Digital Voice and Picture Communication

Prof. S. Sengupta

Department of Electronics and Communication Engineering

Indian Institute of Technology Kharagpur

Lecture - 28

Audio Coding: Basic Concepts

Today we start with a new topic and that is the digital audio. And in fact what we are going to talk in this is the basics of audio coding. Now, we have so far seen the coding and communication aspects of speech signals, we had seen for the images, we have also seen for the video. Now what we exactly mean by audio is somewhat different from that of speech; although the broad definition of audio includes speech definitely but when we talk about the audio communication or audio coding and communication there we are in fact also including the high bandwidth audio because as I had mentioned during the speech coding applications that there the speech bandwidth is always assumed to be limited to 3.4 kHz so that the defect to standard for sampling the speech signals was 8 kHz or a rate of 10 kHz which we used to consider for the speech sampling and speech processing application.

In the case of audio the bandwidth that is under consideration is the complete audio frequency spectrum that means to say that which is up to 20 kHz which includes high fidelity audio including the music transmission so where 20 kHz when the bandwidth is kept then we know that the sampling frequency has to be higher than 40 and the international standard that emerged is to use a sampling frequency of 44.1 kHz, that was a sampling rate which was used for the stereo quality audio sampling, for the CD storage application.

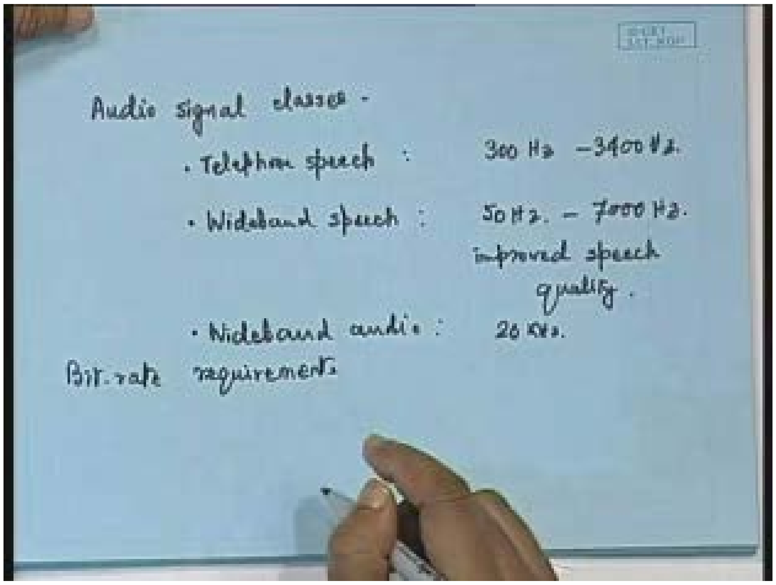
Now we can observe that over the past 15 years there has been a phenomenal growth in the digital audio technology especially with the advent of digital storage medium like the CD ROMs then the DVDs it is always and also we are having the standards already developed for the digital TV for which the audio or rather the digital audio forms an integral part of it so these are some of the major applications major application domains of digital audio and we have seen that over the past 15 years there has been phenomenal growth.

And in fact if we talk about the applications of the digital audio then there are several items that we can put into that list. As an application of digital audio we can mention first of all the digital telephony, then we can talk of satellite broadcasting for the high bandwidth television applications, then the consumer electronics for the CD and DVD quality audio and music. These are some of the application domains. And in fact looking at the different applications of digital audio, one can have a few distinct classes of the audio signals.

So, if we have to talk about the audio signal classes we can mention these few points. The first is of course the telephone speech and telephone speech as you know that because it is a speech application it covers a frequency of 300 Hz to 3400 Hz in the speech frequency range and the intelligibility and the naturalness of the speech that is somewhat poor at 3400 Hz but tolerable because for telephony speech transmission we are not that worried about the quality of the audio; quality in the sense that as far as the distortions are concerned or the naturalness of the voice is not there but as long as the intelligibility is enough we can accept that.

The second audio class is the wideband speech. And for wideband speech the frequency range of coverage is 50 Hz to 7000 Hz and this is for improved speech quality. And the third category is the high fidelity one or rather the wideband audio and for wideband audio the bandwidth as I have already mentioned just now is 20 kHz.

(Refer Slide Time: 7:26)



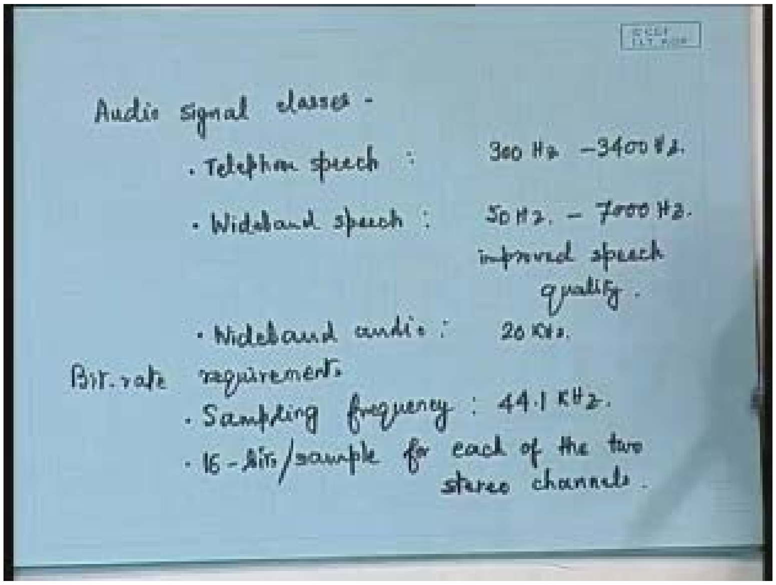
Audio signal classes -

- Telephon speech : 300 Hz -3400 Hz.
- Wideband speech : 50 Hz. - 7000 Hz. improved speech quality.
- Wideband audio : 20 KHz.

Bit.rate requirements

So, taking this wideband audio applications we can calculate the bit rate requirements. In terms of bit rate, actually if we are using, I mean, because the bandwidth is 20 kHz we have to use a sampling frequency which is higher than two times this so we use a sampling frequency of 44.1 kHz that is standard for the digital audio 44.1 kHz and at that rate we use 16 bits per sample and we use stereo quality audio so 16 bits per sample we use for each of the two stereo channels.

