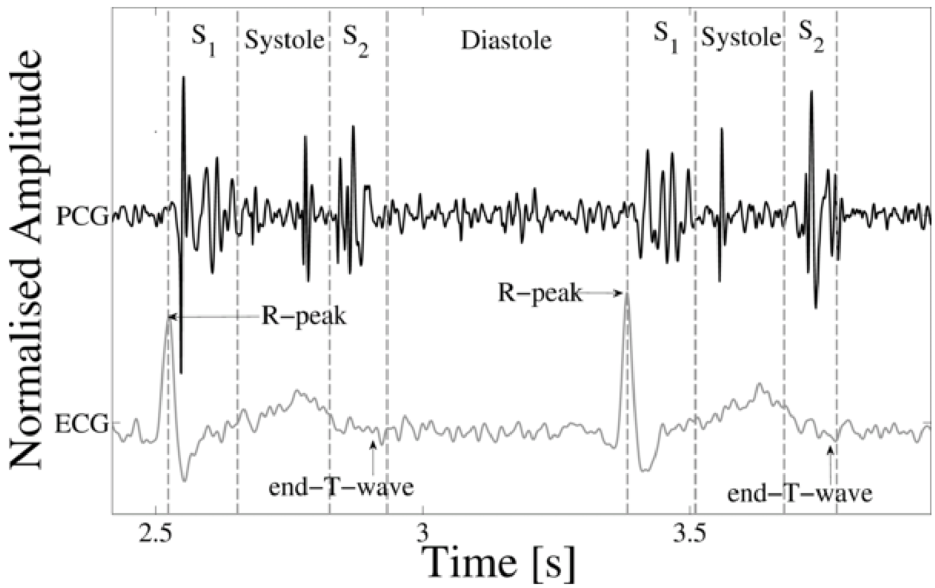
PCG Signal Processing

First Week:

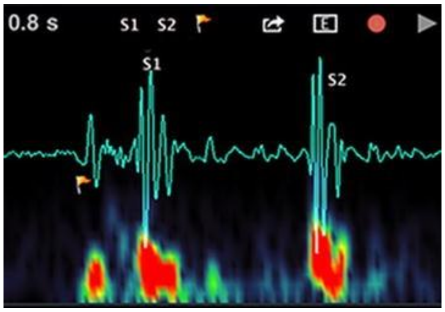
The phonocardiogram (PCG) is the graphical representation of a heart sound recording. The figure illustrates a short section of a PCG recording.



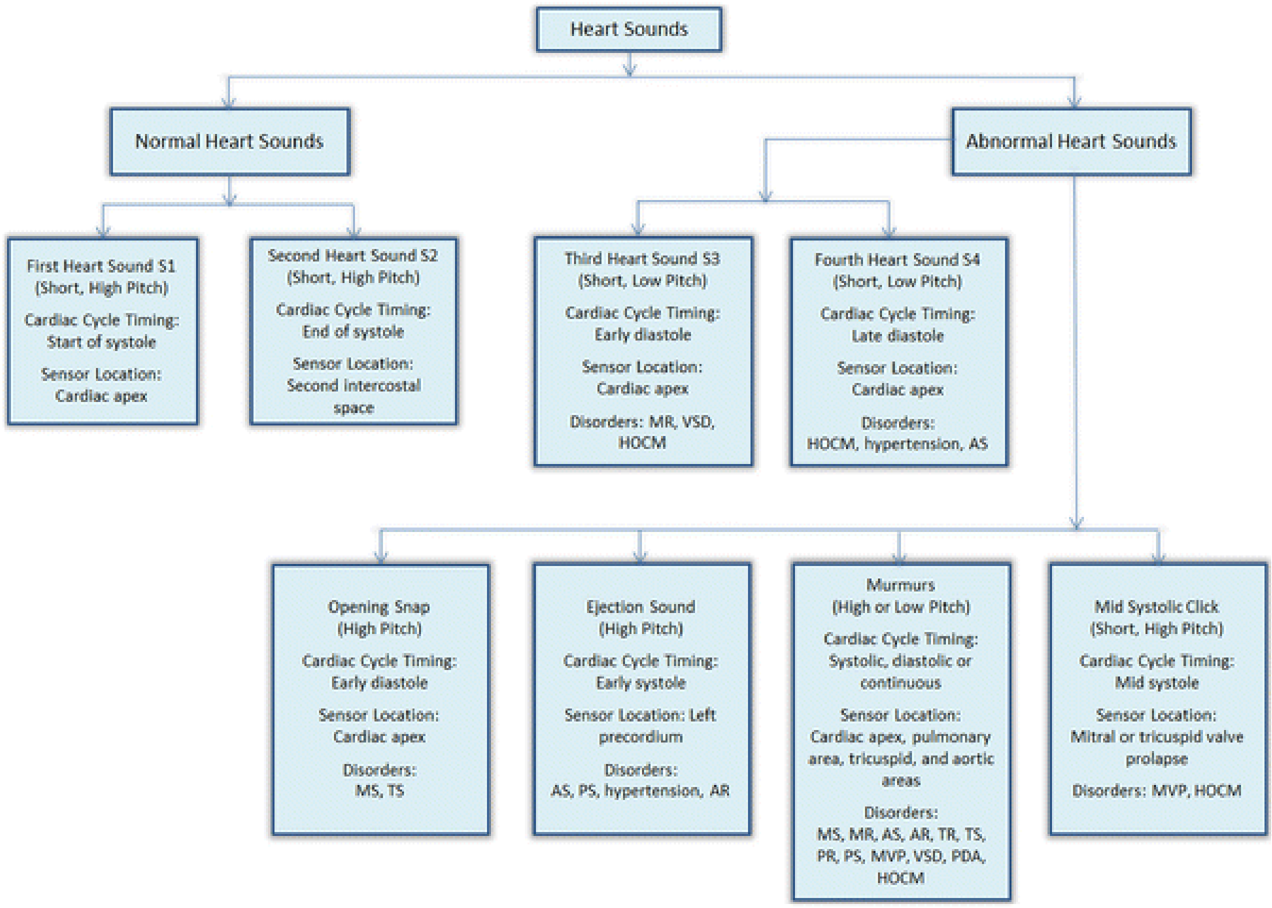
Fundamental heart sounds (FHSs) usually include the first (S1) and second (S2) heart sounds. **S1** occurs at the beginning of **isovolumetric ventricular contraction**, when the **mitral and tricuspid, valves close due to the rapid increase in pressure within the ventricles**. **S2** occurs at the beginning of **diastole** with the **closure of the aortic and pulmonic valves**. While the FHSs are the most recognizable sounds of the heart cycle, **the mechanical activity** of the heart may also cause other audible sounds, such as the third heart sound (**S3**), the fourth heart sound (**S4**), systolic ejection click (EC), mid-systolic click (MC), diastolic sound or opening snap (OS), as well as heart murmurs caused by the turbulent, high-velocity flow of blood.

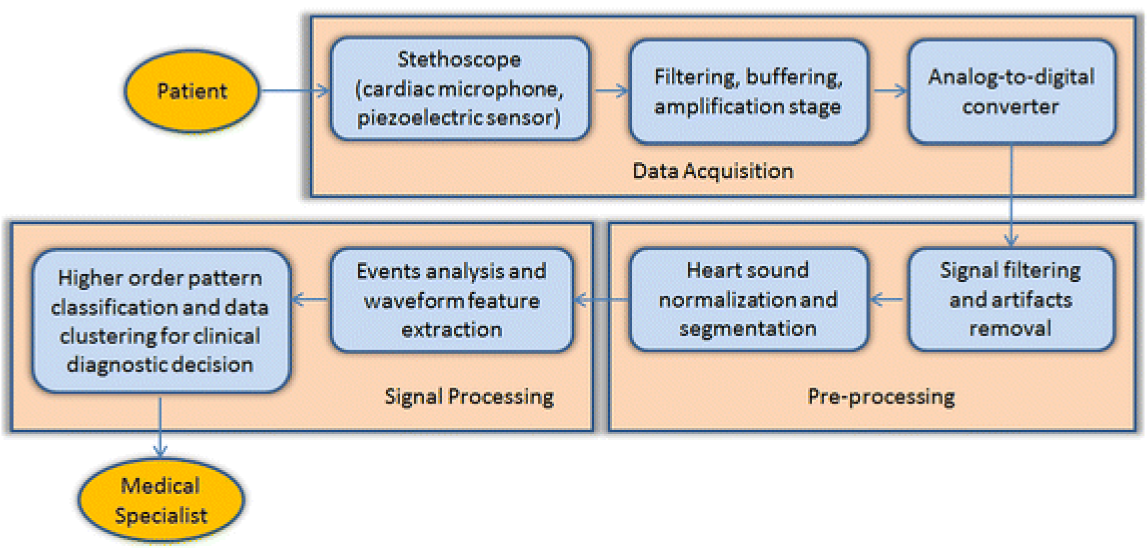
Characteristics of heart sounds:

HSs are generated by the beating heart and the resultant flow of blood through it. In healthy adults, there are two normal HSs (as illustrated in the Figure): the first HS (**S1**), produced by the **closing of the atrioventricular valves**; and the second HS (**S2**), caused by the **closure of the semilunar valves**. In the case of abnormal HS, there could be other several signal activities between S1 and S2 such as S3, S4, murmur, etc. The third HS (S3) is a rare extra sound caused by a sudden deceleration of blood flow into the left ventricle from the left atrium. This sound is normal in children and adults up to age 35–40 years. After the age of 40, a third HS is usually abnormal and correlates with dysfunction or volume overload of the ventricles [12]. The fourth HS (S4) is caused by the vibration of valves, supporting structures and the ventricular walls. S4 is proved to be a sign or symptom of heart failure during diastolic period. In general, the frequency of S1 is lower than that of S2, and the duration of S1 is longer than that of S2. The S3 occurs from 0.1 to 0.2s after S2, while S4 occurs from 0.07 to 0.1 s before S1—both of them are low pitched. In addition to these HSs, numerous heart murmurs may arise mainly from heart problems or diseases. The murmurs are extra or unusual sound heard during a heartbeat and broadly classified as systolic, diastolic and continuous.



