

Digital Voice and Picture Communication
Prof. S. Sengupta
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kharagpur
Lecture - 15
LPC Vocoder

So today we are going to have our last lecture on the LPC, the linear predictive coding and in particular we will be discussing about the linear predictive coder, Vocoder of course. This coder words is coming twice so it is popularly known as the LPC Vocoder. So rather than telling it by that we will be referring to as the LPC Vocoder so that is going to be our topic of today.

We have already learnt several things about the linear predictive coding and there you have seen that essentially what we are trying to do is to utilize those parameters α_1 to α_K which are our prediction coefficients. Now there are few alternative ways of expressing these predictor coefficients. One thing what one can do is to directly send this α_1 to α_K 's or otherwise it could be sent in some modified form which is also equivalently serving the purpose of the LPC but just a different formulation of the LPC.

Now one of that is by having the roots of the predictor polynomial. Now, if we are looking at the predictor polynomial expression, what results is that we can express the predictor polynomial in this form $A(z)$ is equal to $1 - \sum_{k=1}^p \alpha_k z^{-k}$ and that is going to be the product for $k=1$ to p which we can write it we can express in this form that it is $(1 - z^{-k})$. So what we are essentially doing is that this polynomial expression in z we are expressing as a product of several factors, several means they are p factors, so there are p such predictor coefficients so it could be expressed as a product of p terms where all these individual z^{-k} 's that we have written so the roots of this roots of this predictor polynomial or which is given as the set of z^{-i} where i goes from 1 to p and these are actually an equivalent representation of this $A(z)$.

(Refer Slide Time: 3:48)

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k} = \prod_{k=1}^p (1 - z_k z^{-1})$$

Roots $\{z_i, i=1, 2, \dots, p\}$

Therefore, instead of using the alpha k's we could also try to use the predictor polynomial roots. These are the predictor polynomial roots and we may make use of this and in fact these roots are there in the z plane and instead of expressing the roots in the z plain we can express the roots in the equivalent s plane representation also. So, if we are going to convert from the z plane to the s plane then the transformation that we have to use we know is that z i should be written as e to the power s i into T for z to s transformation, for transforming from the z plane to the s plane.

In this case in the s plane we are going to write it as: s i should be written as the sigma i plus j omega i where we know that essentially omega i will be representing the phase and in fact we can express..... this is the s plane roots so the equivalent s plane roots are going to be this s i corresponding to this z i's and if we are going to write this z i's in this form that z i's will be having a real part and an imaginary part then we can write it as z i real plus the imaginary part of this that is z ii and **this is the** this is with the operator j so z i will be written as z ir plus jz ii and using this one can obtain the omega i as..... what is going to be omega i when z i is equal to e to the power s iT what is going to be the omega i? Omega is going to be 1 by T and then tan inverse? Tan inverse **[Conversation between Student and Professor – Not audible ((00:06:06))]** z ii by z ir so this is going to be tan inverse z ii by z ir so this is going to be the omega i.

(Refer Slide Time: 00:06:21)

Predictor polynomial $A(z) = 1 - \sum_{k=1}^p a_k z^{-k} = \prod_{k=1}^p (1 - z_k z^{-1})$

Roots $\{z_i, i=1, 2, \dots, p\}$

$z_i = e^{s_i T}$

$s_i = \sigma_i + j\Omega_i$

$z_i = z_{ir} + jz_{ii}$

$\Omega_i = \frac{1}{T} \tan^{-1} \left[\frac{z_{ii}}{z_{ir}} \right]$

$z \rightarrow s$ transformation.

Now what is going to be the sigma i?