(Refer Slide Time: 8:36)



Audio signal classes -

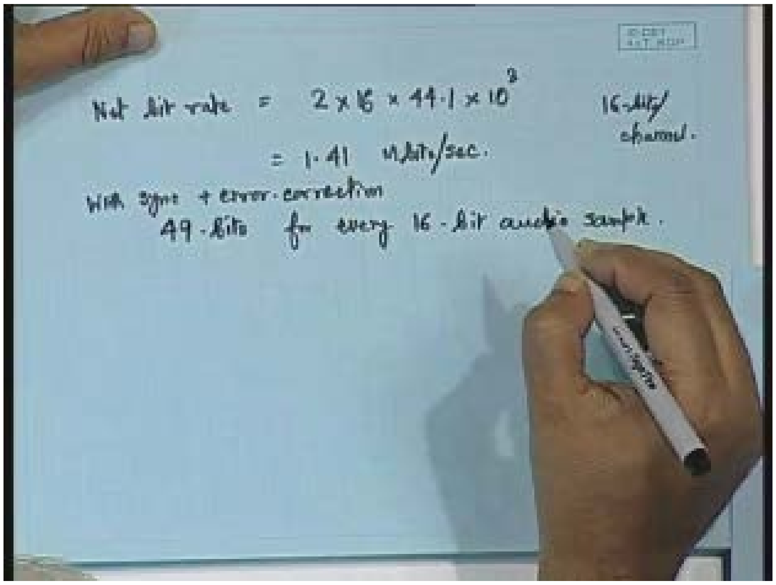
- Telephone speech : 300 Hz -3400 Hz.
- Wideband speech : 50 Hz. - 7000 Hz. improved speech quality.
- Wideband audio : 20 KHz.

Bit-rate requirement.

- Sampling frequency : 44.1 KHz.
- 16-bits/sample for each of the two stereo channels.

Therefore, now if you consider these specifications 44.1 kHz sampling rate, 16 bits per sample per channel and there are two such stereo channels so multiply 2 by 16 by 44.1 we should get the bit rate that will be required for the raw audio transmission; means before any compression is applied and if you find out that, the net bit rate that works out is as follows. So net bit rate is 2 and 2 is because of the two stereo channels each stereo channel requiring 16 bits per sample so 16 and the sampling frequency being 44.1 kHz so 44.1 into 10 to the power 3 and this works out as 1.41 megabits per second 1.41 megabits per second and actually this 16 bits per channel what we are saying, while transmitting we do not transmit 16 bits per channel; in fact what we have to do is to use some synchronization and error correction before transmission and over this 16 bits if we incorporate this synchronization and error correction then we normally have 49 bits. So with synchronization and error correction with sync plus error correction we use 49 bits for every 16-bit audio sample.

(Refer Slide Time: 10:48)



Net bit rate = $2 \times 16 \times 44.1 \times 10^3$ 16-bits/channel.

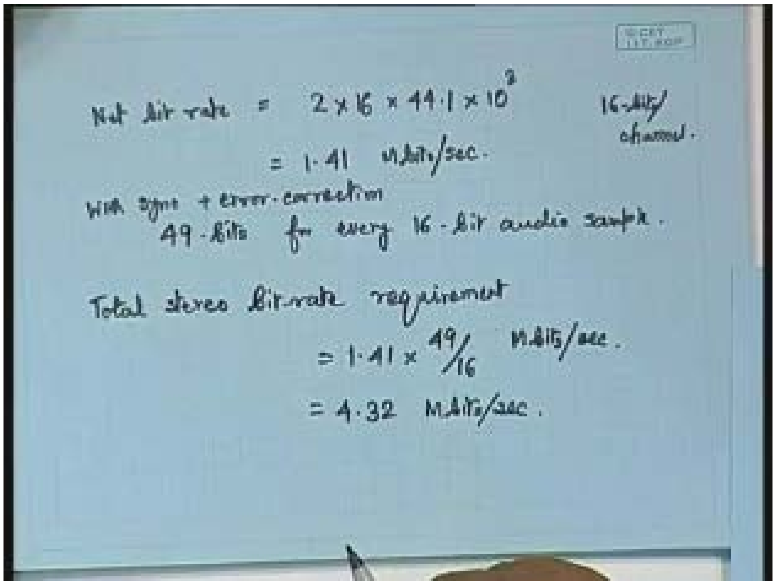
= 1.41 Mbits/sec.

With sync + error-correction

49-bits for every 16-bit audio sample.

Hence, you can be sure that the actual bit rate or the total stereo bit rate requirement should be around three times of this so we should expect....... here it was 1.4 so we can expect in the order of 4.2 megabits per second. Just to make it exact we can see that the total stereo bit rate requirement that should be this 1.41 as we calculated already 1.41 and since there are 49 bits for every 16-bit audio sample then it should be 1.41 into 49 by 16 megabits per second which exactly comes to 4.32 megabits per second.

(Refer Slide Time: 11:43)



Net bit rate = $2 \times 16 \times 44.1 \times 10^3$ 16-bits/channel.

= 1.41 Mbits/sec.

With sync + error-correction

49-bits for every 16-bit audio sample.

Total stereo bit-rate requirement

$= 1.41 \times 49/16$ Mbits/sec.

= 4.32 Mbits/sec.

So again just as in the case of video or in the case of images we had seen that without compression a very large bandwidth requirement is there the same is also the case for audio; although it can appear prima facie okay that the audio's bandwidth is much lower as compare to the video bandwidth but audio sensitivity is again higher and not only that, we cater for the stereo quality sound; especially whenever we are making the full use of digital audio which is supposed to be noiseless the only noise that comes in is the quantization noise and to reduce that to a minimum we go in for 16 bits per sample which is a very fine quantization that we do.

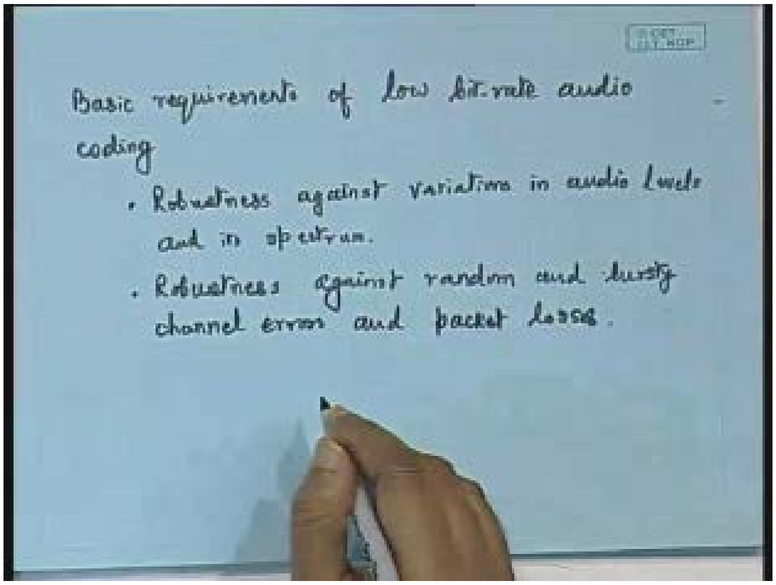
Therefore, considering such high fidelity requirements 16-bit per samples, stereo and a sampling rate of 44.1 to include all the effects of sound all the special effects of sound we require 4.32 which is considerable. That means to say that even for audio also we require good amount of compression.