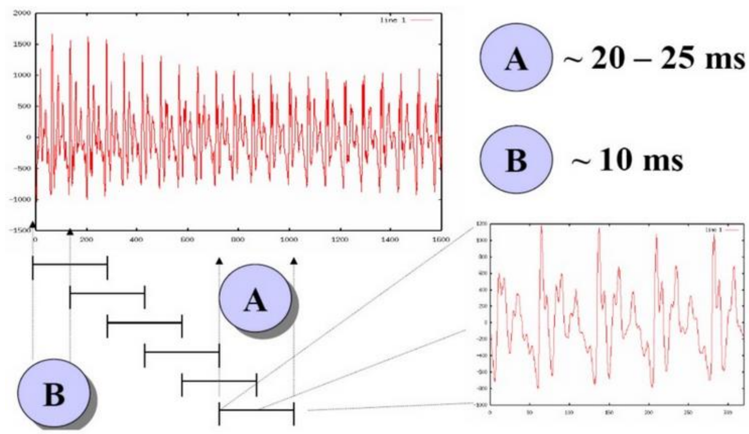
Two primary heart sounds: S1 and S2


Heart Sounds
Normal Heart Sounds
Abnormal Heart Sounds
First Heart Sound S1 (Short, High Pitch)
Cardiac Cycle Timing: Start of systole
Sensor Location: Cardiac apex
Second Heart Sound S2 (Short, High Pitch)
Cardiac Cycle Timing: End of systole
Sensor Location: Second intercostal space
Third Heart Sound S3 (Short, Low Pitch)
Cardiac Cycle Timing: Early diastole
Sensor Location: Cardiac apex
Disorders: MR, VSD, HOCM
Fourth Heart Sound S4 (Short, Low Pitch)
Cardiac Cycle Timing: Late diastole
Sensor Location: Cardiac apex
Disorders: HOCM, hypertension, AS
Opening Snap (High Pitch)
Cardiac Cycle Timing: Early diastole
Sensor Location: Cardiac apex
Disorders: MS, TS
Ejection Sound (High Pitch)
Cardiac Cycle Timing: Early systole
Sensor Location: Left precordium
Disorders: AS, PS, hypertension, AR
Murmurs (High or Low Pitch)
Cardiac Cycle Timing: Systolic, diastolic or continuous
Sensor Location: Cardiac apex, pulmonary area, tricuspid, and aortic areas
Disorders: MS, MR, AS, AR, TR, TS, PR, PS, MVP, VSD, PDA, HOCM
Mid Systolic Click (Short, High Pitch)
Cardiac Cycle Timing: Mid systole
Sensor Location: Mitral or tricuspid valve prolapse
Disorders: MVP, HOCM


Patient
Stethoscope (cardiac microphone, piezoelectric sensor)
Filtering, buffering, amplification stage
Analog-to-digital converter
Data Acquisition
Signal filtering and artifacts removal
Heart sound normalization and segmentation
Pre-processing
Events analysis and waveform feature extraction
Higher order pattern classification and data clustering for clinical diagnostic decision
Signal Processing
Medical Specialist

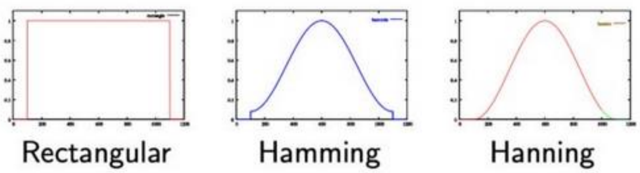
Heart sound segmentation:

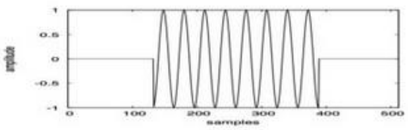
Signal segmentation is usually carried out after denoising stage. The simplest way of **segmentation** is through the use of **sliding window**. This technique simply partitions the whole HS data into a number of segments with the same duration, without considering the positions of starting point and ending point for each segment. The more commonly adopted segmentation technique in HS analysis is performed on the basis of the cardiac cycle, since in most cases the activities in the HS signal relating to given disease are contained in a single interval of cardiac cycle. Specifically, the **segmentation algorithm** divides the HS data into portions, each of which consists of four parts: **S1**, the **systole**, **S2** and the **diastole**.



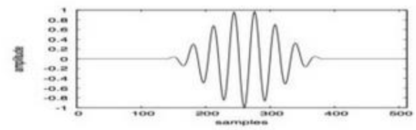
$$x[n] = w[n]\, s[n]$$

sliced frame original audio clip

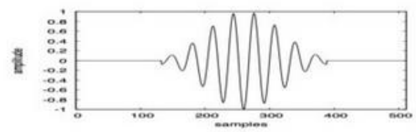




(a) Rectangular window

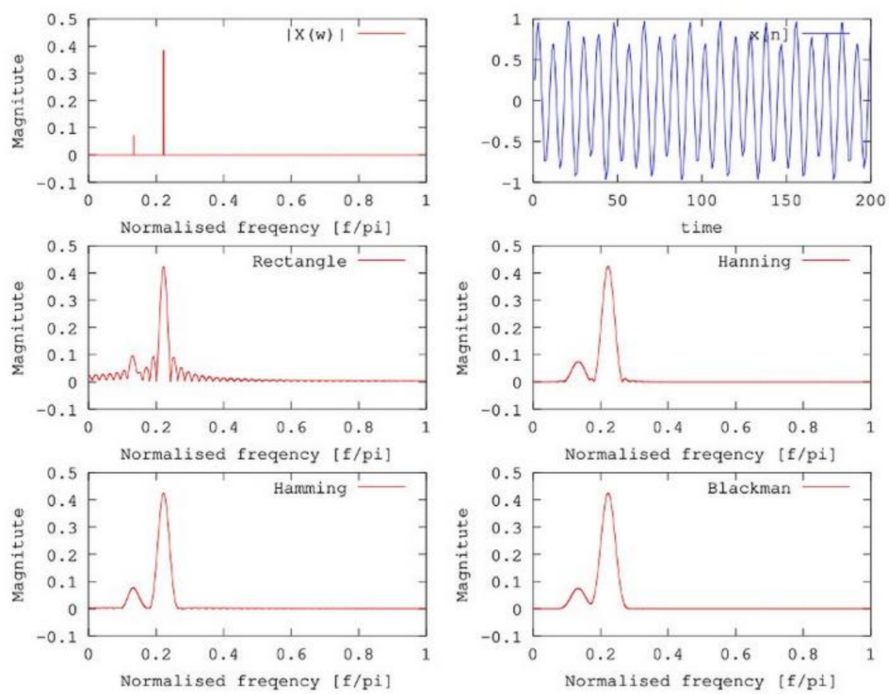


(b) Hanning window

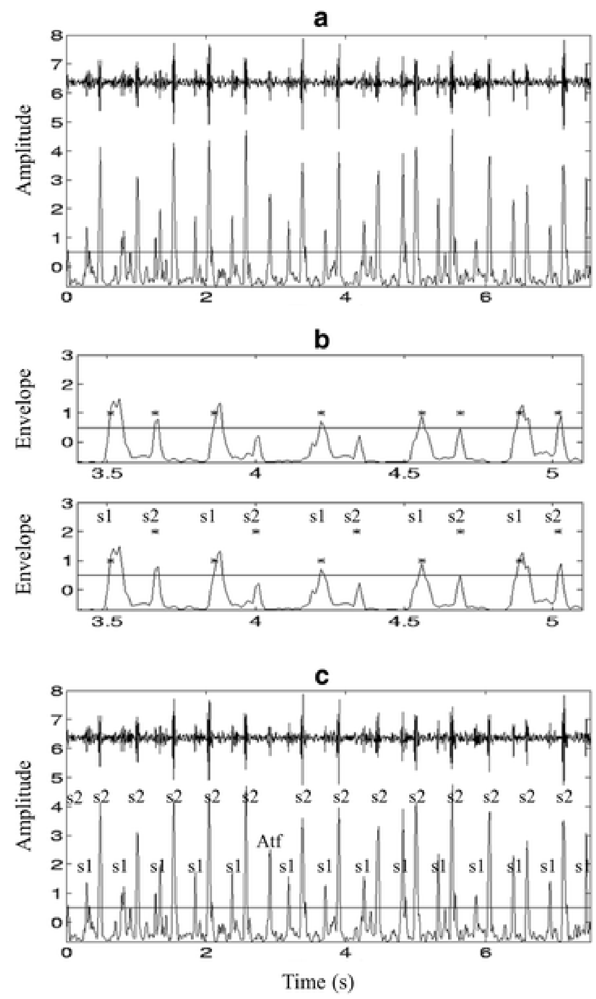


(c) Hamming window

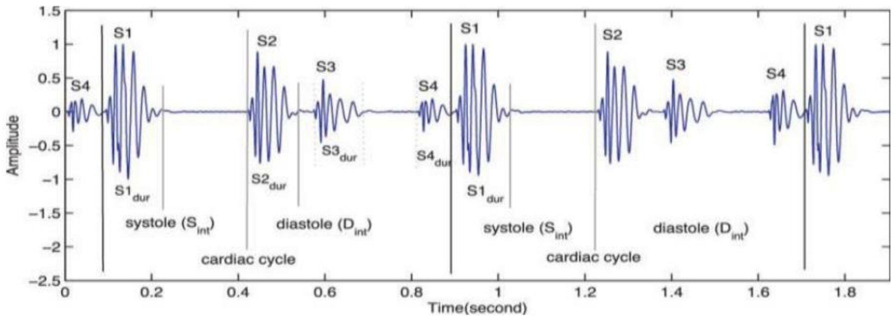
As shown, for Hamming and Hanning window, the amplitude drops off near the edge. (The Hamming window has a slight sudden drop at the edge while the Hanning window does not.)



The majority of the attempts in segmenting the HSs use the ECG signal and/or carotid pulse as the reference signal. This method, often called indirect segmentation, demarcates the HS boundaries using the QRS complex and T-waves of ECG signal. The disadvantage of indirect segmentation is that, apart from the employment of ECG electrodes, the complete segmentation may become difficult because the T-wave is too weak to be identified in some patients. Another disadvantage of using ECG as a reference is that the timing between electrical and mechanical activities in a cardiac cycle is not constant for all patients, because of a variety of pathological conditions.