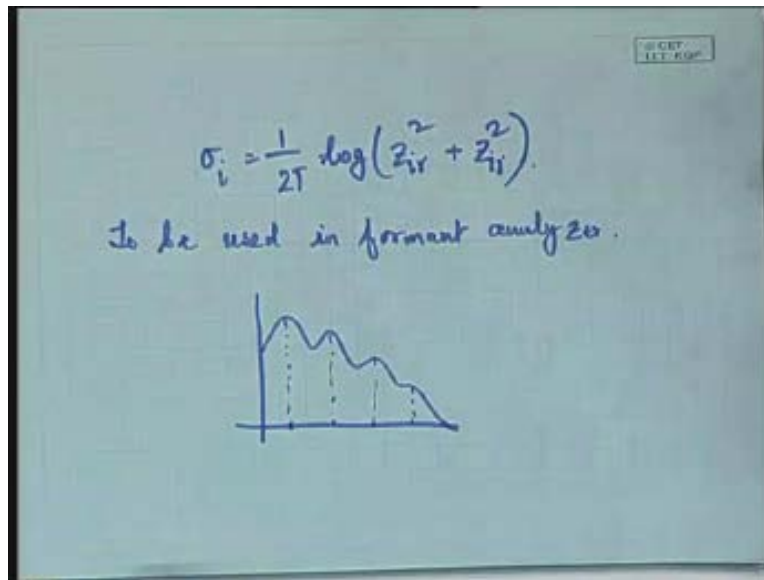
The sigma i will be written as $\frac{1}{2T} \log \left(\frac{z_{ir}^2 + z_{ii}^2}{z_{ii}^2} \right)$. This is going to be the sigma i. So these equations are actually useful for the formant analysis applications of the LPC. So this equivalent form would be used for formant analysis. Now one aspect which we have to consider for a Vocoder realization is that how to obtain the formants. That is going to be a very important aspect and so far we have not..... I mean, although we have talked about the pitch period estimation and the solution of the predictor coefficients but we have not essentially addressed that how to analyze this that how to obtain these formants.

Now essentially in a spectrum in a speech spectrum we are going to have the spectrum like something in this form where there will be existence of four or five different formants would be there. Now how to obtain..... in such a kind of case what is required for us is that, in order to analyze the speech we must be having ideas about the center frequencies of this formants and we should also know about the bandwidth of this formants.

Now how to obtain this? Is there any direct relationship between the LPC coefficients and the formants?

Well, there is of course some kind of an indirect relationship which we can expect. Now, given the predictor coefficients of alpha k's, if we use this expression that A(z) is equal to 1 minus summation k is equal to 1 alpha k z to the power minus k this if we are factorizing in that case this z k's what we obtained as the roots of this polynomial is equivalently representing this alpha k's.

(Refer Slide Time: 9:00)



Thus, the roots of this polynomial if we are obtaining then from these roots it should be possible for us to obtain the central frequency and also the bandwidth of the individual formants and in fact the roots may be obtained quite accurately so we have to, I mean, using this roots we should be able to model the formant. Typically what happens is that if we are having p such predictor coefficients in that case we are going to allow up to p by 2 number of complex roots because the roots in this case will occur in complex conjugate pairs so there will be p by 2 number of complex conjugate roots from this p coefficients then you can know that if there are let us say, if p is chosen to be let us say 6 or 8 something like that in that case we are going to have three or four formant frequencies and the formant bandwidths that can be obtained from this kind of an analysis.

In fact the factorization..... this this basically involves the factorization of the predictor polynomial and this factorization is not too much complicated simply because there is existence of..... only a limited number of p 's are considered because anyway from a prediction point of view we have already discussed this point that beyond 6 or 7 values of past samples are not having any correlation with the present sample so there is no point in having p which is greater than 6 or 7 so we can restrict then this p and if p is restricted in that case the number of roots that is also restricted. So it should be possible for us to factorize and obtain this.

Yet another alternative way of finding out the formants is that what one can do is to have a spectrum of the speech and from this spectrum one can obtain the..... I mean, by a peak picking process one can obtain all the local peaks and using this local peaks one can find out that what is going to be the center frequency of the formants and what is going to be the bandwidths of the formants.