In fact finally you will be surprised that as far as the bandwidth requirement is concerned we may have to reduce this bit rate to as low as up to 1-bit per sample or for very wide bands speech wide band audio application wideband speech applications we reduce it to 1 to 2 bits per sample so it is a considerable bit rate reduction that one has to achieve and there are some standards which have been already in effect for the digital audio.

But before we talk about the standard specifically let us first highlight some of the basic requirements of low bit rate audio coding. The first requirement is that, you see the audio levels can change as the music goes on or the high quality audio goes on there are many variations in the audio intensity levels, the loudness levels so it should not be that the characteristics of the coder is specifically matched to a particular amplitude and this will get affected when the amplitude changes. So there should be robustness against variations in audio levels because variations will be there so your audio coder should be robust against that; variations in audio level and not only the audio level but also its spectrum audio levels and its spectrum; spectrum means the frequency spectrum.

And then this kind of digital audio we are also using for the mobile communication. So when we are using the digital audio for mobile communication it should be subject it may be subjected to some burst error conditions. So the audio coding must be robust against such burst errors. Any burst error present should not make the encoding process and the decoding process topsy-turvy. It is not that one single presence of burst error disturbs everything that follows. The effect of the burst error should be minimal, minimal in the sense that its extent of damage should be as low as possible. So it is robustness against random and bursty channel errors and packet losses, packet losses also may happen.

(Refer Slide Time: 17:15)



Basic requirements of low bit-rate audio coding

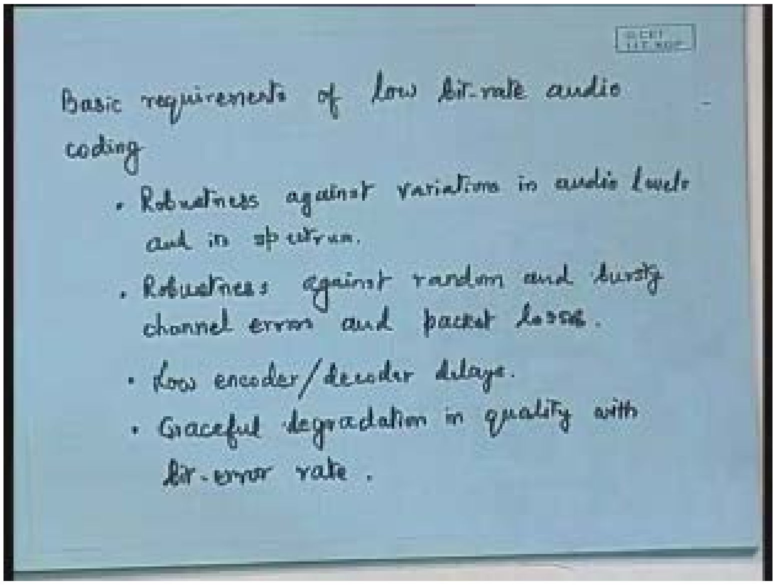
- Robustness against variations in audio levels and in spectrum.
- Robustness against random and bursty channel errors and packet losses.

Then obviously we have the delays in the encoding and the decoding process that should be minimum, why; because when we are going in for some conversational applications some multiparty conversation is going on then it cannot happen that one party is getting the audio signal considerably delayed with respect to the other parties. Because if that is the case in that case there will not be any coherence; means another part you will not be able to make out that when the response actually comes from party number 2 then what was party number 1 saying before or whether it is a reply to party number 1's question or party number 3's question so it is really very difficult to see that is why the basic requirement of low bit rate audio coding should be that it should have as minimum encoder and decoder delays as possible, so low encoder and decoder delays this is another requirement.

And I was mentioning about the kind of burst errors can happen, burst errors or in general I should say channel errors. And the channel errors not only we should talk about the robustness but also the effect of channel errors should not be catastrophic; means without channel errors under ideal situation it is the best form of encoder and in presence of channel errors it is just losing its track. It should not happen. The degradation that results in the quality should always be a graceful degradation.

In fact we always talk about the degradation characteristics to be graceful. So we have graceful degradation of quality with increasing bit error rates. That is why it is a good thing to study, I mean, when you design an audio coder the good thing should be to study its quality as the bit error rate is varied. It should have a smooth variation and a graceful degradation not that up to certain bit error rate it works fine and beyond certain bit error rate it never works.

(Refer Slide Time: 20:00)



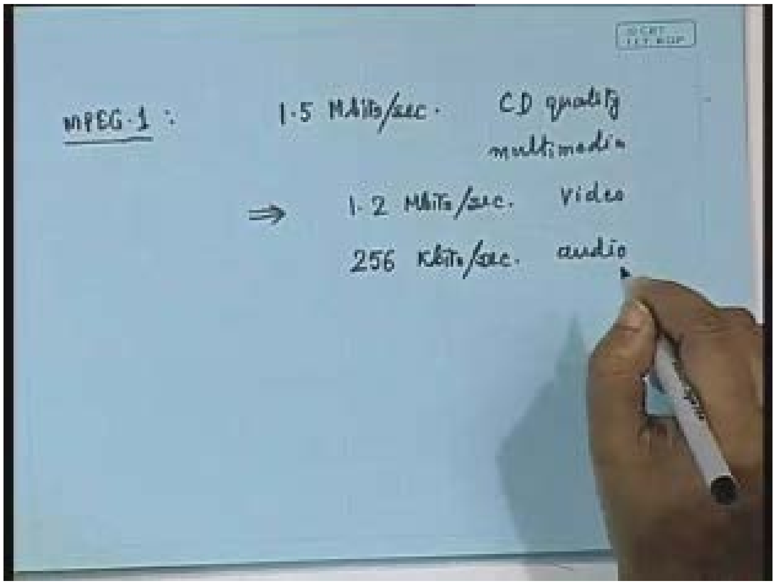
Basic requirements of low bit-rate audio coding

- Robustness against variations in audio levels and in spectrum.
- Robustness against random and bursty channel errors and packet losses.
- Low encoder/decoder delays.
- Graceful degradation in quality with bit-error rate.

So with these requirements in mind the low bit rate audio coding standards were framed. And in fact the standards where framed in different domain. One was the effort that was made by the MPEG community because MPEG is actually a standard for the multimedia communication. Being a multimedia communication standard it has to address both the audio as well as video signals and that is why the MPEG standards also includes the audio coding as a part of it.