Envelogram based heart sound segmentation an original signal and its average Shannon energy, b recovering the “lost” peaks, and c identified S1 and S2 and artifacts.



Methods used for Processing:

<u>Mel-frequency cepstral coefficients (MFCC):</u> The **MFCC feature extraction** technique basically includes windowing the signal, applying the DFT, taking the log of the magnitude, and then warping the frequencies on a Mel scale, followed by applying the inverse DCT.

Classifying heart sounds using tf (tensor flow):

Feature Extraction → mfcc

Tensor Flow $\begin{cases} setting\ parameters \\ training\ process \rightarrow gradient\ descent\ optimizer \end{cases}$

Classifying heart sounds using sklearn:

Feature Extraction → mfcc

Training the data → SVM method

MATLAB Code Implementation:

- ✓ Setting the parameters:

  The sampling frequency at which to extract signal features: 1000

  The down-sampled frequency: 50

  Tolerance for S1 and S2 localization: 0.1 seconds

- ✓ the heart rate and the systolic time interval from a PCG recording
  Heat Rate Calculation:
  From Schmidt:

The duration of the heart cycle is estimated as the time from lag zero to the highest peaks between 500 and 2000 ms in the resulting autocorrelation. This is performed after filtering and spike removal.

- Filtering the primary signal with a Low pass Butterworth filter.
- Filtering the signal with a high pass Butterworth filter.
- Spike removal from the original signal
- Finding the homomorphic envelope with Hilbert
- Finding the autocorrelation
- Finding the systolic time interval:
  From Schmidt: The systolic duration is defined as the time from lag zero to the highest peak in the interval betwFind the autocorrelation in 200 ms and half of the heart cycle duration.

✓ Setting the Min and Max values for the S1 and S2 sounds using the mean and standard deviation of systole and diastole times

✓ Finding the spectrogram of the signal

✓ Extracting PCG features: an array of derived features

- 25-400Hz 4th order Butterworth band pass (a butterworth low pass filter with the frequency of 400Hz and a butterworth high pass filter with the frequency of 25Hz).
- Spike removal from the original signal
- Find the homomorphic envelope with Hilbert
- Downsample the envelope
- normalise the envelope
- Power spectral density feature (Finding the mean PSD over the frequency range of interest. PSD is the array of maximum PSD values between the max and min limits, resampled to the same size as the original data.)
- finding the discrete wavelet transform at level N for signal X

Second Week:

Respiratory Rate calculation:

How do you calculate respiratory rate from ECG?
One gets respiratory rate by **measuring the number of ECG samples in R-R interval** and its advantage lies in its simplicity. The other detects the rate by

measuring the size of R wave in QRS signal. This algorithm can detect the rate more robustly but it is complicated and requires the ECG signal base line to be stabilized. The preliminary study in laboratory environment showed that the precision of these algorithms was over 97%.

As a first step the respiratory signal is extracted from the ECG-signal sampled with the sampling frequency fs. For this purpose, a Hamilton-Tompkins algorithm is used to detect all QRS-complexes in the ECG signal (the below figure). Next, the amplitude of the R-to-S peak is computed, which provides a robust estimation of the signal amplitude with respect to the baseline. These extracted amplitudes are then interpolated by means of cubic splines to obtain the respiratory signal with the same sampling frequency as the ECG signal. An example for an extracted respiratory signal can be seen in Figure 1. Aside from a phase shift there is a high correlation between the extracted and the reference signal.

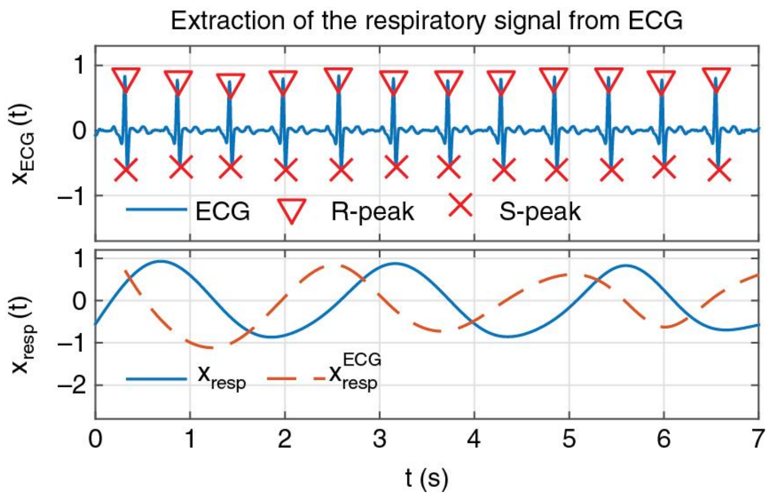


Figure 1

Top: baseline-corrected ECG signal with marked R- and S-peaks for the estimation of the R-to-S amplitude; bottom: reference respiratory signal (blue) compared to ECG-derived respiratory signal (orange). Amplitudes are normalized, arbitrary units.

Normal Heart Rate Interval: [60,100] beats per minute

Normal Respiration Rate Interval: [12,16] breaths per minute

Third Week

Processing Lung Sound Signals:

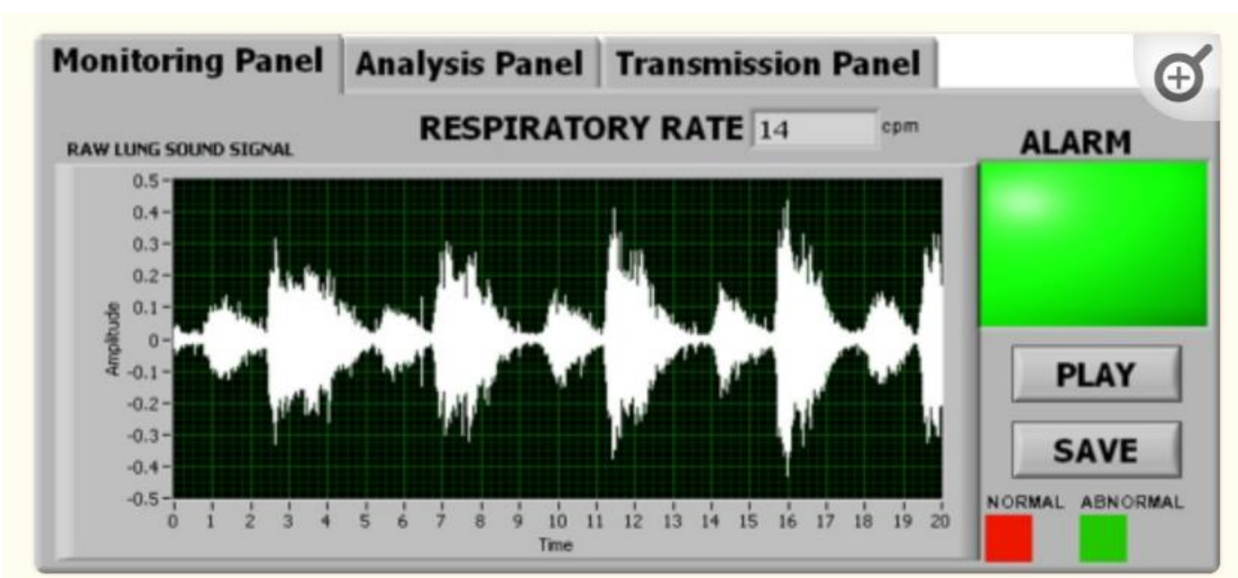
In adults, the normal respiratory rate is roughly **12 to 20 breaths per minute**.

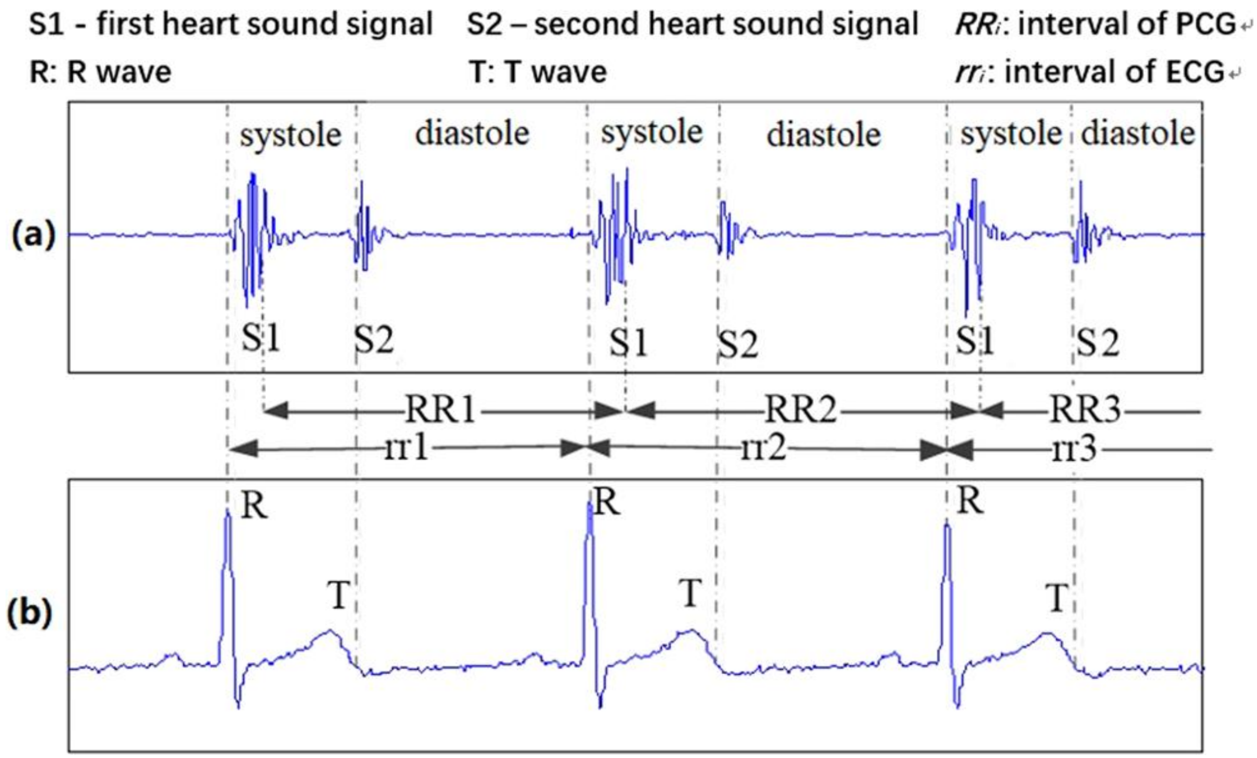
A rescue breath should last about 1 second.

