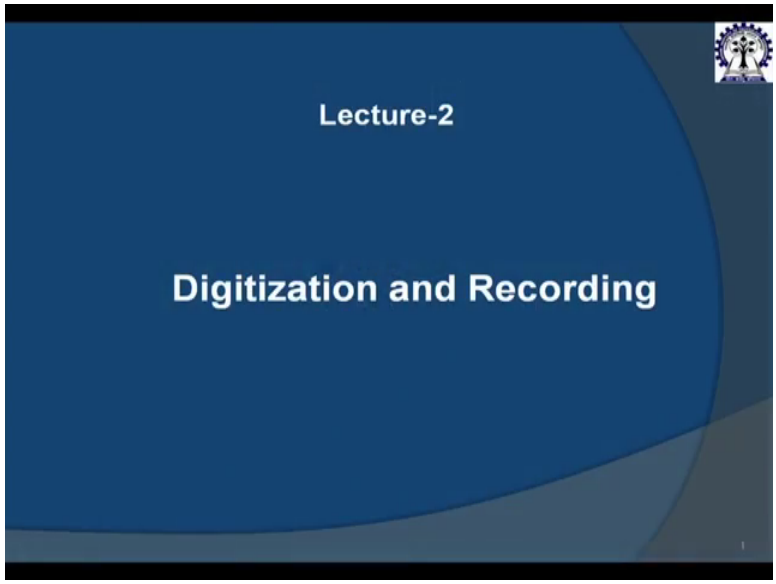
Digital Speech Processing
Prof. S. K. Das Mandal
Centre for Educational Technology
Indian Institute of Technology, Kharagpur

Lecture – 02
Digitization and Recording

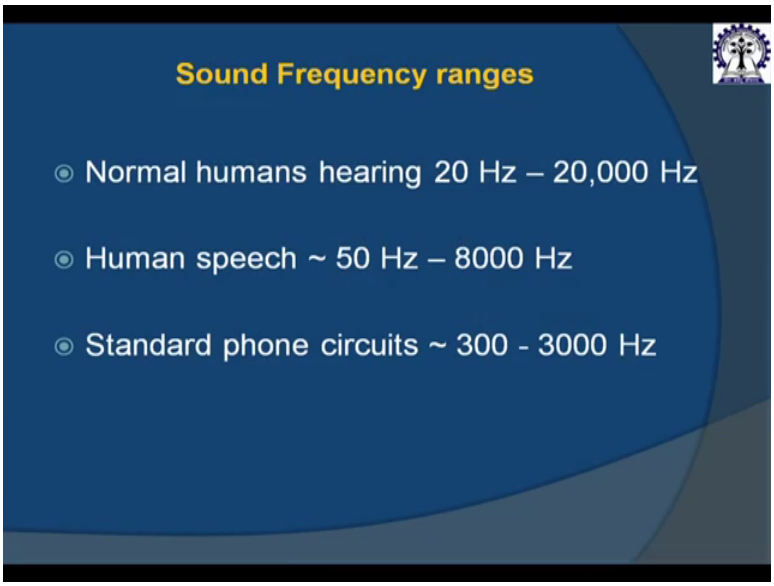
So I will come back. So, this lecture is deal with the digitization and recording.

(Refer Slide Time: 00:24)



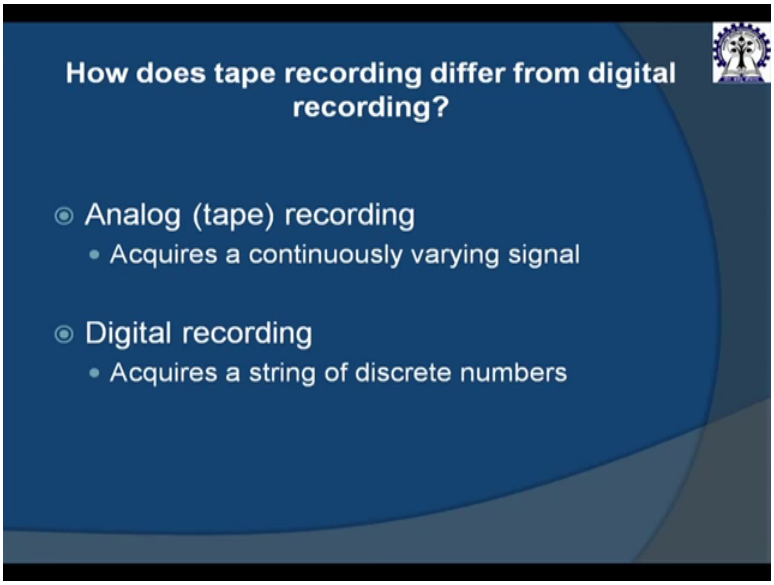
What I will discuss before you go to the any speech or this kind of things. I should you should familiar how to record the speech. Many people know that how to record the speech, but at least some you can say the skill kind of things I will explain here, not details of digitization details of this that kind of things I is not clear. But some what do you mean by digitization or kind of things I should explain.

(Refer Slide Time: 00:53)



So, if you see the normal human hearing is 20 hertz to 20 kilo hertz, we know that. And human speech is 50 hertz to 8 kilo 800 or 8 kilo hertz and standard phone circuits or telephone speech is 300 hertz to 3 kilo hertz or 3.3 kilo hertz, sometimes say 3.3 kilo hertz. So, this is the bandwidth of the human speech, normal human speech 20 hertz to 20 kilo hertz ok.

(Refer Slide Time: 01:21)



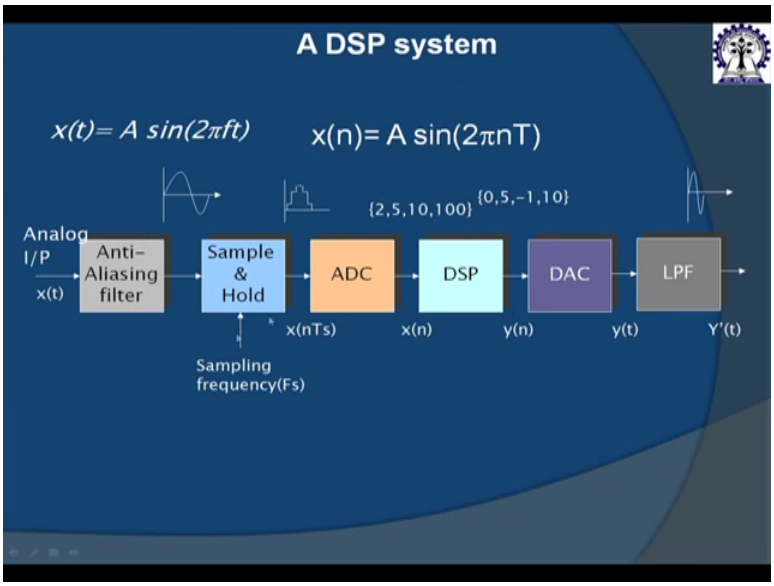
Now, what I want that I want the digitization of the speech. What I said that speech is in acoustic speech what about the human being is produce that human speech is in acoustics

wave. So, once the human being produce the speech any to any kind of speech processing I should have to digitize the speech because we said the digital speech processing. So, I say something I have to converted that acoustics wave to digital speech domain first analogue signal then analogue to digital conversion. So, how do we do that is we use a micro phone and amplifier and then micro phone amplifier then analogue to digital conversion to convert the acoustics wave to a digital signal in this speech.