So, as the first standard we have the MPEG-1 and as you know that MPEG-1 is designed for the CD quality multimedia storage at a bit rate of 1.5 megabits per second. This is for CD quality multimedia video as well as audio that should be stored at 1.5 megabits per second and out of this 1.5 megabits per second 1.2 megabits per second according to the standard 1.2 megabits per second is allocated to the video and the rest that means to say that roughly 300 kilobits per second to be very precise leaving aside the synchronization bit requirements and the header information that goes into the packet payload-wise we get 256 kilobits per second to the audio.

(Refer Slide Time: 22:07)



MPEG-1 : 1.5 Mbits/sec. CD quality multimedia

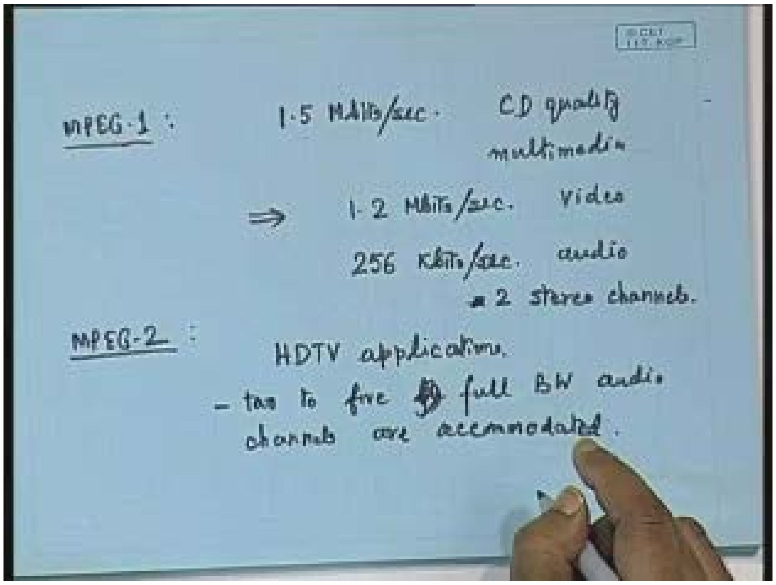
⇒ 1.2 Mbits/sec. video

256 Kbits/sec. audio

And in MPEG-1 it caters for two stereo channels. So, when two stereo channels are accommodated which means to say that every stereo channel gets a bit rate of around 128 kilobits per second so this includes 256 includes two stereo channels so one 128 each so totally 1.456 and with all the other ancillary information like the synchronization and the header information everything put together you reach a bit rate of 1.5 megabits per second which is the standard for MPEG-1.

In MPEG-2 there is an enhancement to this. In fact as I had mentioned about the MPEG-2 while discussing the video coding aspects that MPEG-2 basically caters for the HDTV applications and for HDTV applications the video bandwidth requirements I said is very high and in its audio part in the audio part of MPEG-2 2 to 5 full bandwidth audio channels are accommodated; 2 to 5 full bandwidth audio channels they are accommodated under the MPEG-2.

(Refer Slide Time: 23:58)



MPEG-1 : 1.5 Mbits/sec. CD quality multimedia

⇒ 1.2 Mbits/sec. video

256 Kbits/sec. audio

= 2 stereo channels.

MPEG-2 : HDTV applications.

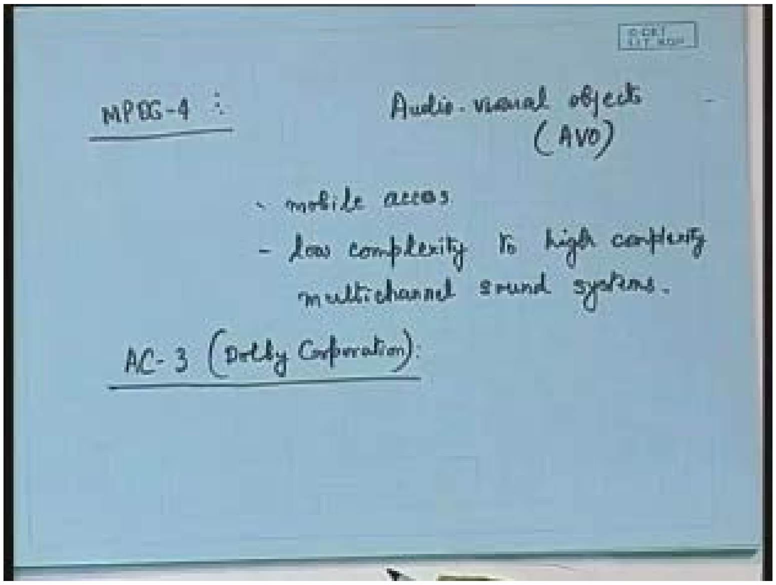
- two to five full BW audio channels are accommodated.

And there are some collection of tools which have been, I mean, certain coding tools that has been specified under MPEG-2 and that comes under the Advanced Audio Coding of MPEG-2 so Advanced Audio Coding or in short form it is called as the AAC. So, when you can refer to MPEG-2's AAC the Advanced Audio Coding this will talk about the collection of audio coding tools that are used. So this is about MPEG-2. So MPEG-2's audio part is an advanced version of what MPEG-1 supports.

And then, also there is further enhancement which is done in the MPEG-4. So, in the MPEG-4 audio, basically MPEG-4 as I also made a brief mention while discussing the video coding that MPEG-4 is a standard that goes in for the object oriented coding and there the encoding is done on the basis of the audio visual objects audio visual objects or what is known in short form as the AVOs. So there the encoding is done on the AVOs and the AVOs may be either a video object or the AVOs may be some audio objects also.

Now this basically has to cater for different application domains like the mobile access, then multi-channel sound systems and all these things; mobile access, then low complexity multimedia terminals, low complexity, I mean, it ranges from low complexity to high complexity multi-channel sounds systems. So this is what MPEG-4 addresses. But we will not go into the depth of each of these MPEG standards. But we will rather make a coverage of one audio coding standard which is in very wide usage especially for the HDTV application. Because when the HDTV standardization effort was made that time the audio coding standard which is known as the AC-3 AC stands for the Audio Coding and it is called as the audio coding version 3 and this was in fact announced by Dolby Corporation. You must have heard about Dolby Corporation because they are the international leaders about the high fidelity audio systems. In fact you will be finding that the major advanced movie theaters and all that they use the Dolby sound system and also the high end musical equipment in the commercial electronics domain they all follow the Dolby quality of sound. So Dolby has adopted this AC-3 standard which we will be talking of.

(Refer Slide Time: 28:23)



MPEG-4 : Audio-visual objects (AVO)

- mobile access
- low complexity to high complexity multichannel sound systems.