Now that may be somewhat more realistic, although this is model-wise much more simplistic and given that our predictor polynomial factorization is indeed possible we can get the center frequencies quite easily from this alpha case or rather to say from the predictor polynomial itself. Now, one thing which we must note about this kind of an approach is that whenever we are factorizing it like this then essentially this is having an inherent assumption that the system is having all poles so it is essentially considering all pole model; although this aspect is debatable because it is not very well-known that whether all the difference types of speech sounds which are produced..... say for example, some particular cases like the nasal sounds, whether it could be modeled as the all pole model or not is not very clear; in fact there may be some cases where the all pole model is not going to yield a good result but yes, if we are going in by the LPC method of formant analysis in that case we have to accept the all pole model but otherwise it is quite a simplified model.

Whereas if we are going in by the direct spectral approach in that case we are not as such accepting any mathematical model but we have to go in for a more complicated process like obtaining the spectrum which will involve taking the Fourier transforms or any other spectrum determining techniques and then the peak detection that also we have to use whereas if you are

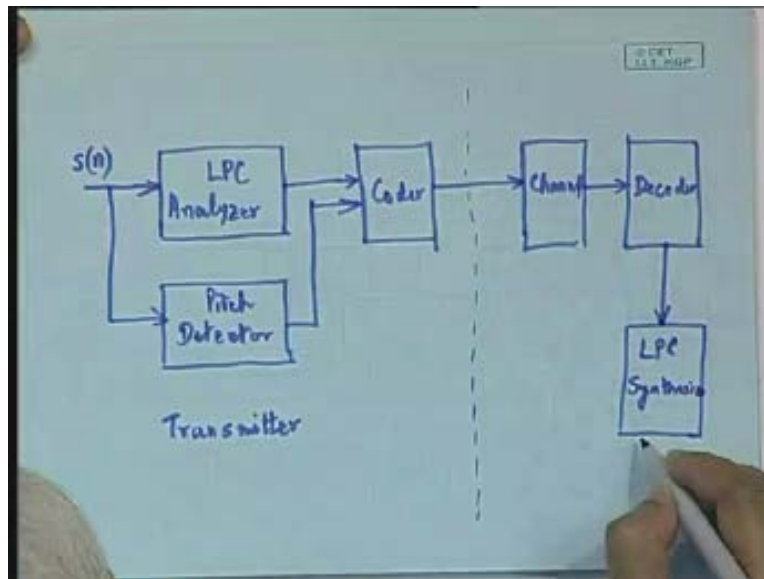
anyway having the predictor coefficients out of this LPC you can have it to implement the formant frequencies relatively easily.

Now we are in a position to present the LPC Vocoder structure. All the individual blocks have been discussed so what we are going to do is simply to place those blocks one after the other and realize an LPC Vocoder. So what it is going to contain is that at the input we are going to feed the speech segment which we are going to write as s of n and then we are going to have the LPC analyzer.

What this LPC analyzer is going to do is that it will take all this α_1 to α_k 's or its equivalent parametric forms because rather than the α_1 to α_k 's one can also send the roots of the predictor polynomial. It is just one and the same of equivalent representation and also this block is going to yield us some ancillary information, not this but rather the pitch detector block because there is yet another processing which will do in parallel and that is the pitch detector and this pitch detector will not only estimate the pitch frequency but also it will generate some information like whether it is the voiced segment or the unvoiced segment and then it will also tell us about the estimated gain. So the gain estimation and the voiced/unvoiced flag that will be from this pitch detector whereas the LPC analyzer will send the coefficients.

Now everything, the pitch detector output as well as the LPC analyzer output that has to be properly coded. So what we need to do is to have a coder after this. So up to this part we will be having the transmitter. This will compose the transmitter part and then we are going to have the channel. This will be followed by the channel and after the channel we are going to have the decoder and I am drawing it this way since I can't go any further on that side that is what we use that after the decoder we are going to have the LPC synthesizer.