(Refer Slide Time: 02:02)

So, how do we do that? You know that this is a basic theory this is called sampling theory that sampling frequency the this is the complete circuits that.

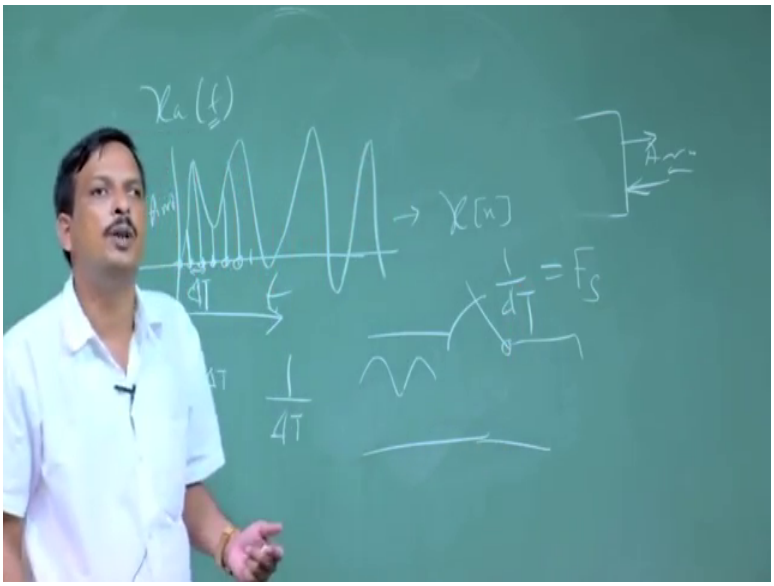
(Refer Slide Time: 02:25)



This is the normal DSP systems. So, I can say analogue speech do not I am not details of the details of this every blocks. So, those are the signal processing domain, but you should know what is there in computer. So, there is a analogue speech which is coming from the micro phone may be converted to the going to the anti aliasing filter then sampling hold then ADC analogue to digital conversion. Then you do the digital signal processing then again if I say the human being human listeners cannot listen that electrical signal. Again I have convert that electrical or digital signal to analogue signal and analogue should be played in the loud speaker to produce the acoustics wave.

Now so, sampling frequency, analogue to digital conversion sampling frequency is one of the important issue. So, basically what it is it is that may be some of some of you are there who do not know the digital signal processing or analogue to digital conversion. So, just I explain the very briefly that one.

(Refer Slide Time: 03:28)



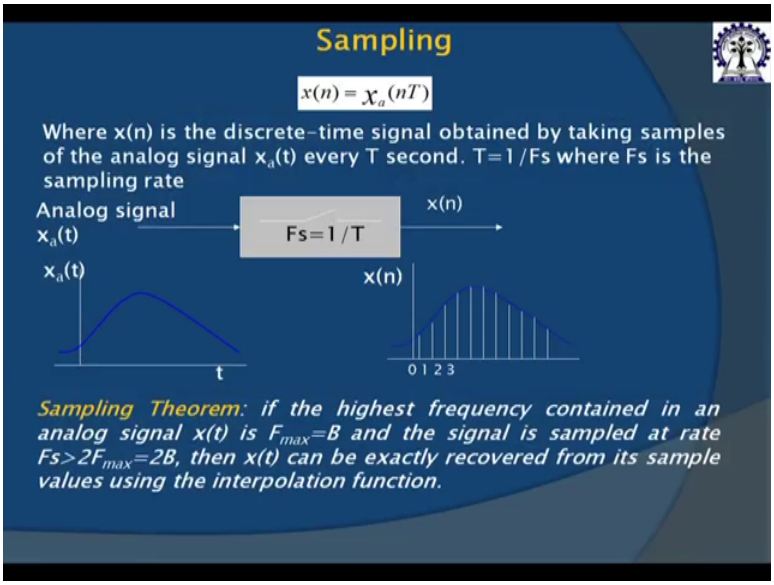
So, let us x a t is a time domain signal. What do you mean by time domain signal? Here is the time and here is the amplitude. So, the signal is varying like this way. So, that is the time domain representation of the signal analogue signal. Since analogue signal is continuous if you see this is a continuous signal. Now I want to convert this signal to a digital signal which is x n. I convert this same signal to a analogue digital signal which is x n. Instated of x t, I convert to x n number what is nothing but a, so digital signal is nothing but a integer number.

So, how do you do it? First there is a time which is continuous; I have to convert this time to a some instant with a fixed interval, that is delta t is a fixed interval delta t. So, I can say that I am taking the x a t as a sample or I can think about a switch. That suppose this is a switching circuits. I there is a x a t is going on, I am closing the switch with a frequency with a duration of delta t. So, for this delta t I take the signal then again I had take the signal here, I take the signal in here, I take the signal in here take the signal in here.

