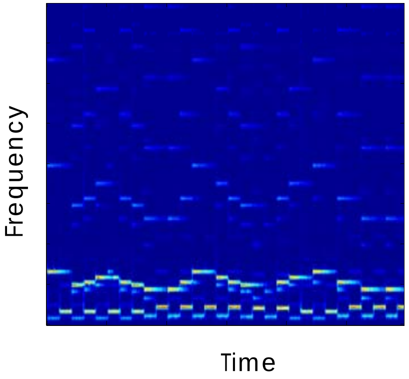


Unpitched Sounds

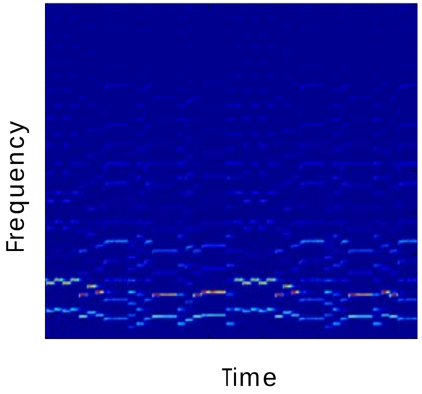


Spectrogram

8-bit Digital Sound

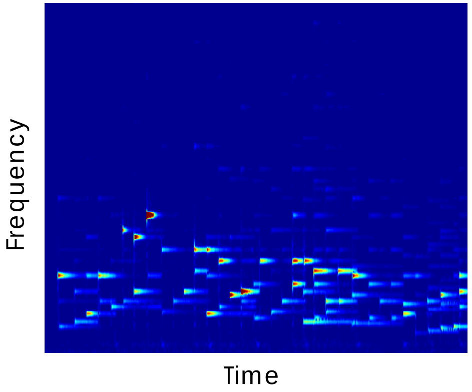


Saxophone

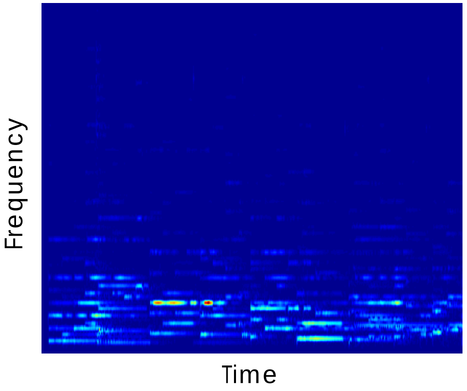


Spectrogram

Piano

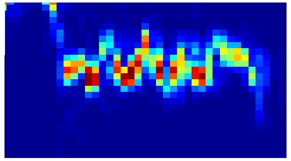


Guitar

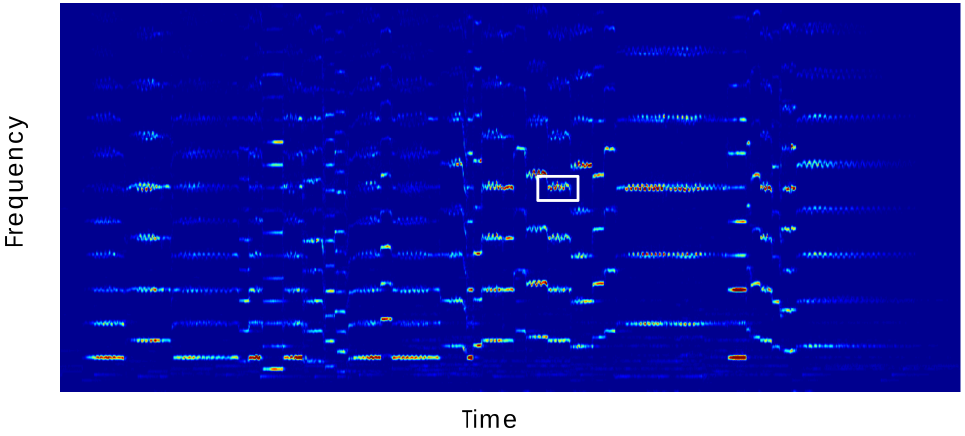


Spectrogram

High Vibrato & Tremolo

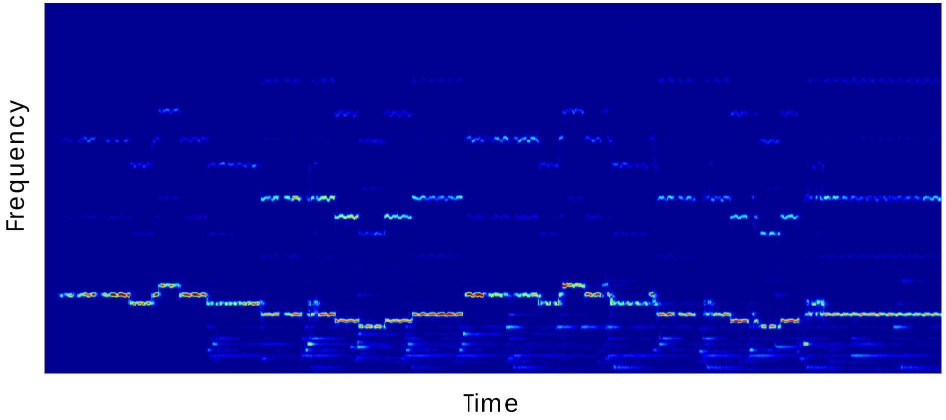


Violin