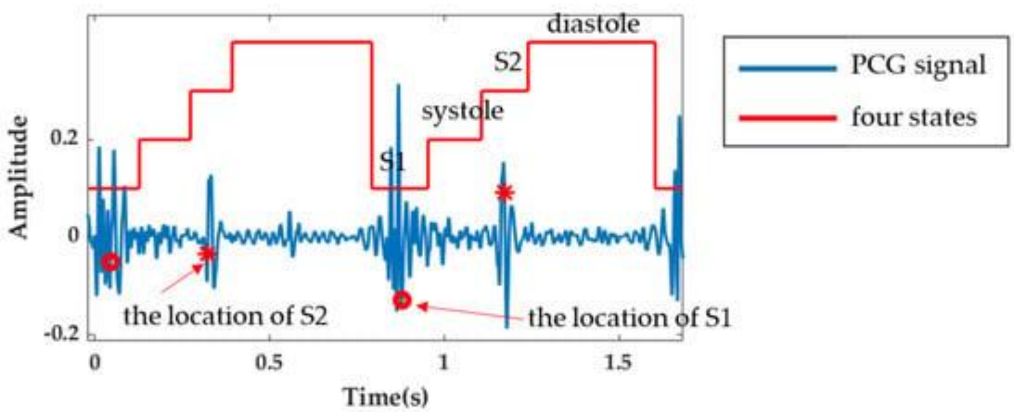
When we are at rest this is how normal breathing, usually appears: Breathing in (inhalation) for 1 to 1.5 seconds. Breathing out (exhalation) for 1.5 to 2 seconds. An automatic pause of almost no breathing for 1 to 2 seconds.

Preprocessing Step

At this step, the acquired raw lung sound data are down-sampled to 4000 Hz to set up a coherent feature set. Even on normal lung sound in the dataset which includes various noises such as intestinal and heart sounds and motion artefacts, 12th order Butterworth band pass filter with 120 and 1800 Hz cut-off frequencies is applied to minimize noise effects.







https://www.mdpi.com/2076-3417/13/2/1170

÷https://www.researchgate.net/publication/321135587_An_Automated_Lung_Sound_Preprocessing_and_Classification_System_Based_OnSpectral_Analysis_Methods

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3157943/

Week 4

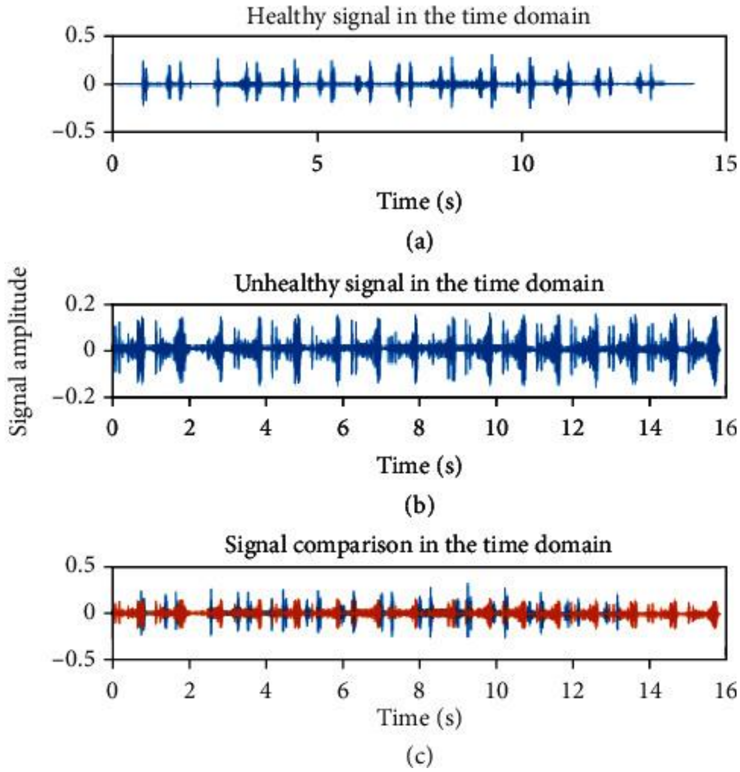
Different heart sounds:

https://depts.washington.edu/physdx/heart/demo.html

S1: corresponding to the beginning of ventricular systole is due to the closure of atrioventricular valves.

This sound is composed of two internal components: the mitral component (M1) associated with the closure of the mitral valve, and the tricuspid component (T1) associated with the closing of the tricuspid valve and **S2:** marking the end of ventricular systole and signifying the beginning of diastole, is made up of two main components: the aorticcomponent (A2) corresponding to the closure of the aortic valve, and the pulmonary component (P2), corresponding to the closure of the pulmonary valve.

S3 and **S4**, with lower amplitude than S1 or S2, appear occasionally in the cardiac cycle by the effect of disease or age.



Normal Category

In the Normal category there are normal, healthy heart sounds. These may contain noise in the final second of the recording as the device is removed from the body. They may contain a variety of background noises (from traffic to radios). They may also contain occasional random noise corresponding to breathing, or brushing the microphone against clothing or skin. A normal heart sound has a clear “lub dub, lub dub” pattern, with the time from “lub” to “dub” shorter than the time from “dub” to the next “lub” (when the heart rate is less than 140 beats per minute). Note the temporal description of “lub” and “dub” locations over time in the following illustration:

…lub……….dub……………. lub……….dub………. lub……….dub……. lub……….dub…

In medicine we call the lub sound "S1" and the dub sound "S2". Most normal heart rates at rest will be between about 60 and 100 beats (‘lub dub’s) per minute. However, note that since the data may have been collected from children or adults in calm or excited states, the heart rates in the data may vary from 40 to 140 beats or higher per minute. Dataset B also contains noisy_normal data - normal data which includes a substantial amount of background noise or distortion. You may choose to use this or ignore it, however the test set will include some equally noisy examples.

Murmur Category

Heart murmurs sound as though there is a “whooshing, roaring, rumbling, or turbulent fluid” noise in one of two temporal locations: (1) between “lub” and “dub”, or (2) between “dub” and “lub”. They can be a symptom of many heart disorders, some serious. There will still be a “lub” and a “dub”. One of the things that confuses non-medically trained people is that murmurs happen *between* lub and dub or *between* dub and lub; not *on* lub and not *on* dub. Below, you can find an asterisk* at the locations a murmur may be.

…lub..****...dub……………. lub..****..dub ……. lub..****..dub ……….…. lub..****..dub …

or

..lub…….dub…******….lub……. dub…******….lub ……. dub…******….lub ……….dub…

Dataset B also contains noisy_murmur data - murmur data which includes a substantial amount of background noise or distortion. You may choose to use this or ignore it, however the test set will include some equally noisy examples

Extra Heart Sound Category (Dataset A)

Extra heart sounds can be identified because there is an additional sound, e.g. a “lub-lub dub” or a “lub dub-dub”. An extra heart sound may not be a sign of disease. However, in some situations it is an important sign of disease, which if detected early could help a person. The extra heart sound is important to be able to detect as it cannot be detected by ultrasound very well. Below, note the temporal description of the extra heart sounds:

…lub.lub…….dub…..………. lub. lub……….dub…………….lub.lub……..…….dub…….

or

…lub……dub.dub………….lub.……….dub.dub………….lub……..…….dub. dub……

Artifact Category (Dataset A)

In the Artifact category there are a wide range of different sounds, including feedback squeals and echoes, speech, music and noise. There are usually no discernable heart sounds, and thus little or no temporal periodicity at frequencies below 195 Hz. This category is the most different from the others. It is important to be able to distinguish this category from the other three categories, so that someone gathering the data can be instructed to try again.

Extrasystole Category (Dataset B)

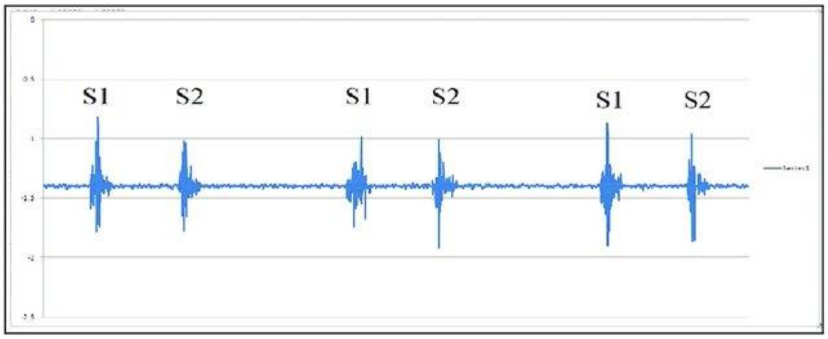
Extrasystole sounds may appear occasionally and can be identified because there is a heart sound that is out of rhythm involving extra or skipped heartbeats, e.g. a “lub-lub dub” or a “lub dub-dub”. (This is not the same as an extra heart sound as the event is not regularly occuring.) An extrasystole may not be a sign of disease. It can happen normally in an adult and can be very common in children. However, in some situations extrasystoles can be caused by heart diseases. If these diseases are detected earlier, then treatment is likely to be more effective. Below, note the temporal description of the extra heart sounds:

…........lub……….dub…………. lub...……….dub…………….lub.lub……..…….dub…….
or
…lub……….dub......…………….lub.…………dub.dub……….lub……..….dub.……

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9348924/

https://www.kaggle.com/code/mychen76/heart-sounds-analysis-and-classification-with-lstm

Abnormal Heart Sounds



1. Abnormal heart sounds
   a. S_1 (e.g., mitral stenosis, atrial fibrillation)
   b. S_2 (e.g., hypertension, aortic stenosis)
   c. S_3 (e.g., heart failure)
   d. S_4 (e.g., hypertension)
   e. Abnormal splitting (e.g., atrial septal defect)

2. Systolic murmurs
   a. Ejection murmurs (e.g., physiologic, aortic stenosis)
   b. Pansystolic murmurs (e.g., mitral regurgitation)

3. Diastolic murmurs
   a. Early (e.g., aortic regurgitation)
   b. Mid-diastolic (e.g., mitral stenosis)

4. Pericardial friction rubs

https://mcc.ca/objectives/expert/key/62/

What are the types of heart murmurs and other abnormal sounds?