AC-3 (Dolby Corporation):

But before coming to the specific aspects of the audio coding standards like the MPEG audio coding or the AC-3 audio coding it is a very much important to talk about certain fundamental aspects of the audio coding.

Now you see, every coding........ of course as I have mentioned that one of the major objectives of this coding will be to achieve a significant amount of compression. Now that challenge we had in everything. Whenever we had speech we faced that challenge and there what we did was that we tackled it using two aspects: one is the inter sample redundancy which we had exploited through the waveform coding of speech signal using the predictive coding of speech and another aspect which was exploited to a very large extent was the vocal tract model. We considered a vocal tract model and based on the estimation of parameters for the vocal tract speech sounds could be or speech signal could be synthesized and it is in the form of a parametric coding, so since there are very few parameters to be encoded that is how we could achieve very significant compression in the speech.

For still images also we could exploit the redundancy in the form of spatial redundancy; that when we consider the adjacent pixels in either of the two directions horizontal or vertical, when we consider the redundancy present in the neighboring pixel that could be exploited to achieve very significant amount of compression and that is what we did by using transform operators and thereafter we had used the energy compaction property of the transforms then we could achieve compression by encoding only those coefficients having very high energy.

Video we could have compression by making use of the inter-frame redundancy also; means the temporal redundancy was exploited in addition to the spatial redundancy.

Now, coming to audio we also must have we must have a property of that that which property of it should give a good amount of compression. Is it just the inter-sample redundancy alone? Perhaps we have to exploit something beyond the inter-sample redundancy.

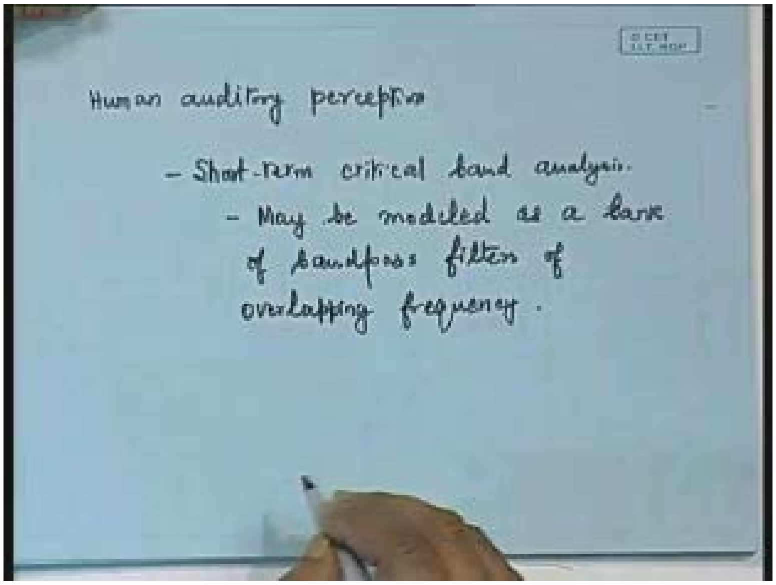
Now in speech to some effect so to some extent we had used inter-sample redundancy the waveform coding but then we also had a strong model on the speech synthesis which we made use of in our speech coding applications.

In the case of audio we cannot have any audio generation model. Whatever we could do for speech the same cannot be done for audio because there is no audio generation model that we can talk of from which we can go in for an audio synthesis. The audio generation process is much more complicated and less understood as compared to the speech generation process where the modeling happens to be much simpler.

Now here, in order to achieve a very significant compression what one makes use of is some psychoacoustic properties psychoacoustic properties and psychoacoustic properties have been studied over the years and if one goes through the findings of this psychoacoustic properties there are some very interesting points which we can note and try to make use of those in the audio encoding process.

Now one of the psychoacoustic property that one finds is that; basically the psychoacoustic studies were first made for the human auditory perception. If you follow the human auditory perception then from the studies it has been found that it performs the human auditory response system that performs short time critical band analysis so the human auditory system performs short term critical band analysis. And in fact what I mean by that is that the human auditory system can be modeled as a bank of band pass filters of overlapping frequencies.

(Refer Slide Time: 34:46)

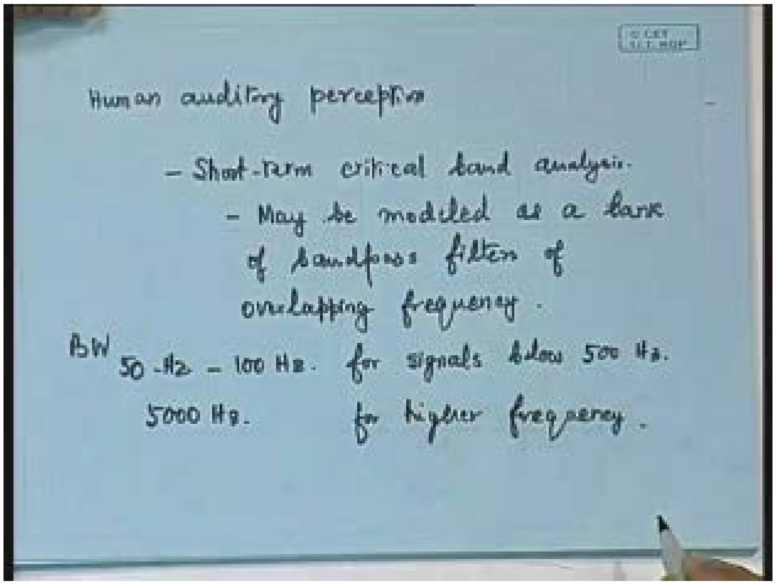


Human auditory perception

- Short-term critical band analysis.
  - May be modeled as a bank of bandpass filters of overlapping frequency.

Now it is found out from different psychoacoustic experiments that these bandwidths that is in the range of... I mean, for all these band pass filters the bandwidths of these filters are in the range of 50 to 100 Hz for............ so the bandwidth is 50 to 100 Hz for signals below 500 Hz and for higher frequencies means of the order of 10 kHz or 20 kHz for higher frequency these bandwidths of this channel the bandwidths of this channel may become as high as 5000 Hz. So as you go over to the higher spectrum higher end of the spectrum you find that the bandwidth of this band pass filters they happen to be quite on the higher side; means if you are taking the total audio spectrum range say this is the frequency and these are the human auditory system responses then you will be finding that there is presence of such band pass filters and all these band pass filters are of overlapping characteristics. So you will be finding that initially it would be like this but as we go in for higher frequencies we will be finding that the bandwidth gets wider and wider so this could be the situation that up to a bandwidth of 20 kHz; on the higher side we will be having wider bandwidth and on the lower frequency side we will be having 50 to 100 Hz of bandwidth.