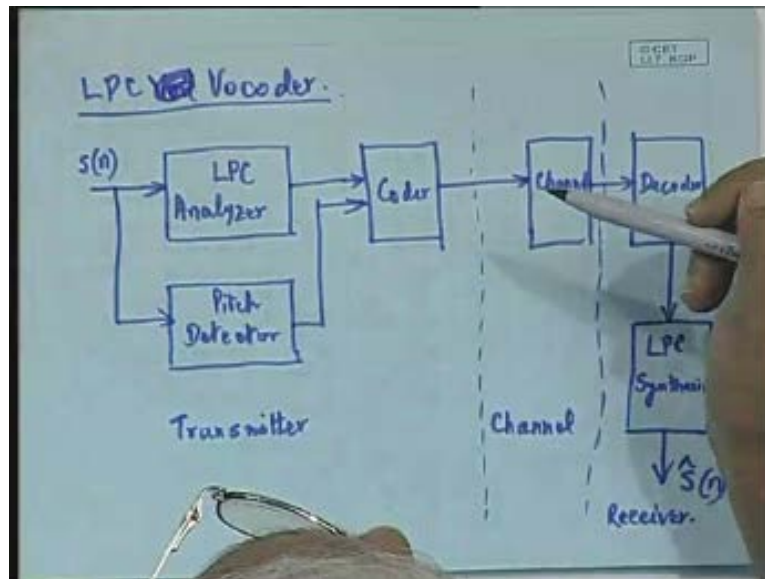
(Refer Slide Time: 17:43)



LPC synthesizer we already discussed in the last class. LPC analyzer and also LPC synthesizer both we already discussed and the LPC synthesizer output is going to be s_{cap} of n . Here it is s of n and LPC synthesizer output is the s_{cap} of n . So here we are going to have..... this is the channel part (Refer Slide Time: 18:02) and on this side we are having the receiver, so together this composes the LPC Vocoder. It is quite similar to the other Vocoders that we have discussed. We have already discussed the channel Vocoders. There, you remember that what exactly we did was to analyze the signal into, I mean, there also we had a kind of analysis filter bank and it was analyzed into different filter banks I mean, 10 or 12 filter banks we had used and then we also had the same thing that is to say the pitch detector was required there and then we encoded everything.

The idea was very similar, but the only thing is that the bank of filters that is being replaced in this case by the LPC analyzer. It was actually a frequency domain analysis whereas the LPC analyzer is going to give us a time domain analysis. So otherwise it is the same but LPC Vocoders are seen to be highly efficient; or rather to say in terms of compression efficiency of speech signals it is seen to be really very efficient and we can find out, we can have a rough calculation as to what bit rate it is going to generate.

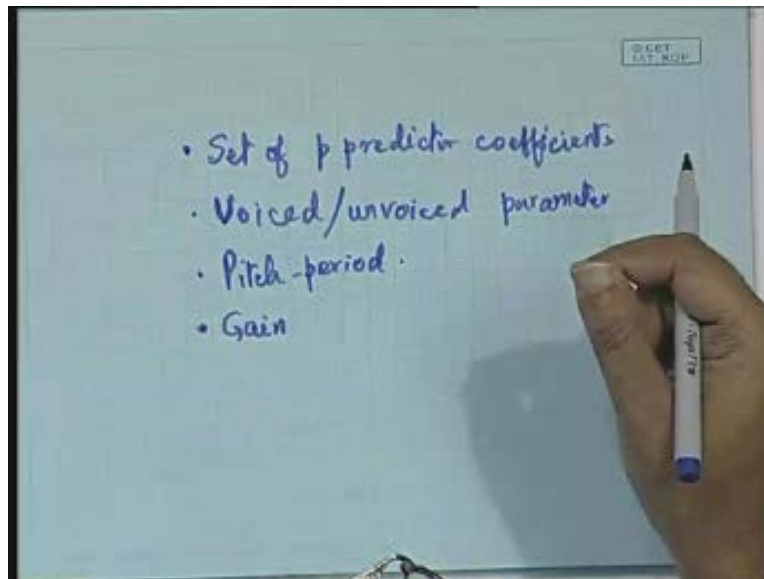
(Refer Slide Time: 18:17)



Now, in order to do that let us see that what are the different parameters that we are required to send in the channel as the coded information. One is that what are the basic LPC analysis parameters?