A normal heartbeat has two sounds, a lub (sometimes called S1) and a dub (S2). These sounds are caused by the closing of valves inside your heart.

If there are problems in your heart, there may be additional or abnormal sounds.

Heart murmurs

The most common abnormal heart sound is a heart murmur. A murmur is a blowing, whooshing, or rasping sound that occurs during your heartbeat.

There are two kinds of heart murmurs:

- innocent (also called physiological)

  It's caused by the sound of blood moving normally through the heart.

- Abnormal

  An abnormal murmur in a child is due to congenital heart malformations, which means they're present at birth. It may need to be corrected with surgery.

  An abnormal murmur in adults is usually caused by problems with the valves that separate the chambers of your heart.

Galloping rhythms

Other heart sounds include a "galloping" rhythm, which involves additional heart sounds, S3 and S4:

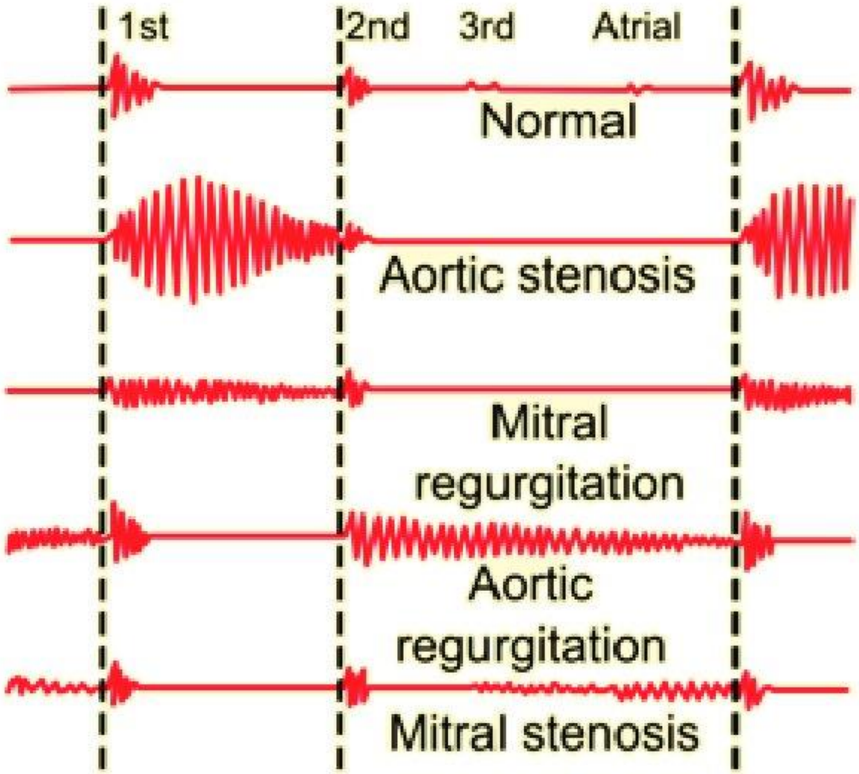
- **An S3 gallop** or “third heart sound” is a sound that occurs after the diastole S2 “dub” sound.
- **An S4 gallop** is an extra sound before the S1 systole “lub” sound. It’s always a sign of disease, likely the failure of the left ventricle of your heart.

You can also have both an S3 and an S4 sound. This is called a “summation gallop,” which can occur when your heart is beating very fast. A summation gallop is very rare.

Abnormal heart sounds are called heart murmurs. A heart murmur may occur in between regular heartbeats and sound like one of the following:

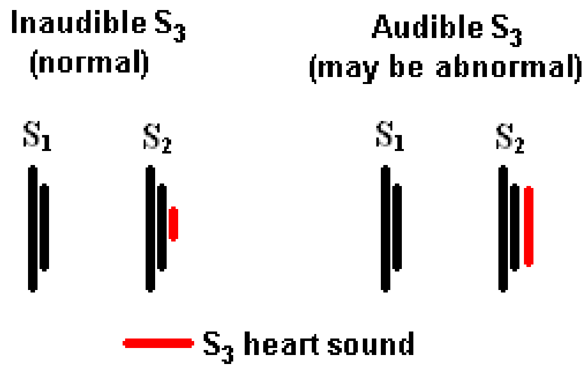
- a rasping تند تند زدن
- a whooshing
- a blowing دمیدن

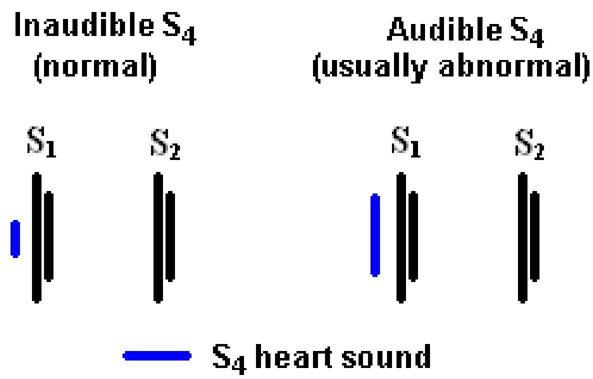
https://www.healthline.com/health/heart-murmurs-and-atrial-fibrillation#what-is-a-heart-murmur