(Refer Slide Time: 35:41)

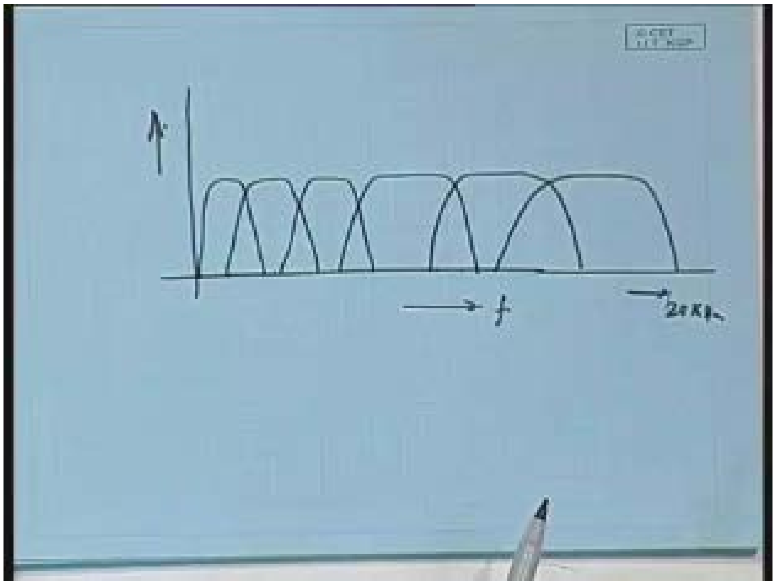


Human auditory perception

- Short-term critical band analysis.
  - May be modeled as a bank of bandpass filters of overlapping frequency.

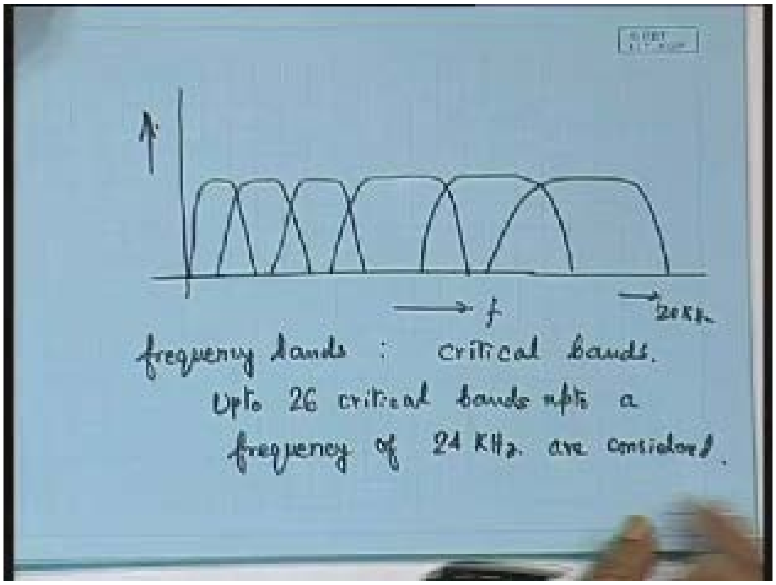
BW 50 Hz – 100 Hz. for signals below 500 Hz.
5000 Hz. for higher frequency.

(Refer Slide Time: 36:45)



So, in effect it is all such type of band pass filters and the frequency bands which we get for these individual filters these frequency bands are referred to as the critical bands these frequency bands are known as the critical bands and up to 26 critical bands up to 26 such critical bands with frequency range 24 kHz are considered. This is the first point to be noted that the human auditory systems being modeled as a bank of band pass filters that is one point to be noted and another very interesting and important aspect which permits us to do considerable amount of audio compression is what is known as the masking phenomenon in audio signals.

(Refer Slide Time: 38:10)

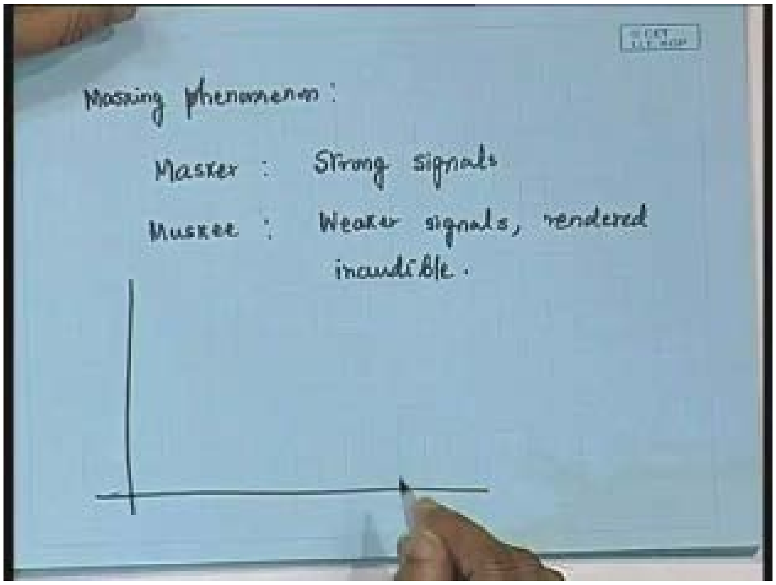


frequency bands : critical bands.
Upto 26 critical bands upto a
frequency of 24 KHz. are considered.

What is meant by masking phenomenon is that if some audio signal is present, at a particular instant if we find the presents of a very strong audio signal at some frequency within the audio frequency range then that strong signal tends to mask off the other audio signals which may be present in and around or in the near vicinity of the strong audio signal which is creating such masking effects which means to say that we have some strong signals which are called as the masker so maskers are the strong signals which cause masking of weaker signals; maskers are the strong signals and the signals that get masked or which are rendered inaudible they are referred to as maskee.

Maskees are the weaker signals in the vicinity of the maskers which are rendered inaudible or you can say that they are masked off. This is a very interesting effect. this is something like this that supposing you first consider that there is no audio signal that is present; means conduct some experiment in absolutely quite environment absolutely noise free sound free environment and then we study a characteristic that what is the audible level of the human auditory system.

(Refer Slide Time: 40:32)



Masking phenomenon :

Masker : Strong signals

Maskee : Weaker signals, rendered inaudible.