The first is the set of p predictor coefficients, then there is a voiced/unvoiced parameter **voiced/unvoiced parameter** then we also have to send the pitch period and then we also have to send the gain information.

(Refer Slide Time: 20:53)



Now for all these things we require bits. Voiced/unvoiced parameter is only a flag. So it is just a 1 bit information and for gain it is seen that **total of** total of about 5 bits is good enough so this is 1 bit, this is 5 bits, 5 bits distributed on a logarithmic scale, 5 bits which is distributed on a logarithmic scale that is seen to be sufficient. The pitch period that is also expressed in a quantized form and they are also normally 6 bits of quantization in the pitch period is generally considered to be okay so that really takes us to 6 plus 1 plus 5 12 bits already but we have not yet sent the predictor coefficients.

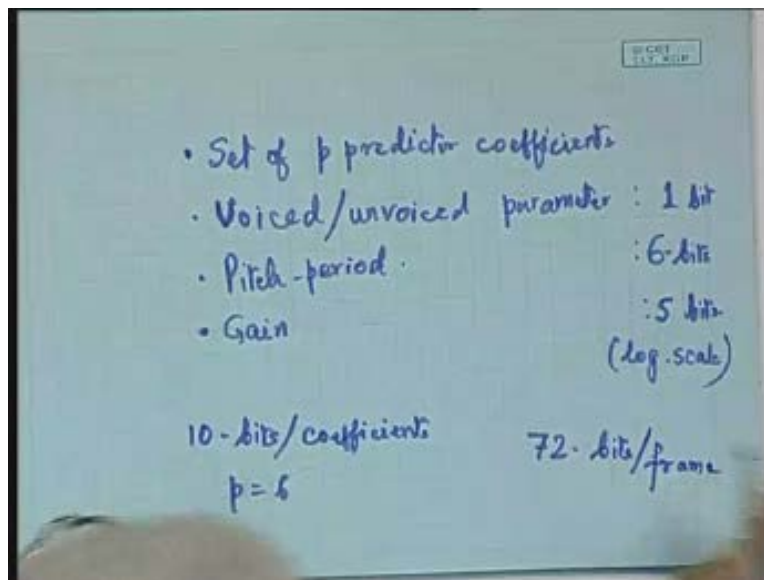
Now **how many predictor so** how many bits are really necessary for predictor coefficients?

One of the approaches that one can take is to directly quantize these predictor coefficients. But that approach is not very welcome because you know, in the process of truncating the predictor coefficients may lead to certain problems like the predictor coefficients may alter the position, I mean, any quantization error could result in a more sensitivity to this location of the pole positions; if you are factorizing this, if you are factorizing the predictor polynomial and obtaining the roots of the predictor polynomial and calculating the center frequency and the bandwidth, this center frequency and bandwidth will be very much sensitive; the roots and the calculation of the center frequency and bandwidth will be very much sensitive to any deviation

of the predictor coefficients. So it should not be very abruptly truncated and in fact a very good precision has to be used.