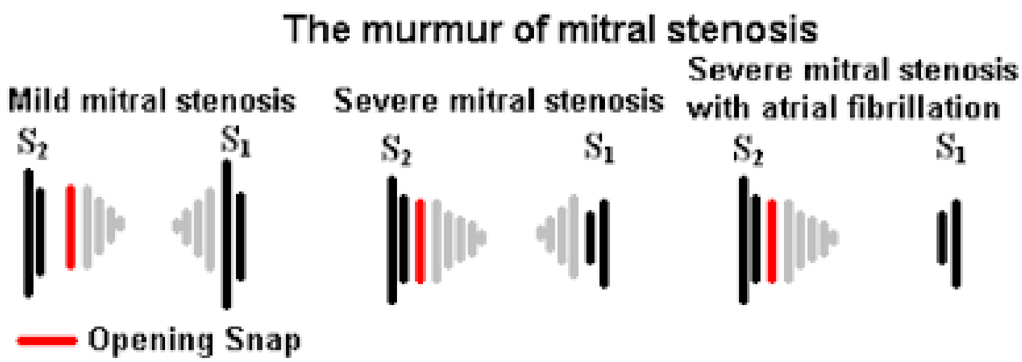


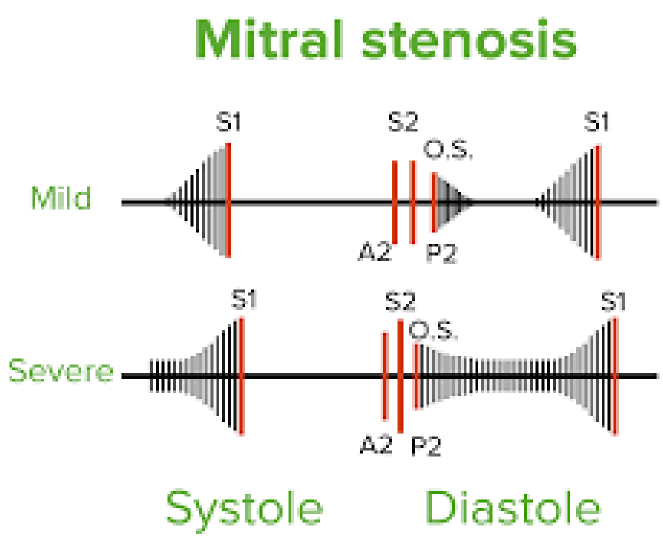
aortic stenosis: تنگی آئورت

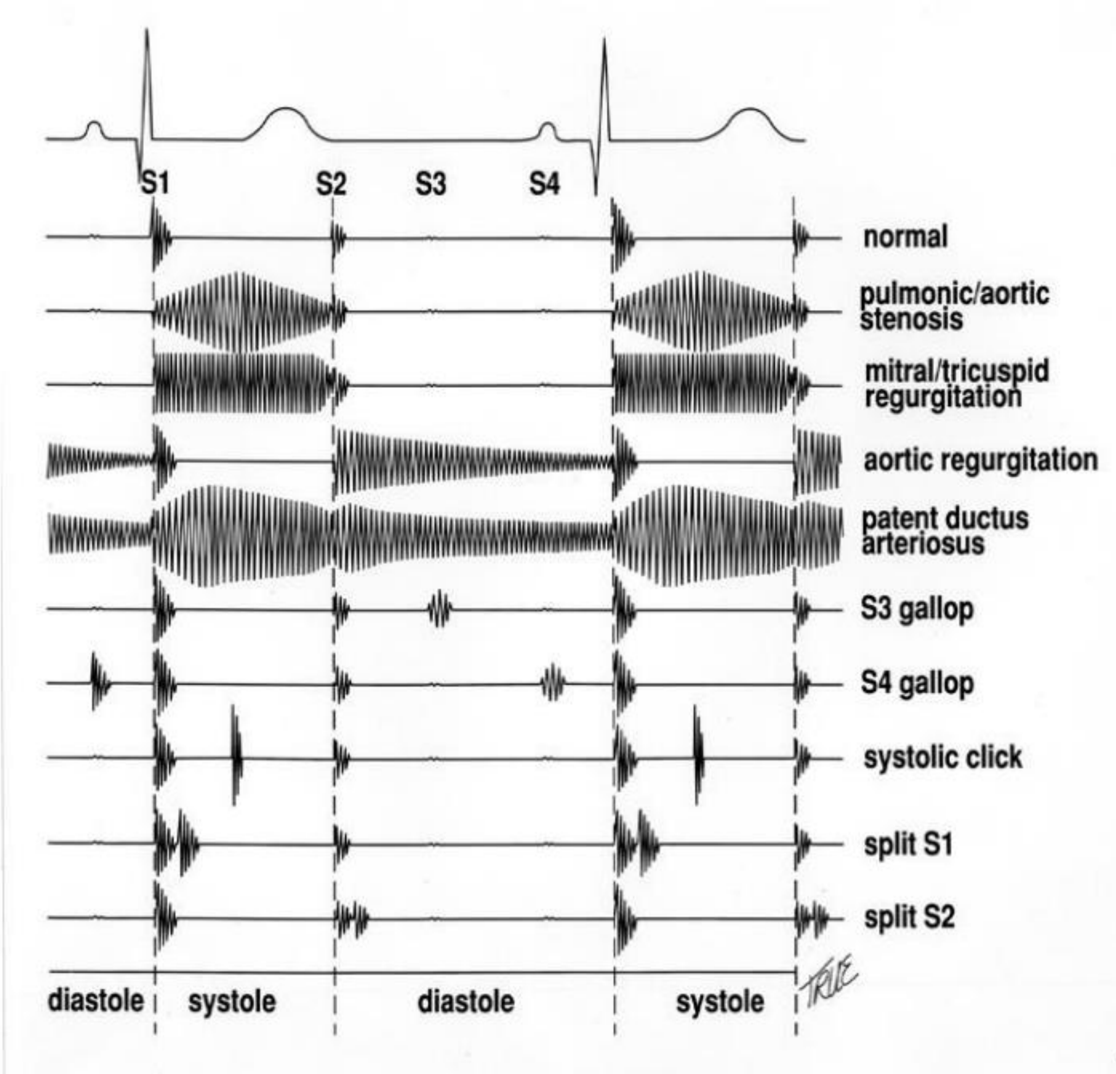
Mitral regurgitation: نارسایی میترال



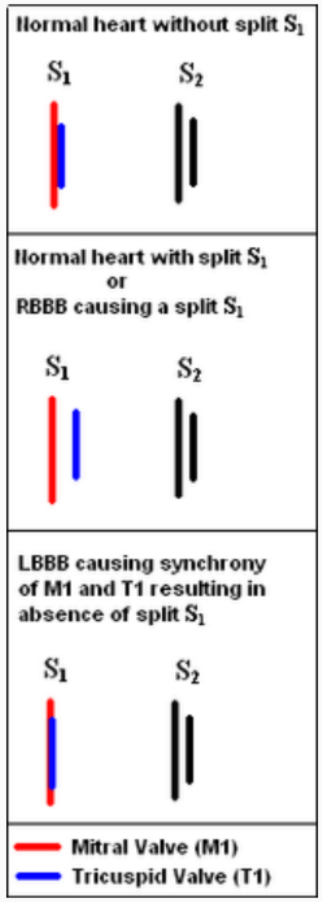

Inaudible S4
(normal)
Audible S4
(usually abnormal)
S1
S2
S1
S2
S4 heart sound


The murmur of mitral stenosis
Mild mitral stenosis
Severe mitral stenosis
Severe mitral stenosis with atrial fibrillation
S2
S1
S2
S1
S2
S1
Opening Snap


Mitral stenosis
Mild
S1
S2
O.S.
S1
A2
P2
Severe
S1
S2
O.S.
S1
A2
P2
Systole
Diastole

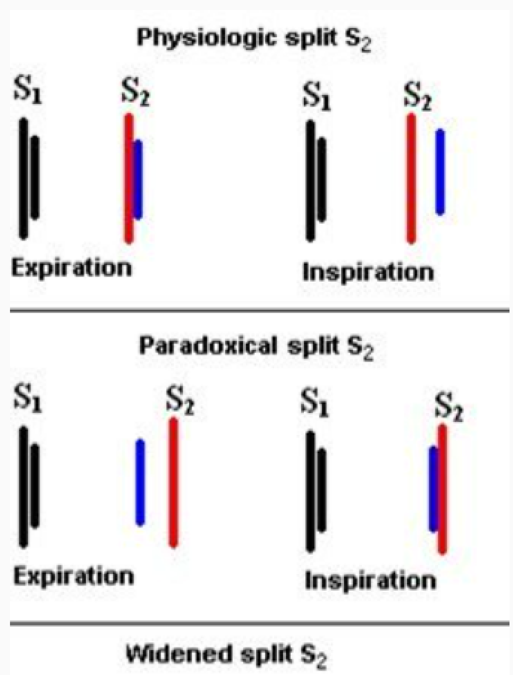


https://www.isvma.org/wp-content/uploads/2017/10/Cardiovascular_Physical_Eaxamination.pdf

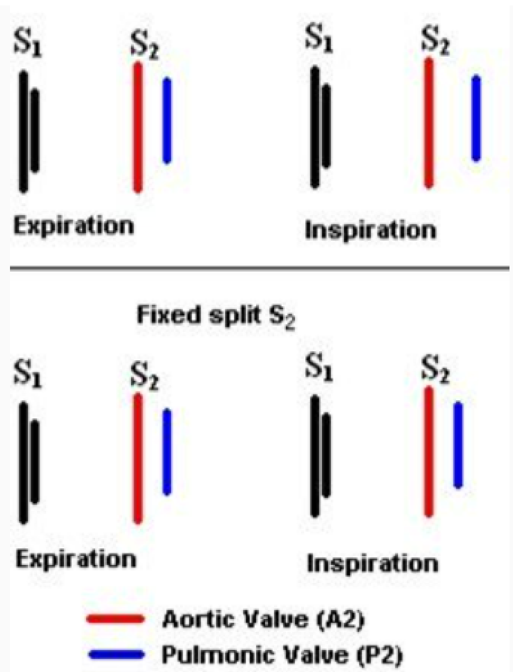


Second Heart Sound (S2)

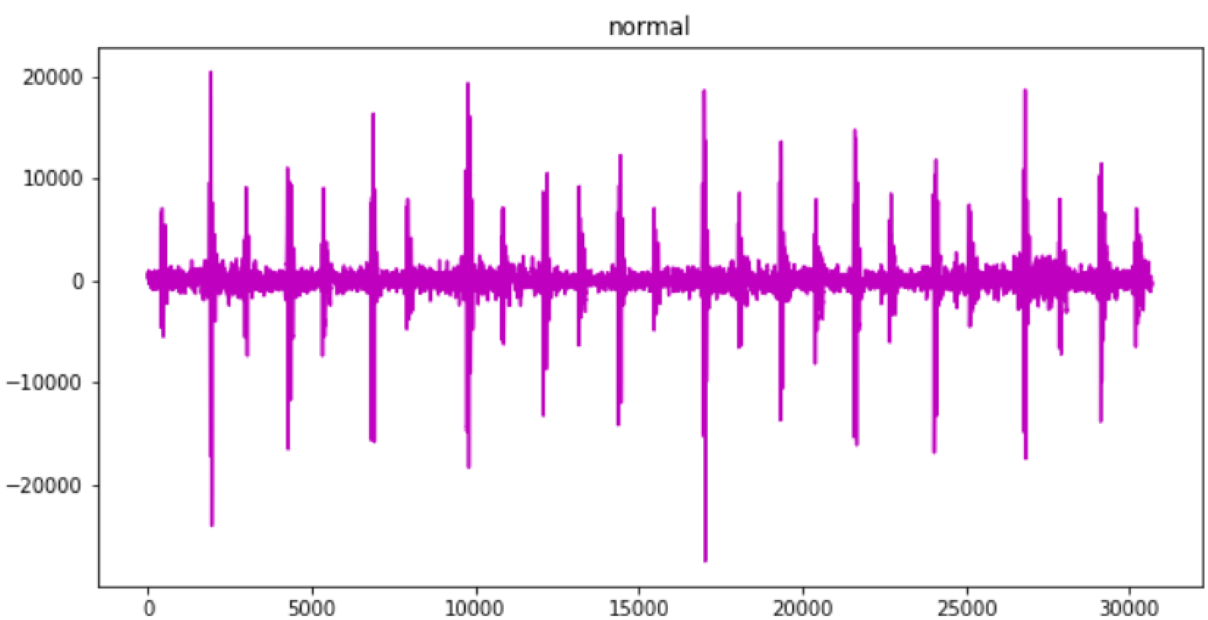
The second heart sound is produced by the closure of the aortic and pulmonic valves. The sound produced by the closure of the aortic valve is termed A2, and the sound produced by the closure of the pulmonic valve is termed P2.