So, what I am doing? I converting the continuous t with a n t with a duration of delta T. So, n delta T. So, n 0; that means, here 1 means here 2 means here 3 means here 4 means here 5 means here. So, if I think about the frequency representation of this delta T 1 by delta T is a frequency, how do what is the frequency operation of this switch, with a

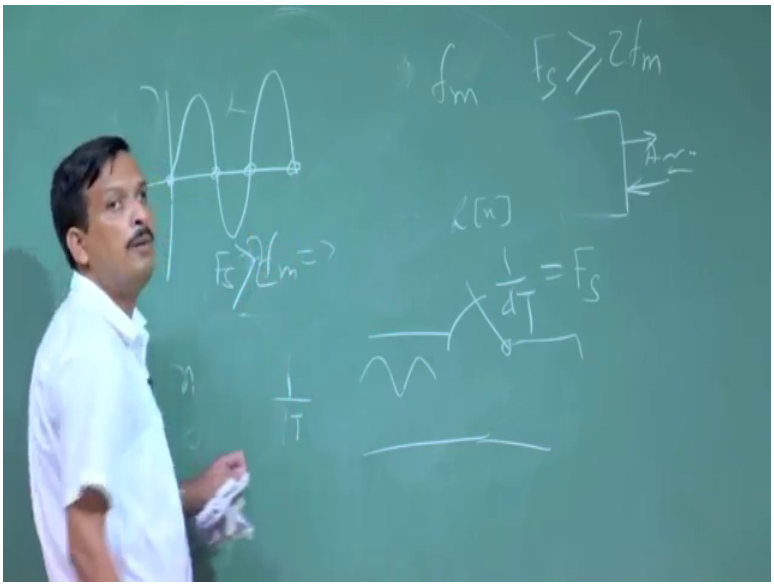
frequency is 1 by delta T. So, which is nothing but a called F s sampling frequency, F s. So, sampling theorem if you say if the there is a sampling theorem.

(Refer Slide Time: 05:53)



This is a sampling theorem which explain this slide I am not saying again and again. So, I can say if the sampling theorem say that, it is possible to completely recover this signal from this sampled signal. There is a some header or some constant what is those constant? It constant says that if the baseband frequency of this signal is bandwidth is B. Baseband or you can say the bandwidth is B means let us the highest frequency component of this signal is f m. If I sample this signal which is F s which is which must be greater than or equal to 2 F m. This the header. It is possible to recover the signal if it is greater than or equal to twice f m. Why it is greater than or not all time equal to? Because let us think about a pure sinusoidal wave, if it is a pure Sinusoidal wave.

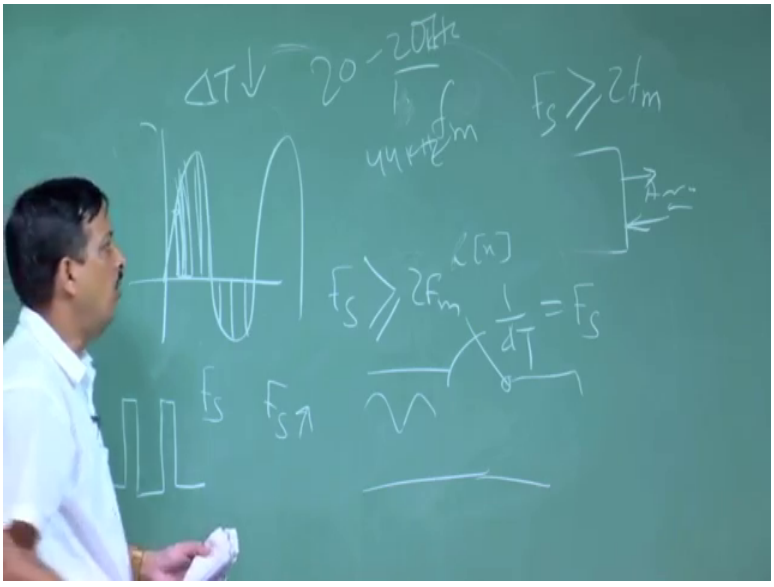
(Refer Slide Time: 07:02)



If it is 2 f if f m is equal to F s is equal to 2 f m then all the sample will be here. So, I cannot recover the signal. So, in that case I required F s must be greater than twice f m ok.

So, instead of taking the analogue signal, so on the analogue signal I am taking some sample instant.

(Refer Slide Time: 07:30)



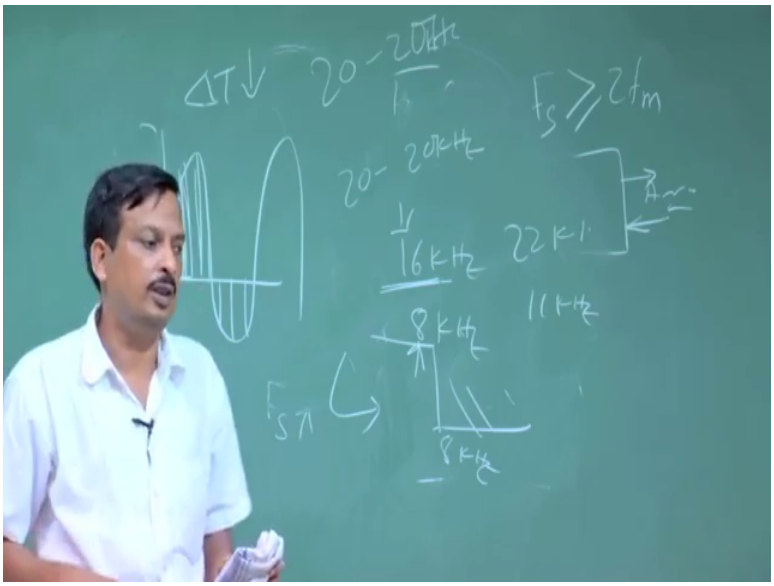
That is called that; how do we take the sample instant? Multiplying this signal with a sampling frequency impulses impulse signal whose frequency is f s. So, when I can say

the impulse is there when the impulse is present, I take the measurement of the signal, that is f s. Now if you see if my F s is increases, this gap will be decreases delta s delta T will be decrease if F s is increases delta T will be decreases.

So, if I use the if my sampling frequency is much much larger, then I can get accurately the signal. But right at the I cannot lowest I the guider is that Fs must be greater than equal to 2 F m, but I can take any sampling frequency. So, let us human speech has a 20 hertz to 20 kilo hertz. I can take the sampling frequency just 44 kilo hertz I can take it which is much much above the 2 times of the which is much above the 2 times of the 20 kilo hertz 44 40 kilo hertz 44 kilo hertz I have taken. Now if I increase the sampling frequency, I said that I can accurately take the signal ok.

Now, sometime you see although human speech is 20 hertz to 20 Kilo hertz.