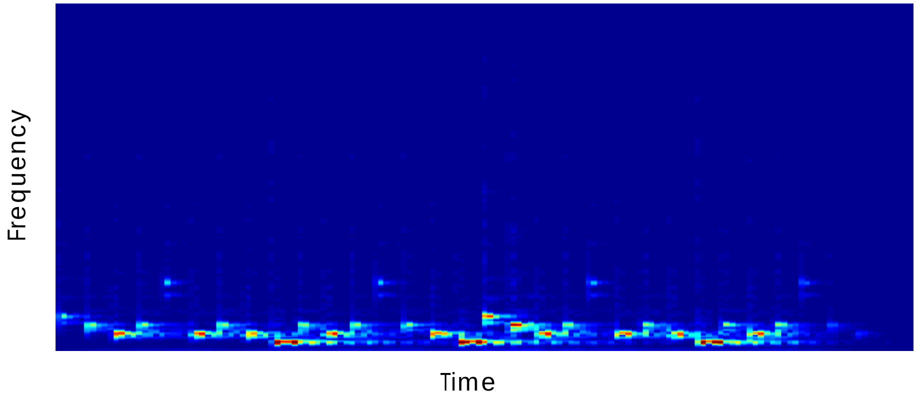
Spectrogram

Thai Flute



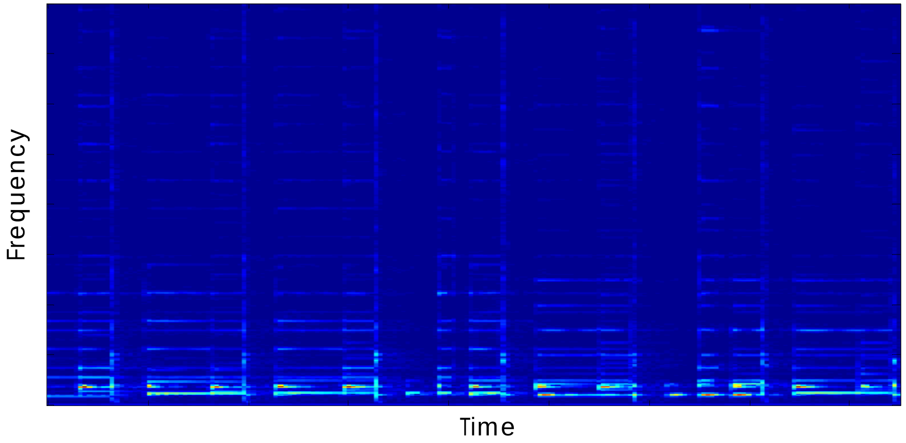
Spectrogram

Hand Drum



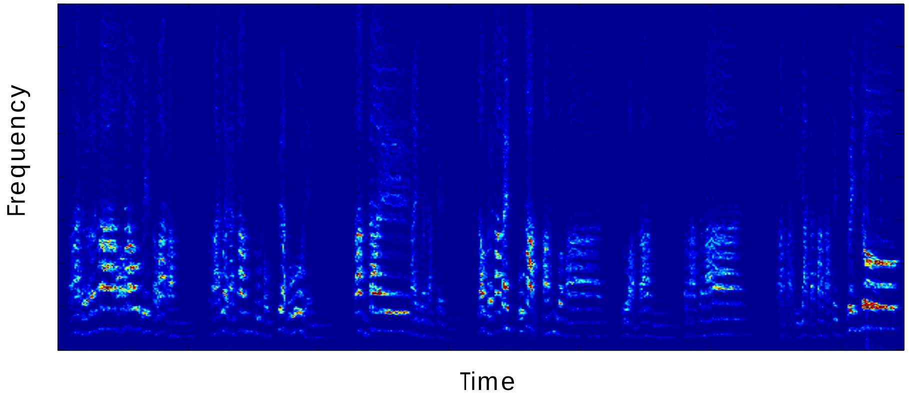
Spectrogram

Guitar + Tapping



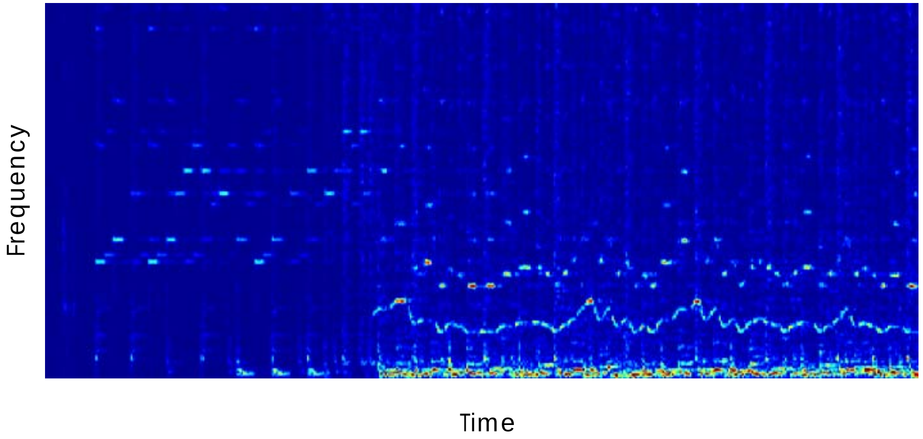
Spectrogram

Sing-along



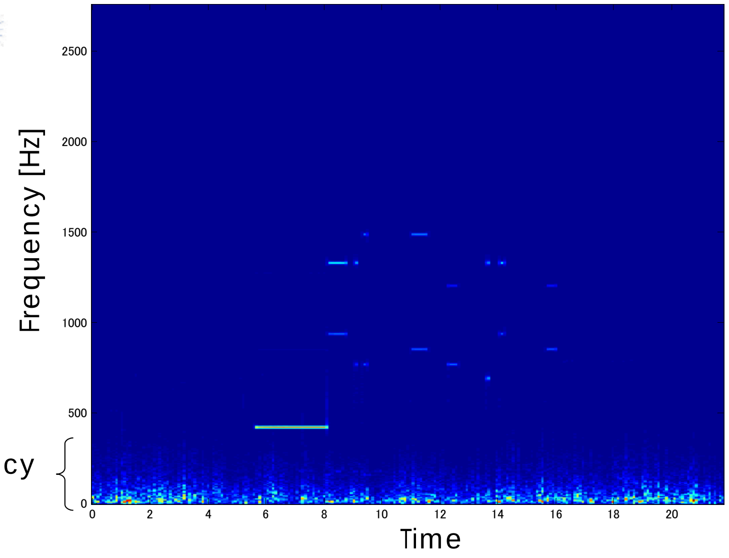
Spectrogram

Polyphonic



Frequency Selective Filter

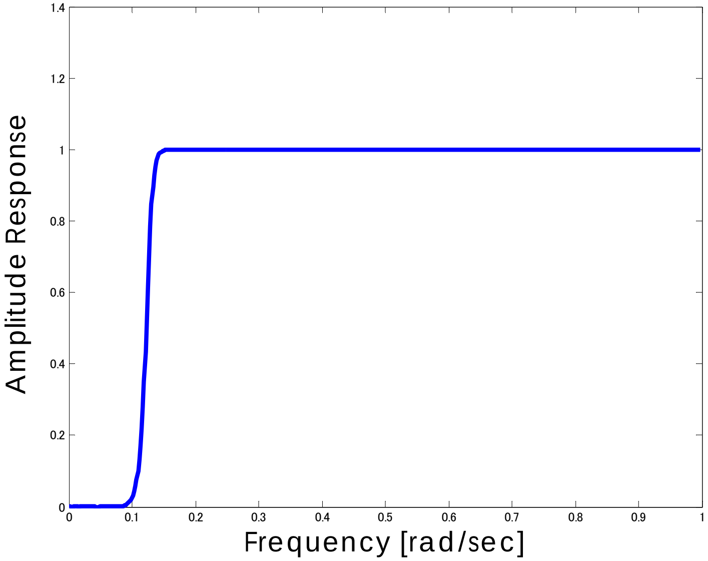
Noisy Signal



Low Frequency Noise