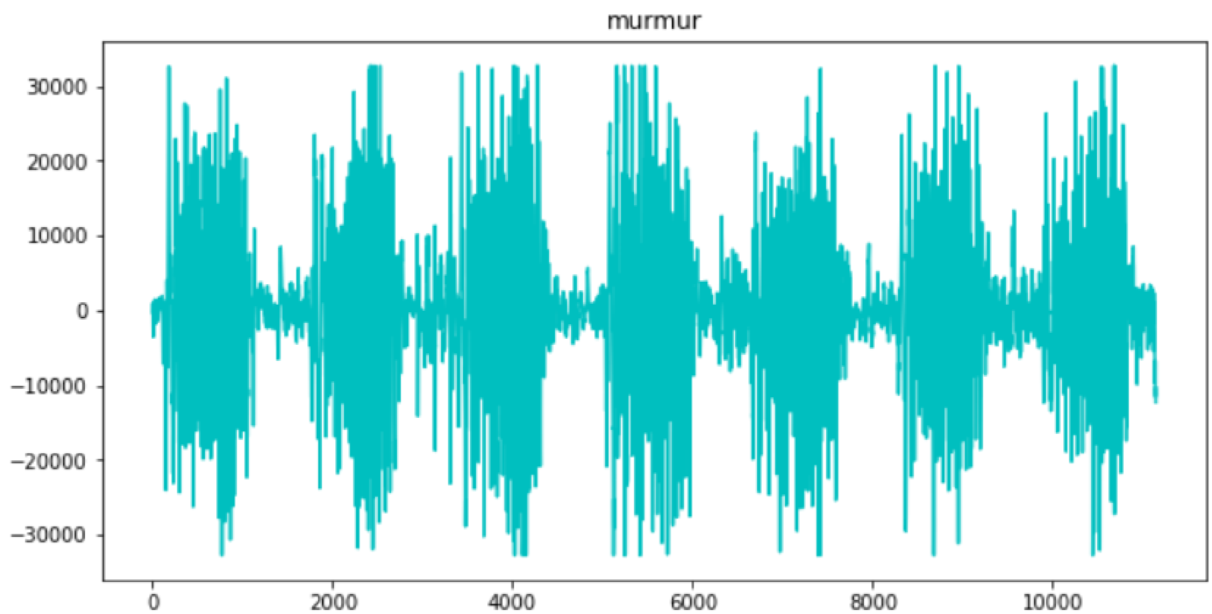


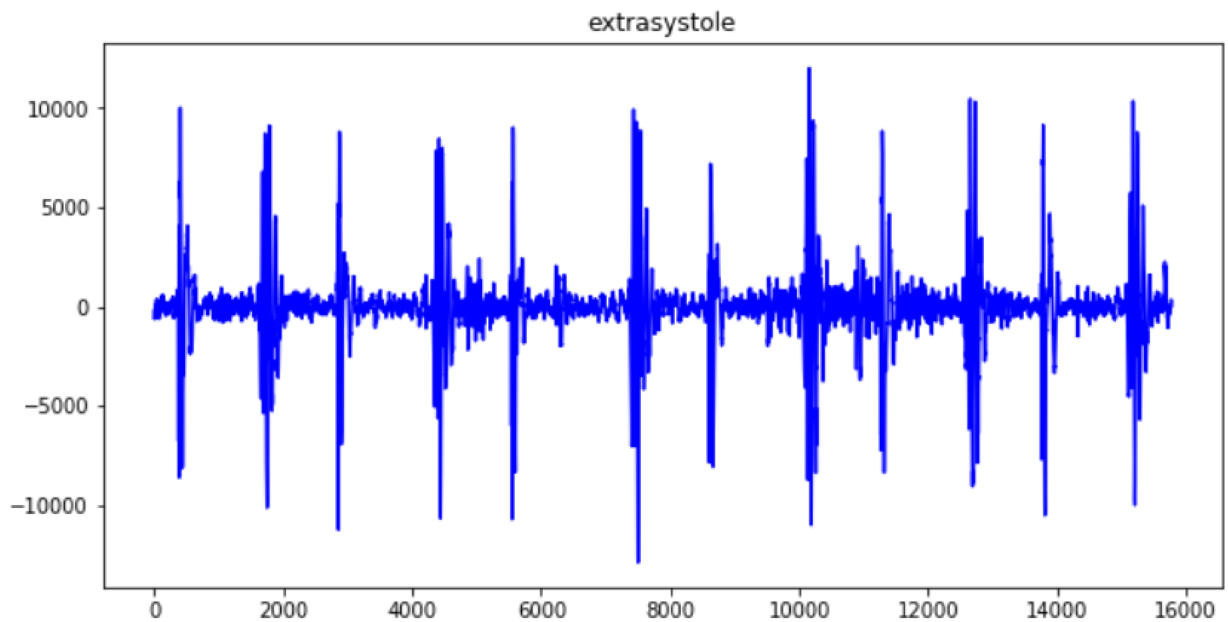
The A2 sound is normally much louder than the P2 due to higher pressures in the left side of the heart; thus, A2 radiates to all cardiac listening posts (loudest at the right upper sternal border), and P2 is usually only heard at the left upper sternal border. Therefore, the A2 sound is the main component of S2.



https://www.healio.com/cardiology/learn-the-heart/cardiology-review/topic-reviews/heart-sounds







Symptoms

Harmless (innocent) heart murmurs usually don't cause any other symptoms.

Symptoms of worrisome heart murmurs depend on the cause. Heart murmur symptoms may include:

- Blue or gray fingernails or lips
- Chest pain
- Cough that doesn't go away
- Dizziness
- Swollen liver
- Swollen neck veins
- Fainting
- Heavy sweating with little or no activity
- In infants, poor appetite and lack of growth
- Shortness of breath
- Swelling or sudden weight gain

Quality Index Using SNR

The range of SNR (Signal-to-Noise Ratio) for heart and lung sound signals can vary depending on the recording conditions and the quality of the equipment used.

Generally, a higher SNR value indicates a better quality signal with less noise interference. In the context of heart and lung sound signals, an SNR value above 0 dB is usually considered acceptable for clinical analysis.

For **heart sound** signals, SNR values in the range of **10-20 dB** are generally considered **good quality signals**, while values **below 10 dB** may indicate a **poor quality signal** with a high level of noise.

For **lung sound signals**, the SNR values can vary depending on the specific recording conditions, the location of the microphone, and the respiratory cycle of the patient. SNR values in the range of **5-15 dB** are typically considered **acceptable** for lung sound analysis, although higher values are preferred. Values **below 5 dB** may indicate a **poor quality signa**l with a high level of noise.

It's important to note that the interpretation of SNR values requires expert knowledge in auscultation and signal processing. Therefore, it's recommended to consult with a healthcare professional or a signal processing expert for accurate interpretation and diagnosis of heart and lung sound signals.

The normal **heart and lung** sounds range from **20 to 1200 Hz**. Heart sounds are in the lower frequencies, and thus generally easier to hear for those with a high-frequency hearing loss.

It is commonly admitted that lung sounds' frequency is in the frequency range [50, 2500 Hz], and that tracheal sounds can reach up to 4000 Hz; this allows to define a sampling frequency at 8 kHz.

Heart Sound	Duration(sec)	Frequency Range(Hz)
S1	0.1-0.12	30-45
S2	0.08-0.14	50-70
S3	0.04-0.05	<30
S4	0.04-0.05	<20

Cardiac Signal			Frequency Range (Hz)
Heart sounds	First heart sound	Normal	100–200
	Second heart sound	Normal	50–250
Heart murmurs	Systolic murmur	Aortic stenosis	100–450
		Pulmonary stenosis	150–400
		Mitral regurgitation	60–400
		Tricuspid regurgitation	90–400
		Atrial septal defect	60–200
		Ventricular septal defect	50–180
	Diastolic murmur	Mitral stenosis	45–90
		Tricuspid stenosis	90–400
		Aortic regurgitation	60–380
		Pulmonary regurgitation	90–150
	Continuous murmur	Patent ductus arteriosus	90–140

Lung sounds	Frequency Range	Temporal Features
normal	Low pitched, 50 Hz to 1000 Hz, up to 2500 Hz	Soft, longer and loader inspiration over expiration
crackles	Low (fine) or high pitched (coarse) 100 Hz-2000 Hz	Duration (inspiration + expiration) < 20 ms
wheezes	100 Hz to 1 kHz	80 ms < Duration < 250 ms
ronchi	< 300 Hz	Duration > 100 ms

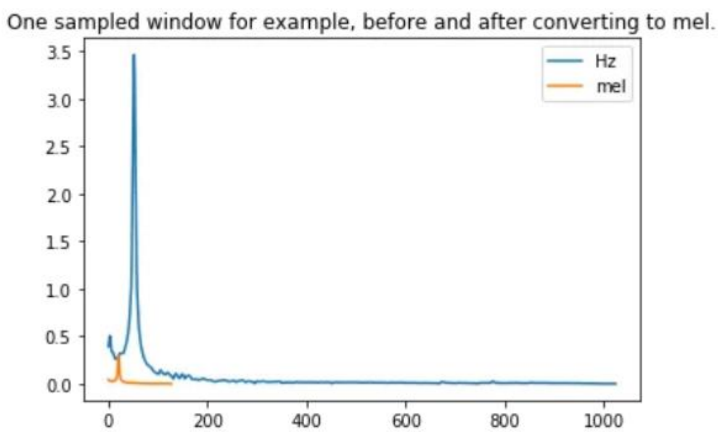
Mel Spectrogram

Creating the Mel-Spectrogram

The difference between a spectrogram and a Mel-spectrogram is that a Mel-spectrogram converts the frequencies to the mel-scale. According to the University of California, the mel-scale is **"a perceptual scale of pitches judged by listeners to be equal in distance from one another"**. If you are familiar with playing or reading music, this may help you visualize and understand the conversion and reasoning. Let's go ahead and picture this as notes on a musical scale:

- From C to D is one whole step, and from D to E is another whole step. Perceptually to the human ears, the step sizes are equal.
- However, if we were to compare these steps in hertz, they would not be equal steps. A C is around 261.63 Hz, a D is 293.66 Hz, and an E is 329.63 Hz.
- C to D difference = 32.03 Hz
- D to E difference = 35.37 Hz

As the notes go higher in octave, the difference between the steps dramatically increases. Mel-spectrograms provide a perceptually relevant amplitude and frequency representation.



$$\boxed{m = 2595 \cdot log(1 + \frac{f}{500})}$$

$$\boxed{f = 700(10^{m/2595} - 1)}$$

How many mel bands?

40 128 60 90

It depends on the problem!

example like on our like western

Convert frequencies to Mel scale

1. Choose number of mel bands
2. Construct mel filter banks
3. Apply mel filter banks to spectrogram

Mel filter banks

1. Convert lowest / highest frequency to Mel

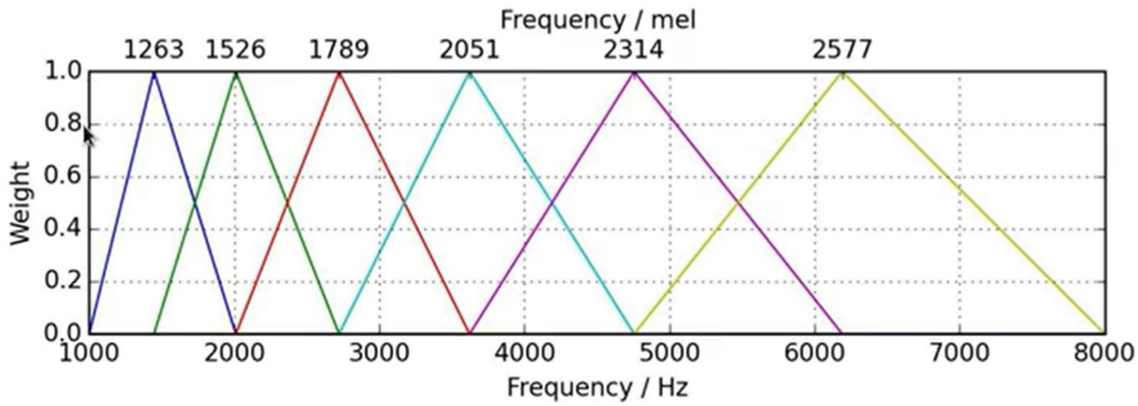
$$m = 2595 \cdot log(1 + \frac{f}{500})$$

1. Convert lowest / highest frequency to Mel
2. Create # bands equally spaced points

1. Convert lowest / highest frequency to Mel
2. Create # bands equally spaced points
3. Convert points back to Hertz

$$f = 700(10^{m/2595} - 1)$$

1. Convert lowest / highest frequency to Mel
2. Create # bands equally spaced points
3. Convert points back to Hertz
4. Round to nearest frequency bin
5. Create triangular filters



As we have six bands equally spaced points, we have six triangular filters as well.