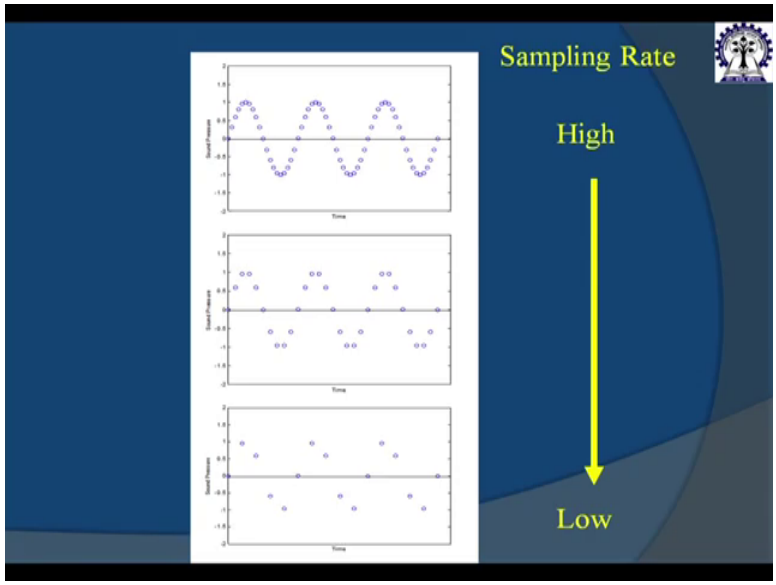
(Refer Slide Time : 09:13)



We have taken a sampling frequency let us 16 kilo hertz. What is meaning? Meaning is that, I am restricted the speech signal, but I am interested the speech signal. So, if it 20 kilo hertz is sampling frequency. So, as frequency of the speech signal is 8 kilo hertz. So, human speech has to be possibly a filter, whose highest cut off frequency is low pass filter is 8 kilo hertz upper up to above 8 kilo hertz all frequency has got down to 0 or you can discard all those frequency. So, sampling frequency based on the sampling frequency you know up to what frequency I can get of this speech signal. So, that is the sampling frequency idea. So, in speech sometime we use 16 kilo hertz; that means, highest

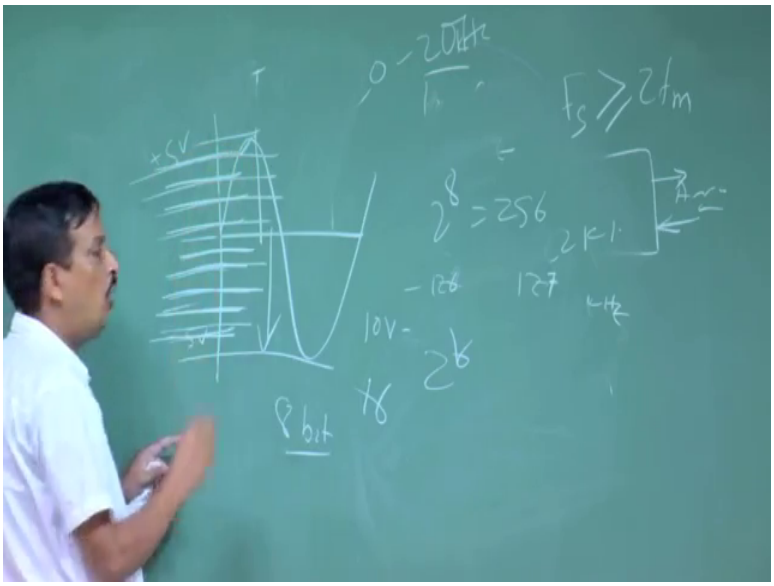
frequency component of the signal is 8 kilo hertz. If it is 22 kilo hertz then it is 11 kilo hertz. So, that is called sampling frequency, other aspect is quantization.

(Refer Slide Time: 10:25)



So, quantization is again I have to what I said I taking instant of this value. This is nothing but a voltage if t is amplitude it is voltage let us pass 5 volt minus 5 voltage here. So, this voltage has to be converted to a some step if you see here some step. So, each step I can represent by a binary number you are computer is (Refer Time: 10:55) binary number. So, I can say let us that my minus 5 volt to plus 5 volt or I an neatly draw it. So, that you can understand it that let us I have a signal whose varies from minus 5 volt to let us this is minus plus 5 volt and this is minus 5 volt ok.

(Refer Slide Time: 11:09)



So, this whole 10 volt I divided into a some level some level. So, how do we decided the level? Let us this whole 10 volt is divided in or you can say that I am representing this 10 volt by a 8 bit number. So, what is highest value of 8? 2 to the power 8, 8 bit number means integer value is 2 to the power 8. So, 2 to the power 8 256. If it is negative and positive. So, one bit is goes 4 side.

So, it is minus 127 to plus 127 or including 0 it will be 2 1 2. So, I am not saying that that part. So, making 0 is middle positive and negative understand 250 sorry, 2 to the power minus 127 or 126 you can take it. So, 8 bit number. So now, if I do that, then 250 I have divided this 10 volt in 256 level. So, plus side and minus side is there. I can divided this thing instead of 8 bit 16 bit. So, 2 to the power 16 level I can divided this thing.

So, once I divided this signal much more smaller gap the accuracy of the quantization will be increases or quantization error will be decreases. I am not detail discussing about the quantization error because that is there quantization error. So, what I am saying that if I recorded the speech signal with a high quantization the accuracy or error quantization error of the signal is reduced. So, when I going for a good quality recording should use quantization level should be very high instead of 8 bit.

(Refer Slide Time : 13:26)

If I 12 bit or 16 bit the quantization error is less, but other aspects is there.

What is other aspects? Once I increase the quantization level what is increasing data size because if it is 8 bit one sample is represented by 8 bit. If it is 16 bit single sample is represented by a 16 bit. Similarly s if it is 8 kilo hertz I can say in one second I have generated 8 k sample. In one second I have generated 8 k sample if each sample is represented by 8 bit. So, for one second speech recording will get 8 kilo bytes why 8 bit is 1 byte. So, 8 kilo byte similarly if I quantize it to the 16 bit t will be 16 kilo byte 16 bit is 2 byte. So, 16 kilo byte.