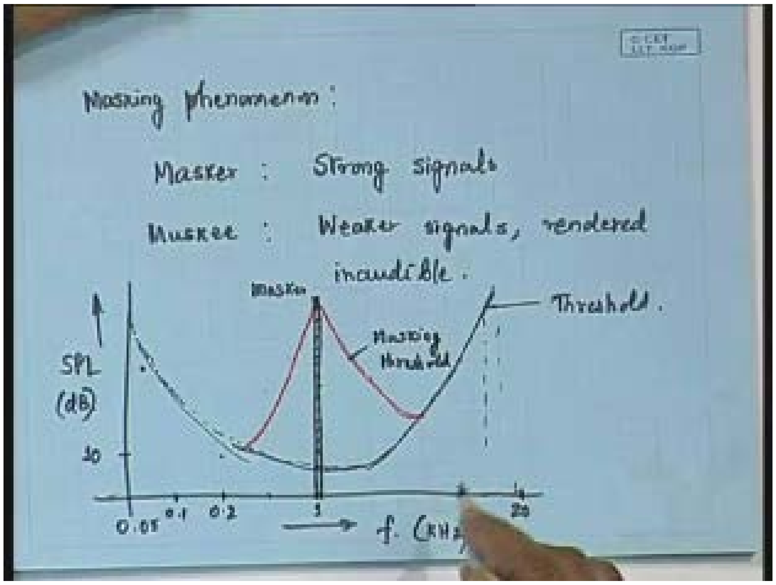
Now we will be finding that if this is the axis that we take for the frequency and here we plot; I mean, on the left hand side we consider the very low frequency like all the frequencies I am plotting in the kHz rate so 0.05 kHz means 50 Hz roughly then the next I am saying as 0.1, the next division is for 0.2 so it is being used in the logarithmic scale up to 20 kHz it is following in a logarithmic scale so the next division will be 0.4 or let's say that here I take 0.5 kHz as follows. Then on this axis on the vertical axis (Refer Slide Time: 41:20) I put the sound pressure level and that is in dB.

Now what we get is that at a very low frequency we require a considerable extent of sound pressure level to make a sound audible. Means I am just drawing a curve like this which you will be able to understand excuse me this is say 20 kHz of sound and here it is 50 Hz so you will see that this is a kind of threshold characteristic so this is the threshold so what I mean to say is that if you have a sound source having this much of sound pressure level at let say 70 Hz or 80 Hz it will be inaudible it has to come above this curve whereas here supposing this level is 10 dB a 0 dB sound at a frequency of 1 kHz may be audible but a 10 dB sound at 200 Hz will not be audible. Hence, this is the kind of a frequency response that we are going to get out of the human auditory system and only the sounds whose sound pressure level crosses this threshold that will be audible.

Now let us suppose that at 1 kHz we have got a very strong signal so at 1 kHz we have got a strong signal that acts as the masker so this is the masker at 1 kHz. Now because of this masker we will be finding that in the vicinity of this the other sound signals will be masked off.

Now, without the effect of masking there is absolutely no difficulty if another sounds source is present at this frequency (Refer Slide Time: 43:57) and another sound source is present at this frequency, they are supposed to be audible. but because of the masking effect what happens is that every masker has got some masking characteristics and the masking characteristic form say masking threshold so it may something like this that we may observe that this line which I have just now drawn with red this is a kind of characteristic which we get for the masking threshold.

(Refer Slide Time: 44:42)



Masking phenomenon:

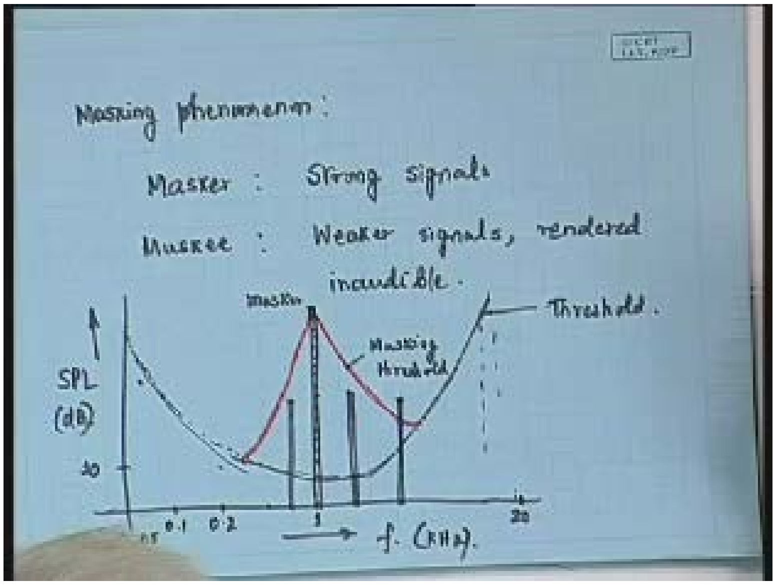
Masker : Strong signals

Maskee : Weaker signals, rendered inaudible.

What it signifies is that, now in presence of this masker any sound which is below this masking threshold will be inaudible. Means if I have a sound source at this frequency having this much of intensity it will not be audible, somewhere here this will not be audible. Normally it would have been audible because it is above the auditory threshold but below the masking threshold if a sound source is present over here (Refer Slide Time: 45:26) with this much of frequency with this much of intensity then it is audible; why, because it is coming above the red line. Thus, in order that it is audible it should be above the red line.

Now here the effect of such masking is only for a range of frequencies around this masker frequency. What we will find is that if the masker frequency in this case it is 1 kHz may be that up to 500 Hz on this side and may be up to 2 kHz on this side or 4 kHz on this side you will find its presence and that too it gradually decreases; means it is very difficult to create another unless we create a sound of higher intensity in the vicinity of this 1 kHz that will not to be audible. It has to be higher than the signal which is present here itself and if that is the case if there is a sound source at 1.1 kHz which is higher than this masker in that case that itself is a strong signal and then that should act as the masker and not this one. So it is always that the stronger signal acts as the masker and the weaker signals with respect to that they are rendered inaudible in the process of masking.

(Refer Slide Time: 46:55)



Masking phenomenon:

Masker : Strong signals

Maskee : Weaker signals, rendered inaudible.

Why this phenomenon is important is that sorry it should be m a s k e e maskee. Now why this is important is that if this is the masking threshold in that case we should see that any noise the noise that is inherently present in the digital audio is the quantizational noise. Now, any quantization noise should be below the masking threshold if you want it to be in audible. So, if you allocate your bits accordingly; see whenever you are encoding an audio sample there you have to decide that how many bits you are allocating to that sample. If you are having more number of bits as you have seen that it is improving..... I mean, the rough figure is that it is 6 dB for every bit for every extra bit of allocation.