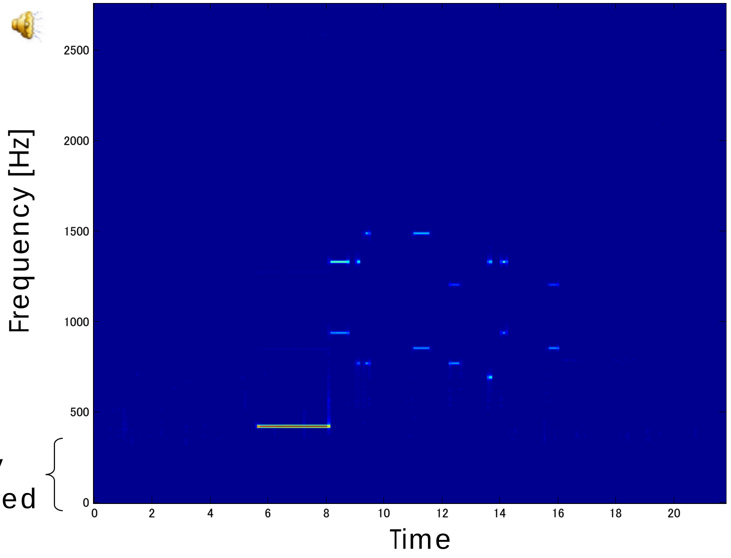
Frequency Selective Filter

High-pass Filter



Frequency Selective Filter

Restored Signal

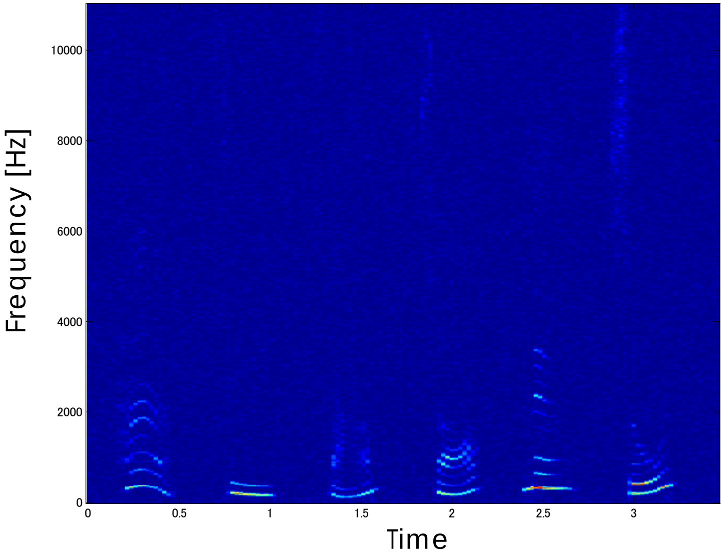


Low frequency noise was removed

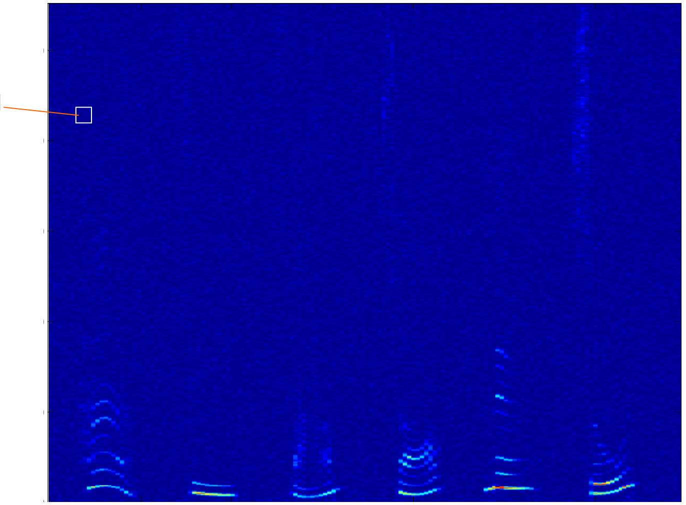
Thresholding

Noisy Signal

Additive White Gaussian Noise



Thresholding

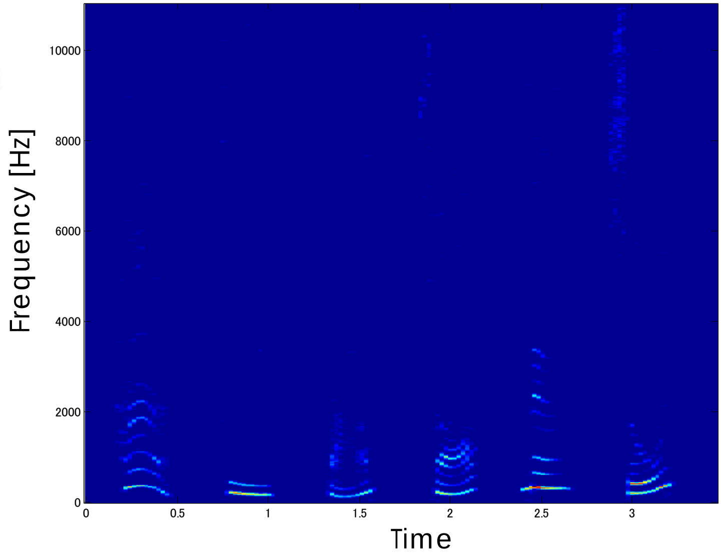


Find the power of signal $\sum |s|^2$ in each block