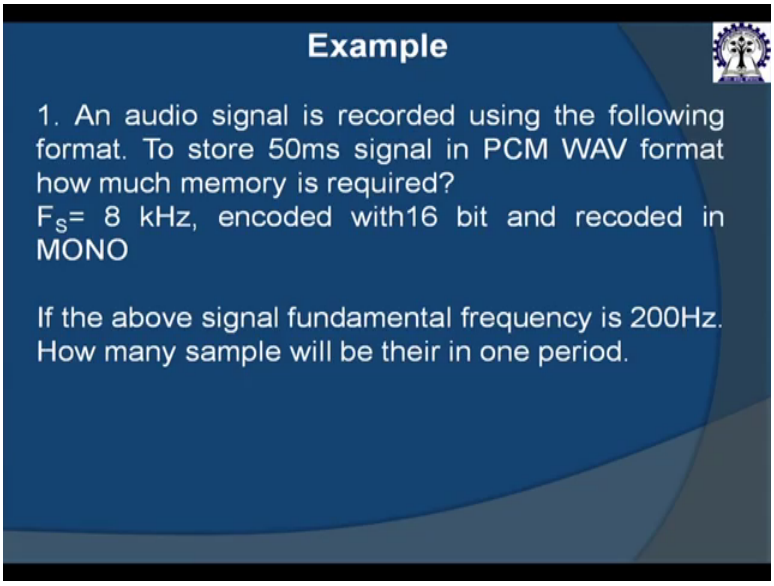
So, when I increase the quantization level although my signal to noise ratio is decreases means increases; that means, I quantization error is decreases, but my memory size is increases. Similarly once I increase the sampling frequency my accuracy of the digital signal will be increases, but what is increases size of the file is increases instead of 8 k if it is 16 k; that means, in one second I have generate 16 k sample if each sample is 8 bit then 16 kilo byte. So, that is a trade of what quantization level I should use what kind of sampling frequency I should use. So, when you record the speech signal suppose you were recording the speech signal for telephone speech which is bandwidth is 300 hertz to 3.3 kilo hertz.

(Refer Slide Time: 15:29)

So, if it is my signal is band limited to 3.3 kilo hertz I should not sample this signal with a very high frequency high sampling frequency. So, my 8 kilo hertz is sufficient even if double this thing nearest is 8 kilo hertz is sufficient to sample this thing.

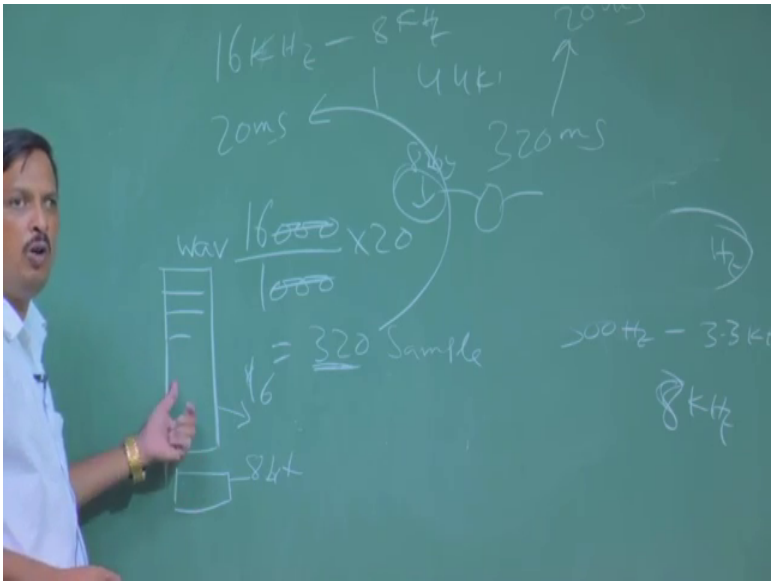
So, that is why the telephone bandwidth is 4 if it is 8 kilo hertz is sampling frequency 4 kilo hertz is the maximum frequency which can present in the speech. So now, I give a problem, let us I am not going details of this digitization things lest this is a problem.

(Refer Slide Time: 16:12)



An audio signal is recorded using the following format to store 50 millisecond signal in PCM WAV format, how much memory is required? Sampling frequency is 8 kilo hertz encoded with 16 bit and recorded in mono invert mono channel. So, if it is 8 kilo hertz 8 bit in one second it will generate 8 k sample. So, in 50 50 millisecond I have to know how much how much sample it will be generate and for each sample is 1 byte how much memory is required I can calculate. Similarly there is a another conversion also I have I should taught in here. Suppose sometime we said number of sample to time conversion is very important. Suppose I have a speech signal. I say a speech signal is sampling frequency is 16 kilo hertz.

(Refer Slide Time: 17:10)

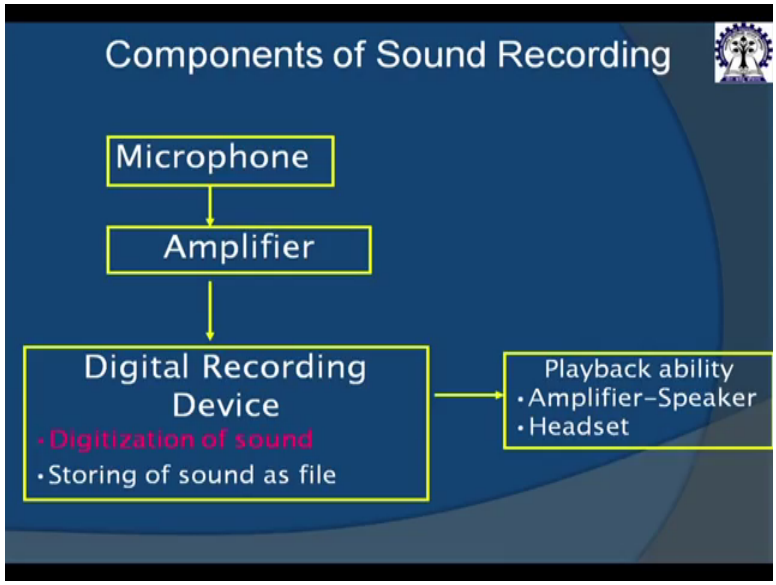


How much sample will be there within a 20 millisecond signal. So, in one second there will be a 16 kilo sample in 20 millisecond in one second into 20 320 sample. So, 320 sample represent 20 millisecond if it is 8 kilo hertz then how many sample will be there half of this half of this. Because it is half half of this if it is 44 kilo hertz then you can calculate how many samples will be there.

So, instead of 20 second I can say 30 second. So, once I say the window size of forty millisecond then you know how much sample will there. Similarly if I say if 320 samples I have taken the time domain representation is 20 millisecond can convert. This will be frequently used in speech processing number of sample to time time to sample. So, this is

the basic idea of digitization of the speech. Now there is a expertise required how do I record the.

(Refer Slide Time: 18:33)



Speech normal recording is there if you see I have connected if my if you see I have a collar mic connected to a since it is wireless mic connected to a wireless transmitter and there is a sound cord. And this is a one side talk this mic is converted to electrical signal this electrical signal goes to the sound cord digitize the sound cord is digitize the signal and recorded it. So, microphone there is a amplifier if it is required sometime microphone directly connected to the sound cord which contain the amplifier then digital recording device or which convert that analogue signal to digital signal then if I want to playback. So, I required a speaker or headphone these kind of things ok.