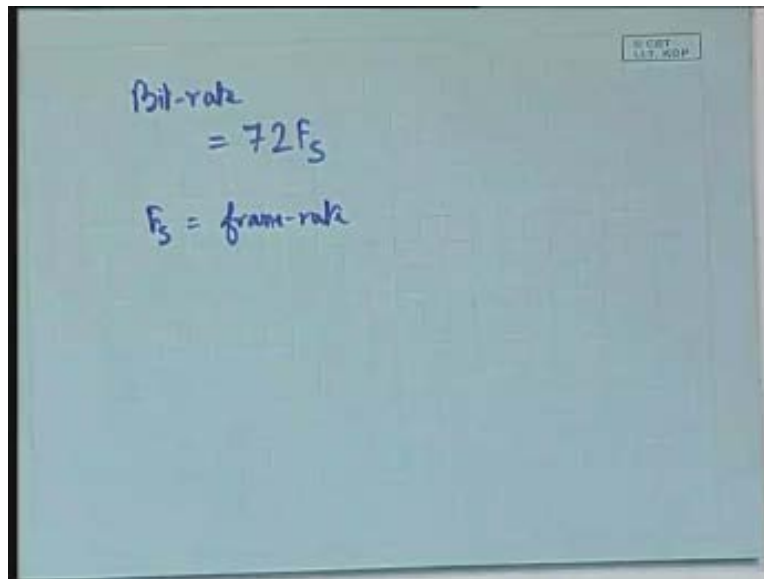
One cannot simply say that, just in order to reduce the bits it will be just truncating the predictor coefficients to 3 bits per coefficient or 4 bits per coefficient; it is seen that in order to preserve a good amount of accuracy at least 10 bits per coefficient is required **10 bits per coefficients** and in order to have a formant analysis as we were just now discussing that we must have at least three formant frequencies with us. So three formant frequencies would mean that we must have up to p equal to 6; beyond p is equal to 6 is not really very much meaningful. So p is equal to 6, so if you take p is equal to 6 that means to say six coefficients and use 10 bits per coefficients then that makes 10 into 6 that is 60 so 60 plus this 12 that makes 72 bits so totally we are adding up to 72 bits of information which must go in one frame. So there will be 72 bits in one frame.

(Refer Slide Time: 24:37)



Now let us see that what is going to be our frame rate. If we are having a frame rate of let say F_s in that case we are going to have $72 F_s$ to be our bit rate. So our bit rate is going to be $72 F_s$ where F_s is going to be the frame rate.

(Refer Slide Time: 25:05)



A photograph of a whiteboard with handwritten text. The text is written in black marker. In the top right corner, there is a small rectangular stamp that reads "SECRET" above "117.40P". The main text consists of two lines: "Bit-rate" followed by "= 72F_S" on the next line, and "F_S = frame-rate" on the line below that.

$$\text{Bit-rate} = 72F_S$$
$$F_S = \text{frame-rate}$$

Now, a typical frame rate is seen to be of the order of 100 or 67 or 33. These are some of the typical values that one uses for F_S . So if you are taking the lowest value of the frame rate let us say 33 frames per second if you are taking, in that case we are going to have a bit rate that is equal to 72 into 33 and that leads to 2400 bits per second. Using 67 it leads to 4800 bits per second and using 100 it is simply 7200 bits per second. So this is with F_S is equal to 100, this is with F_S is equal to 33 (refer Slide Time: 26:01) and this is for F_S is equal to 67.

(Refer Slide Time: 27:21)

Bit-rate
= $72F_s$

$F_s = \text{frame-rate}$

$F_s = 100, 67, 33$

Bit-rate = 2400 Aps	$F_s = 33$
4800 Aps	$F_s = 67$
7200 Aps	$F_s = 100$

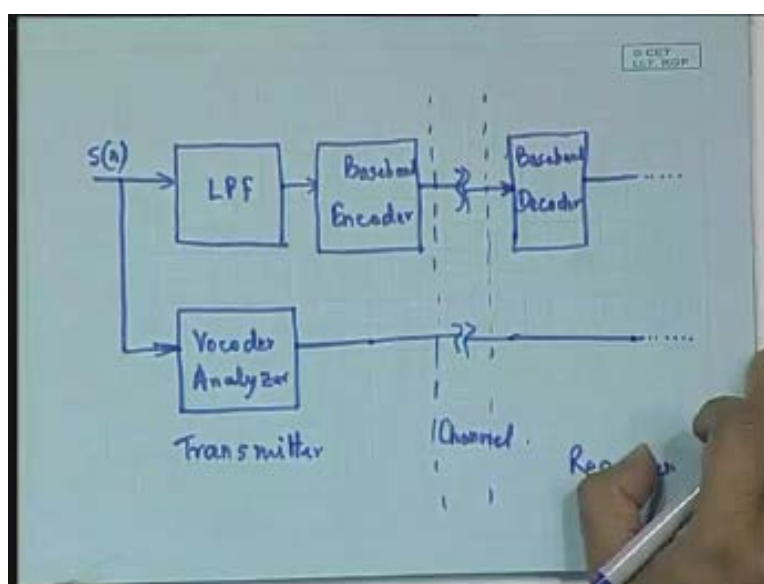
Therefore, you can see that if we are having a bit rate like this, just compare that against the normal speech waveform, there we are having..... basically if we are having a rate of 10 kHz of sampling rate if we assume then we are going to have a..... so with 10 kHz of sampling rate and with 8 bits per sample what does it lead to, it leads to at least 64 kilobits per second, so as against that we can have it in 2400 bits per second; 2.4 kilobits per second only.