If the power is less than threshold, the block is truncated to zero.

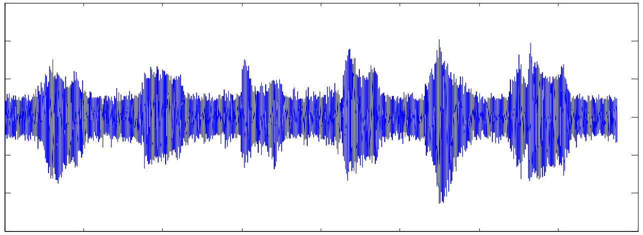
Thresholding

Restored Signal

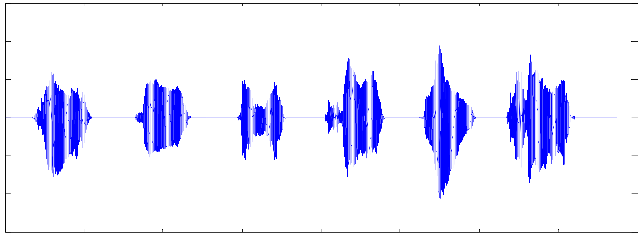


Thresholding

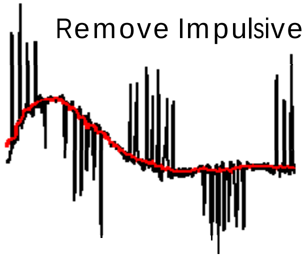
Noisy Signal



Restored Signal



Median Filter



Remove Impulsive Noise

- Computes the **median** of an array over sliding window with a given size.
- Can be used for removing an impulsive noises while preserving edges of the signal.

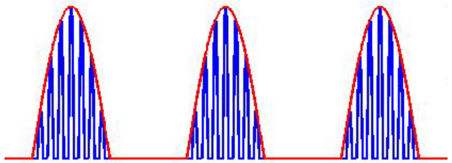
1	2	2	3	5	4	4	9	3	1	1	9	8	7	8	7

Length of window = 5

Median

2	2	2	3	4	4	4	?	3	3	3	7	8	8	7.5	7

Estimate Signal Envelop



Maximum Filter

[Dilation Filter]

- Computes the **maxima** of an array over sliding window with a given size.
- Can be used for finding the envelop.

1	2	3	5	8	9	8	5	6	4	2	5	7	4	2	1

Length of window = 5

Max

3	5	8	9	9	9	9	9	8	?	7	7	7	7	7	4