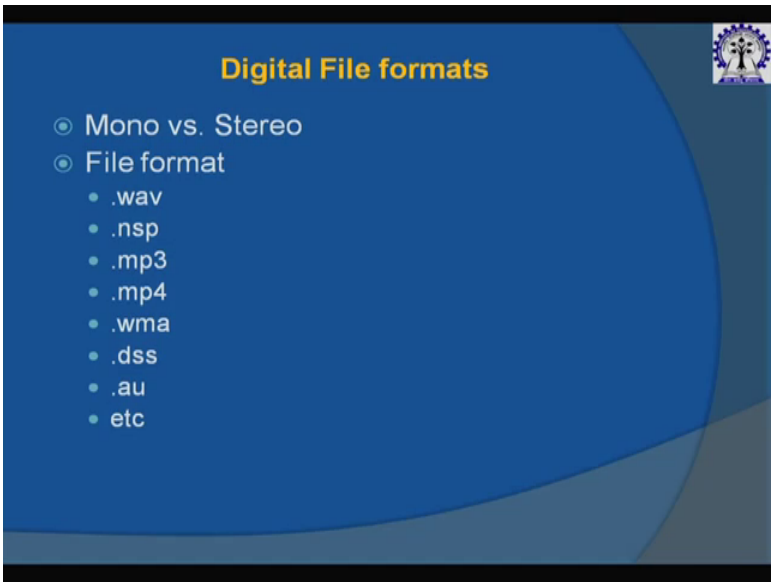
So, in digital recording what is done it is does the digitization and storing of the sound. So, digitization means it requires the sampling frequency and also n number of quantization bit. And also stereo or mono that I have not discussed details, but most cases in sound recording does in mono since it is a single channel that I want to record the human voice. So, when you analyse your voice in computer you should record the signal in mono format not stereo format it is not required.

Because there is no 2 source are there single source is speaking. So, it is mono is sufficient some things that how this is store in computer there is a different kind of file format mono what is stereo that is said stereo mean 2 channel mono means single

channel stereo sound I have not explained in details there is a stereo sound there is a 5.1 channel sound there is a dolby digital sound. So, Dolby digitalis is again one format it is not a sound you can say that it is a sound recording Dolby digital is sound you can say the compression technique or you can say sound compression procedure. So, Dolby digital is a format instead of you can say the it is not stereo mono kind of things. So, 8 5.1 channel sound we have heard mono sound heard stereo sound heard stereo means 2 channel mono means single channel 5.1 means 5.1 channel, I am not discussing that part that is a another course which is audio system engineering there I have discussed the mono and stereo kind of things.

So, what kind of file format is there? If you see that there is a file format lot of file format you see dot wav nsp mp3 mp4 wma dss au etc.

(Refer Slide Time: 20:57)

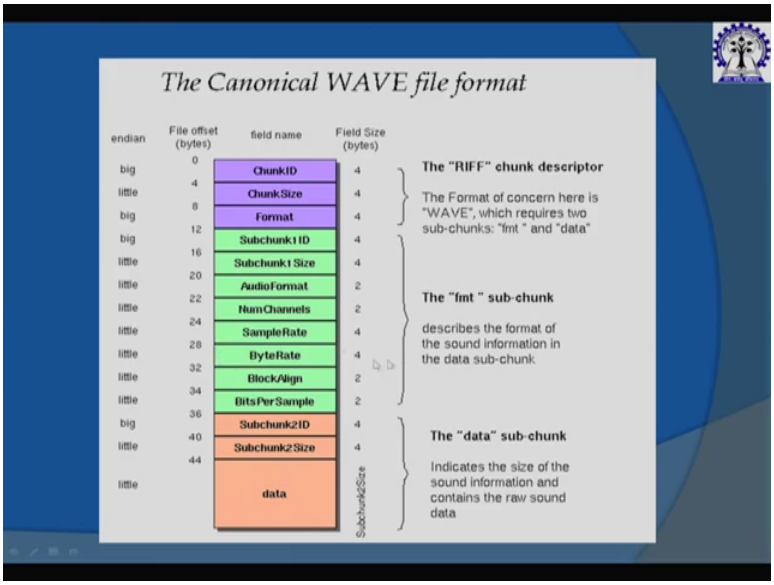


So, all are some or coded speech format some are non coded because it is just some stored the sample. So, if it is the mp3 it is a compressed. So, sound can be stored either compressed or without compress. If I stored the signal let us that in one second it is generate 8 kilo y I can compress it and store it and there is no compression technique and when I want to play the sound. I will decompress the signal and play it in the wave. Similarly when I process the sound I should extract the sound sample by sample. So, if it is a compressed format stored in a complex compressed format I have to decompressed it. And use the sample by sample or processing and again I can compressed it. Now any

audio compression or mp3 mp4 wm wma dss, there is a loss. If I compress the sound I have lost some sound. So, I cannot recover those signal those sound once I do the decomposed it and process the sound sample by sample. So, if I so, if I want to store the just simple recording sample by sample use PCM a wav. So, there is a PCM wav format header file Microsoft PCM a wav format header file.

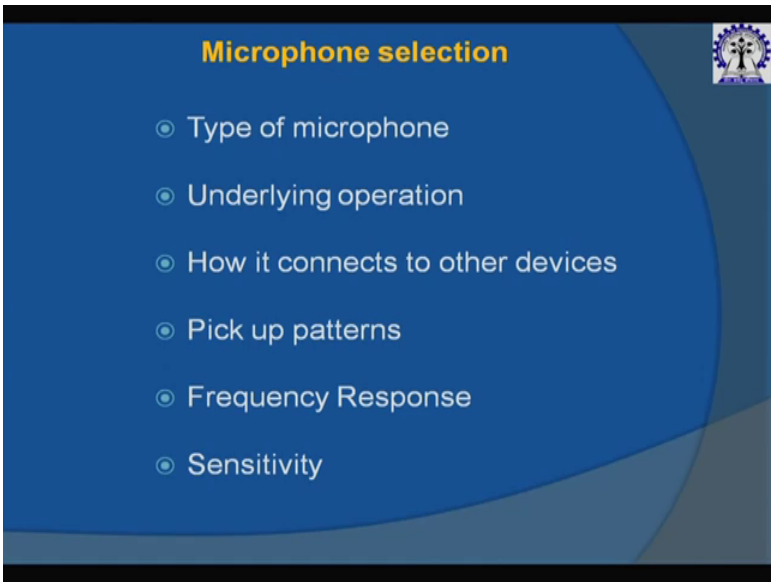
So, once I get the sound file dot wav format this is a binary file it contain whole recording of the sound let us I get this there is a header. So, if you see this header there is a different chunk position has different information. So, I have to know what I the sampling frequency of this sound and what is the encoded bit of this sound. If it is 16 bit; that means, 2 byte represents one sample. If it is 8 bit 8 bit represent one sample. If it is stereo all channels are same. If it is mono it is stereo there is 2 a channel recording is there. If it is mono single channel recording is there and there also sampling frequency is important. Because some unless I do not know the sampling frequency I cannot process it.Digital signal or this digital processing of this sound file. Sound file is required a sampling frequency. So, sampling frequency also stored in this.