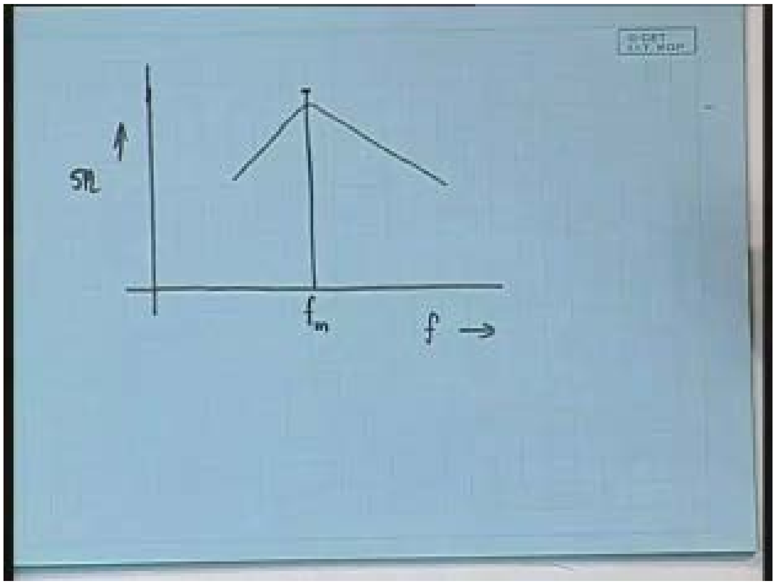
Now the question is that if the masking threshold is this much then by reducing by allocating one more bit I am able to bring down the quantization noise from this one to this one by 6 dB and which is below the masking threshold that should be enough for me. I should not try to make my quantization noise somewhere over here (Refer Slide Time: 48:43). What is the point because a quantization noise of this much is inaudible then why to allocate more bits and waste those bits because you have to be the limiting factor is that you have to come below the masking threshold.

Hence, if there are presence of such maskers take the advantage of that; take the advantage of that in the sense that in presence of such strong masking signals you can afford to keep lower number of bits; because you are massing masking threshold is high so you can afford to keep lower number of bits and can afford to keep your quantization noise sufficiently on the higher side. Whereas say, this much of quantization noise will be inaudible for this frequency (Refer Slide Time: 49:48) but at this frequency if the quantization noise is this much then it gets audible.

So at this frequency you have to bring it down, you have to allocate more bits. So, close to the masking threshold or rather to say close to the masker, closer in frequency to the masker you can afford to have lesser number of bits. This is what one makes use of and in connection with this I would like to mention about some basic definitions of the masking characteristics that how one measures that.

Let us say that this is the frequency and this is the sound pressure level as before (Refer Slide Time: 50:36) only we are just redrawing the diagram and supposing the masking is happening at a particular frequency fm and supposing this is the intensity of the masker signal the masker is having this much of intensity then the masking threshold is seen to be like this.

(Refer Slide Time: 51:04)



Now, one point to be noted that we have already talked of 26 such critical bands in our audio response. Now the pertinent question is that whenever any masker is present does it affect one particular critical band or does it affect more than one critical band. In fact it affects significantly that particular critical band and to a lesser extent to the neighboring critical bands also; the critical bands lying to the left and the critical band lying to the right that also gets affected.

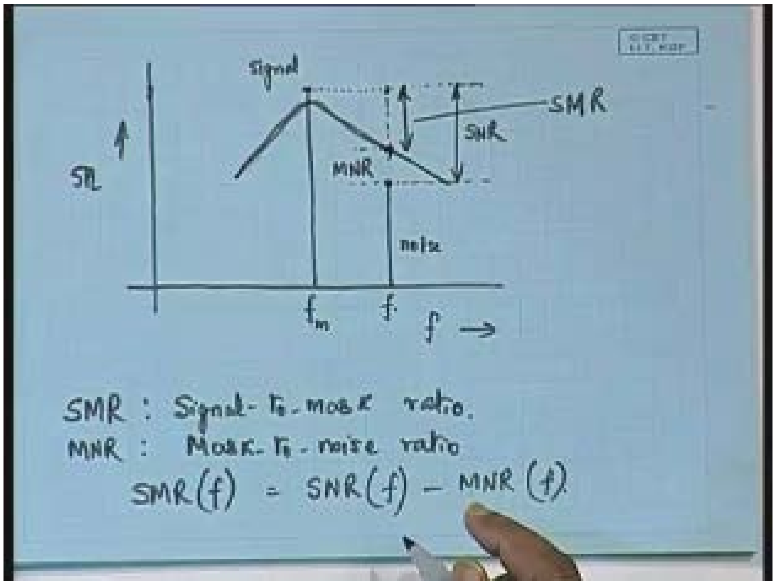
And it is also seen that to the left of this that means to say the lower frequencies generally have a sharper masking threshold falloff as compared to the higher frequency which means to say that lower frequencies are less masked as compared to the higher frequency; means you can see that purposefully I try to keep the slope of the left hand side steeper as compared to the slope on the right hand slope; slope on the right hand side is much flat as compared to the steep slope that I have indicate over here.

Now, supposing we have got a noise source that is present over here. So say this is our noise (Refer Slide Time: 52:41) and the masking the masker that happens to be the signal. So now if we take the difference in the SPL between the signal and the noise at this frequency; let say at a frequency f if you take the difference between this signal level and the noise level, one calls this difference which will be expressed in dB as the signal to noise ratio or the SNR.

Now here you see something interesting that this is the masking threshold that happens at the frequency f. Now if you calculate the difference in height between this signal level and the masking threshold what does that signify? That signifies that your noise should be below this and this particular level what we get as the difference between the signal level and the masking threshold level at a frequency f this is referred to as the signal to mask ratio SMR (Refer Slide Time: 54:00) so SMR stands for the signal to mask ratio and here what you find is that from the masking threshold up to the level of the noise here you have a difference of this much and this difference is referred to as the mask to noise ratio again at the frequency f. So MNR is referred to as the mask to noise ratio.

All these are customarily expressed in the decibel. So this is very clear that the relationship is that the SMR at the frequency f should be expressed as the SNR the signal to noise ratio at the frequency f minus to the mask to noise ratio at the frequency f. So it is SNR minus MNR that gives you the SMR. The significance of the SMR is that whenever we are allocating the bits to the audio samples we must have this SMR considered because we must do the allocation in such a way that the SMR is maximized that means to say that we must keep the noise level below the masking threshold; this curve is the masking threshold (Refer Slide Time: 55:49) so based on that the audio codecs are designed.

(Refer Slide Time: 55:52)



SMR : Signal-to-mask ratio.

MNR : Mask-to-noise ratio

$$SMR(f) = SNR(f) - MNR(f)$$

In the next class we will be studying about how to make use of these psychoacoustic characteristics while designing the perceptual audio codec. We will first present a generalized block diagram of perception based audio codec and then go in for the AC-3 audio codec from the Dolby Corporation, thank you for today.