Applying mel filter banks to spectrogram

M = (# bands, framesize / 2 + 1)

Y = (framesize / 2 + 1, # frames)

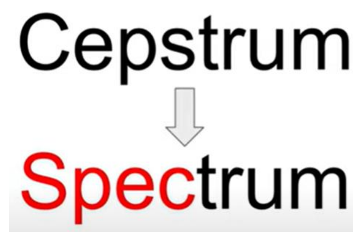
1. Choose number of mel bands
2. Construct mel filter banks
3. Apply mel filter banks to spectrogram

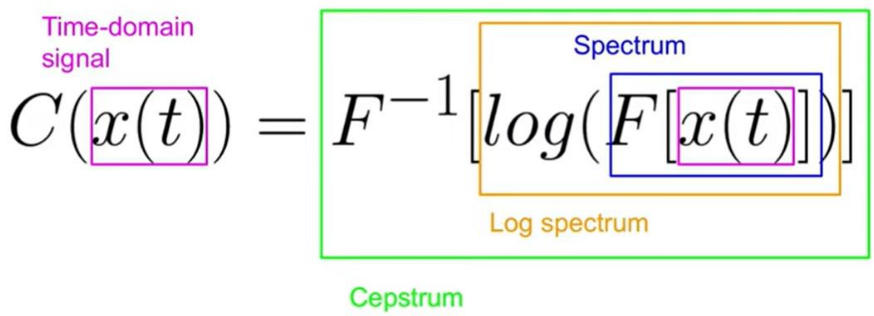
Mel spectrogram = MY

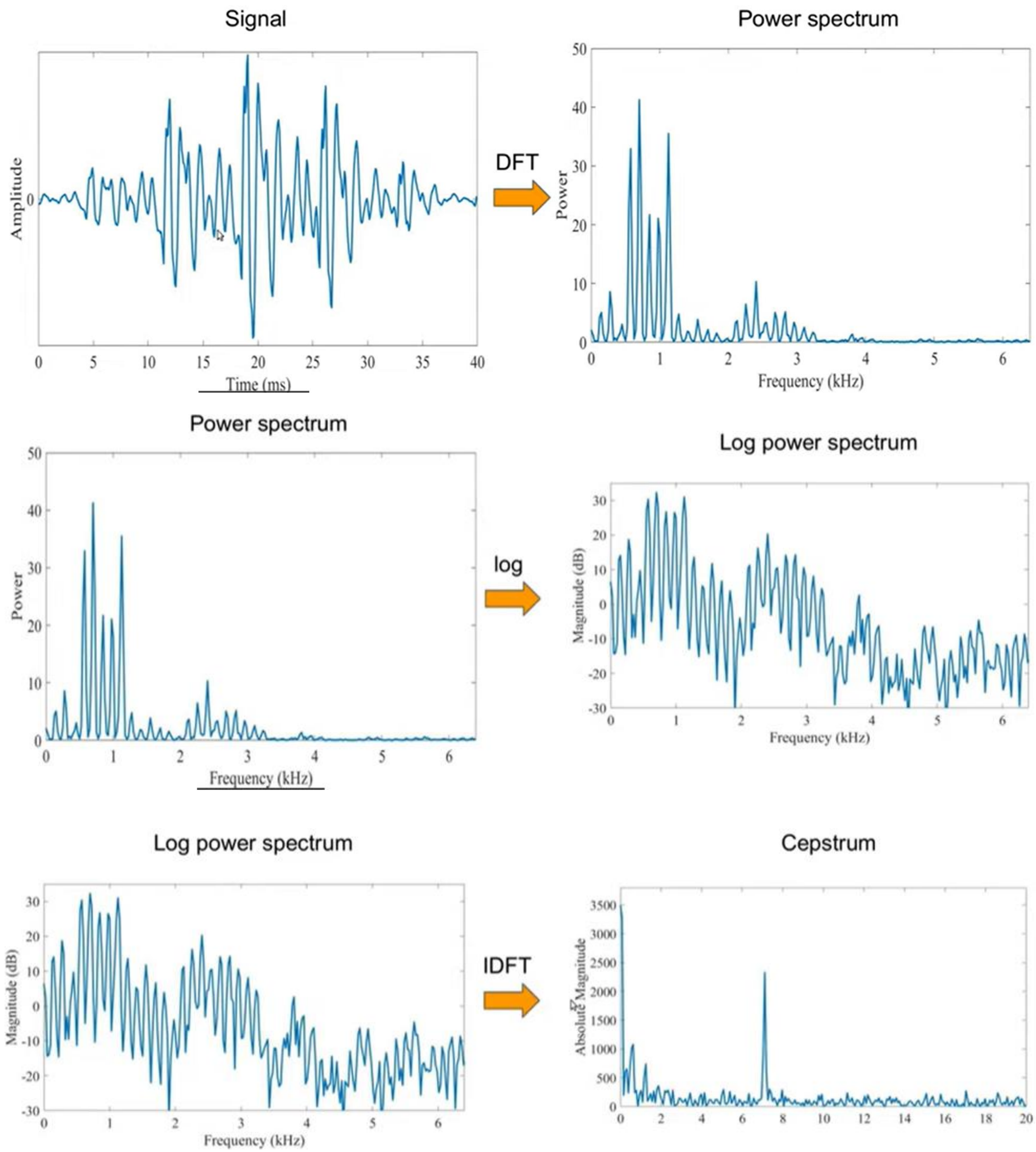
(# bands, # frames)

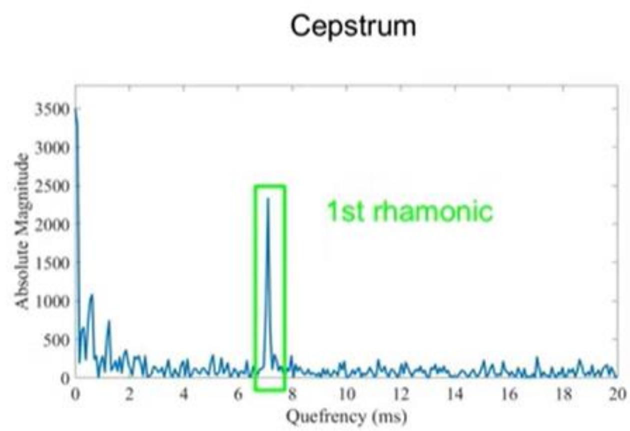
MFCC

The number of features set in mfcc function is the same as the number of rows in the plot of mfcc spectrogram.

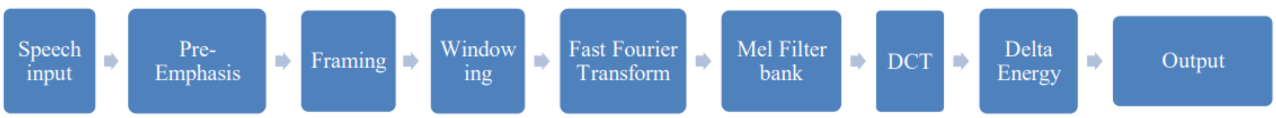


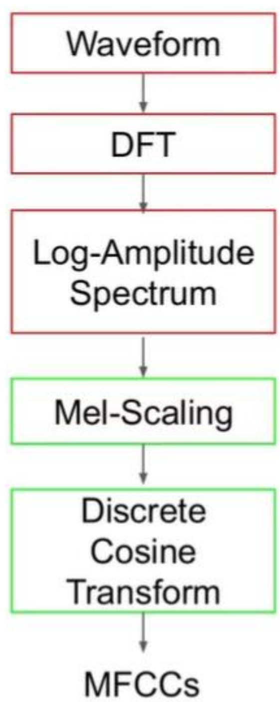



Signal
Amplitude
Time (ms)
DFT
Power spectrum
Power
Frequency (kHz)
Power spectrum
Power
Frequency (kHz)
log
Log power spectrum
Magnitude (dB)
Frequency (kHz)
Log power spectrum
Magnitude (dB)
Frequency (kHz)
IDFT
Cepstrum
Absolute Magnitude



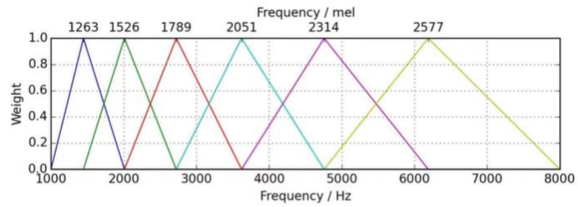
II. FEATURE EXTRACTION





Why Discrete Cosine Transform?

- Simplified version of Fourier Transform
- Get real-valued coefficient
- Decorrelate energy in different mel bands



http://iosrjen.org/Papers/vol4_issue8%20(part-1)/D04812125.pdf

MFCCs are commonly derived as follows:

1- Take the Fourier transform of (a windowed excerpt of) a signal.

2- Map the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows or alternatively, cosine overlapping windows.

3- Take the logs of the powers at each of the mel frequencies.

4- Take the discrete cosine transform of the list of mel log powers, as if it were a signal.

5- The MFCCs are the amplitudes of the resulting spectrum.