(Refer Slide Time : 23:24)



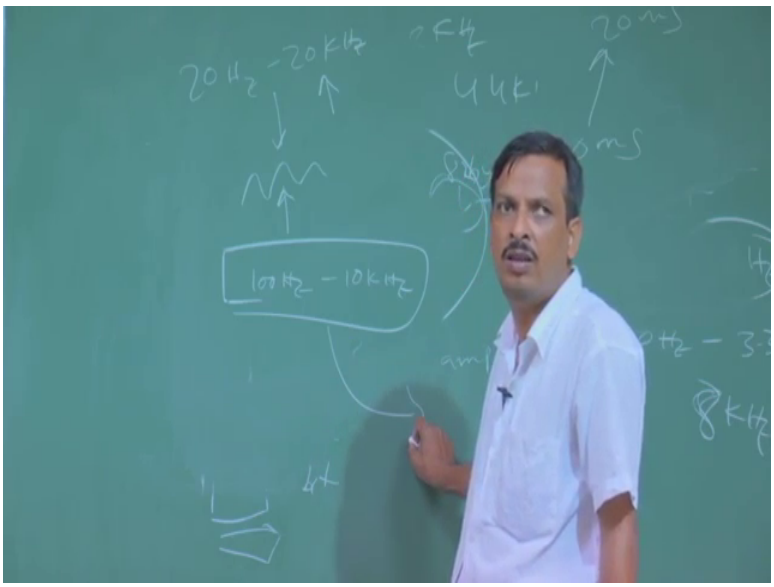
Sampling rate byte rate block again block size sub this. So, those format details are given in these slides. So, digital format uncompressed and compressed format what I have expressed and then I can go for the microphone selection.

(Refer Slide Time: 23:39)



Suppose I want to record I say you to record your name in your computer. So, how to record it? You may use your headphone and connected to the computer sound cord and you record the sound, but once the human being produce the speech it is range is 20 hertz to 20 kilo hertz. The acoustics wave contain 20 hertz to 20 kilo hertz sound.

(Refer Slide Time: 24:05)



Once I put the microphone in the front of the mouth. So, microphone has an limitation. Microphone has it is own property that it has to be convert this acoustic wave to a electrical signal. So, microphone may have a limitation. So, microphone may have a

sensitivity microphone may have a frequency response microphone has a pick up pattern all kinds of restrictions are there who know the details of microphone you know that there is a lot of parameters of the microphones are there.

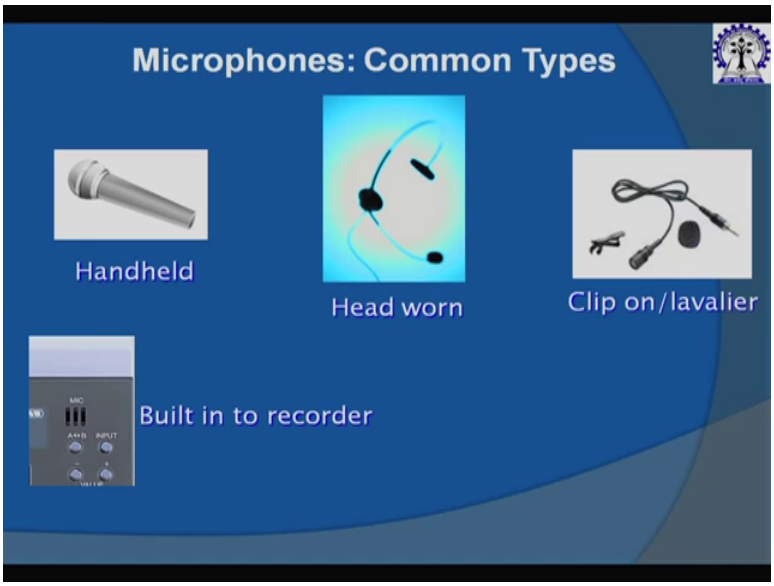
So, if you see I am recording a speech with a microphone whose frequency response is may be 100 hertz to 10 kilo hertz. So, I cannot get the acoustics wave after 10 kilo hertz, when we required in human speech, but I cannot get those speech after 10 kilos because my microphone has a limitation it can only record the signal or it can only transfer the acoustics wave to electrical signal of 100 hertz to 10 kilo hertz.

So, if I apply a 12 kilo hertz signal acoustics wave it may not produce any signal, it will not produce any signal. So, I am not getting any information in the electrical wave once that electrical signal I get I will pass through the digital cord. So, selection of the microphone frequency response microphone type how to connect it all are very important. So, microphone which microphone I should use t depends on what kind of application I want to develop.

Suppose I want to develop a speech coding for telephone channel then I should not use a very high end microphone which frequency response is 20 kilo hertz. Because my signal is 4 kilo hertz band limited. So, I can use 10 kilo hertz sensitivity of a frequency response of the microphone is very good for recording that telephone bandwidth even 5 kilo hertz microphone sensitivity up to 5 kilo hertz is sufficient for recorded the telephone channel signals. Now suppose I want to developed a speech corpus for research purpose. If I want that this corpus only used for scientific research, if I want that that should contain that whole human speech range of frequency then I should use a very high end microphone, which frequency response may be 15 kilo hertz or may be 20 kilo hertz ok.

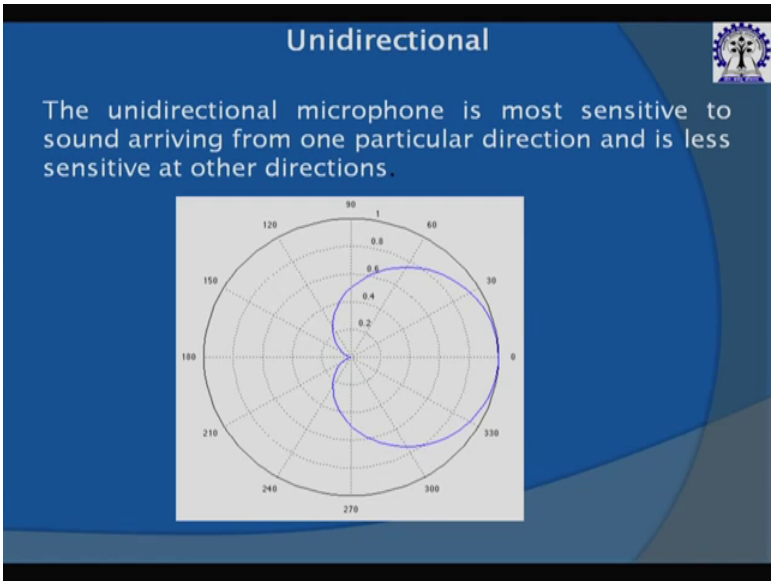
So, the microphone, then the amplifier is also there amplifier frequency response also important. So, amplifier and microphone whole frequency response should be higher than the my desired what frequency I want to record that things. And there is a if you know that there is a different kind of microphone hand handle microphone head mounted microphone all things are there.

(Refer Slide Time: 27:19)



So, all microphones the advantage and disadvantage am not going details of the microphones. Then there is a carbon mic piezoelectric mic dynamic mic condenser mic ribbon mic every mic has his own frequency response his own advantage and disadvantage. Then there is a connector is there then if you see the Directivity.

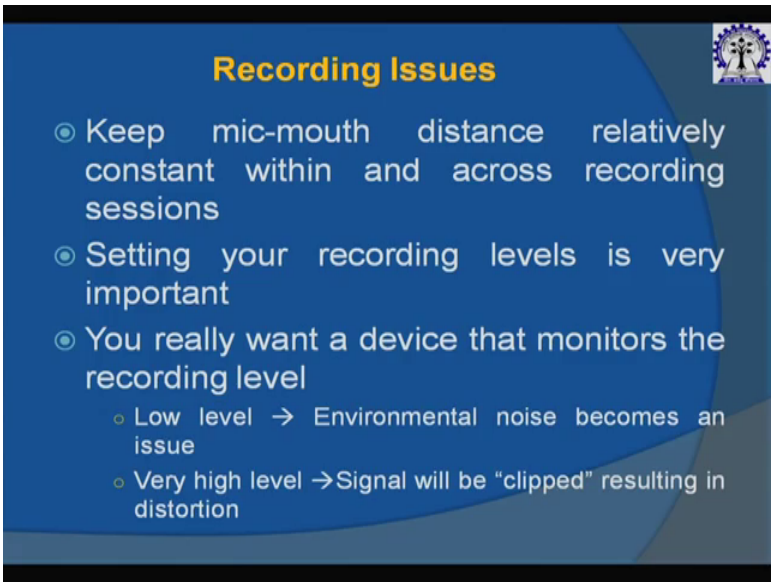
(Refer Slide Time: 27:42)



Microphone has may be omni directional may be bi directional may be uni directional all kinds of directional means you know which point microphone has a pick up; that means, if I use this kind of microphone if you see the omni omni directional; that means, that

microphone can pick up from any sound any direction. Now if it is a unidirectional that can only pick up a sound from in one direction, that is repeated in a polar (Refer Time: 28:11) slides are there this is the frequency response example of a microphone.

(Refer Slide Time: 28:17)



Then recording issue keep mic mouth distance relatively constant with and across recording session. Setting you recording level is very important. Suppose if you see there is a clipping one very important issue I will show you in software. There is a clipping is very important issue during the recording. And maximum people or you can say the maximum people say that one who record the signal that the signal is clipped. What do you mean by clipping?

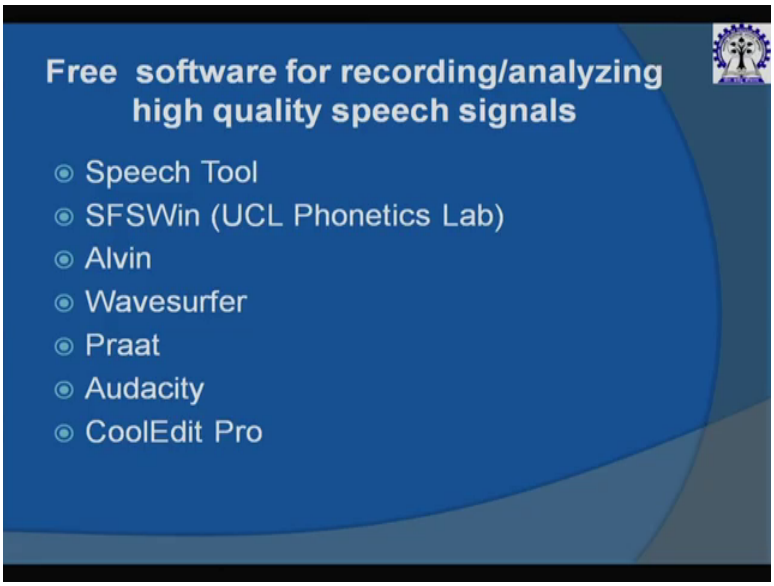
(Refer Slide Time: 28:52)

Now, suppose I have an my ADC or volume of the recording volume which is control the ADC level.

So, suppose this is 5 volt to minus 5 volt. This ADC can handle 5 volt to minus 5 volt. Now if I amplify my sound 7 volt to minus 7 volt. So, suppose 5 volt to minus 5 volt somewhere sine wave let us sine wave it is vary like this, if my limitation is 5 volt if I apply 7 volt. So, what will happen after 5 volt it will be flat. So, after 5 volt it will be flat response.

So, what is if it is flat what is happening? So, this portion I am not getting I am getting a flat. So, if it is clipped; that means, you developed a square wave there. So, original frequency component of this signal I cannot get it. I get the squarer content all the frequency all the frequency content. So, this kind of recording if you record the signal which is clipped. Then if it is clipped then your whole recording purpose is gone. Your recording whole recording purpose is means because it is developed lot of sound frequency also. So, be sure that signal is not clipped, recording level is not clipped. And environmental noise is very important.

(Refer Slide Time: 30:39)



And there is a Free software tools for recording there is a speech tools I have used cool edit pro I have used wave surfer praat all, you can use any software you can use or I can you just go to the Google and typed praat it will praat will be downloaded and store it. Now what session what home tux or you can say what practice I can given that use this recording procedure and record your name let us 5 times record your name in 5 times.

So, that the signal is not clipped and see the highest frequency content of your signal. Next class I will show you using this software how to see the highest frequency content of the signal all those kind of thing and see the how much size it is taken. The think about that if I told you cut a window of this sampled this number sample to this number sample how do you do that in a programming also very important. So, those things you just practice ok.

Thank you.