So naturally it is a great saving in terms of the number of bits. So LPC Vocoders are really very useful for very low bit rate applications and in fact it is used for the telephone quality speech signals. This LPC Vocoders are really a very good choice. And one of the aspects that should be considered about the Vocoder realization; of course when we talked about this pitch period estimation and all (Refer Slide Time: 27:40) then we used it in our LPC synthesizer because LPC synthesizer essentially makes use of the pitch parameters which are obtained over there. So, essentially, good pitch estimation is a requirement for us in order to have a proper synthesis. And unless the pitch parameter estimation is improper the performance of the Vocoders really suffers. So that just leads to conclude that really speaking the pitch parameter estimation or the pitch period estimation is normally the weakest link in an LPC Vocoder.

Now, if we can rather think of some alternative methodology, in order to have the pitch period estimation or rather avoiding the pitch period estimation explicitly using the voice signal itself, that has led to what is called as the voice excited Vocoder; voice excited LPC Vocoder where no explicit pitch detection will be performed but what we will be doing is something like this. So let us follow the block diagram and then we will be discussing the different aspects of it.

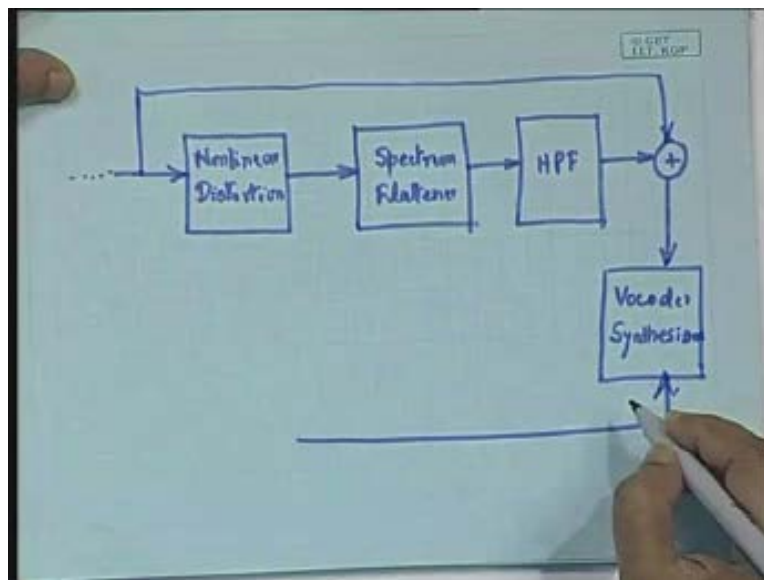
Hence, the input to this is $s(n)$ and then we are going to have a low pass filter and then we are going to have a baseband encoder and then we are going to have a nonlinear. So yes, we will come to the other things later and then parallelly we are going to have another channel which we are calling as the Vocoder analyzer. The Vocoder analyzer will basically extract the α_1 to α_p 's and we do not have the pitch estimator but instead it is the low pass filter version of the voiced signal itself so this also goes into the channels. There are two independent things which are going into the channel: one is the encoded low pass filtered signal and the other is all the Vocoder parameters that have been extracted. They are going independently into the channel so this side is going to be our transmitter then we are going to have the channel over here and then this is where we are going to have the receiver; next to the channel we are going to draw the receiver (Refer Slide Time: 31:16).

(Refer Slide Time: 31:54)



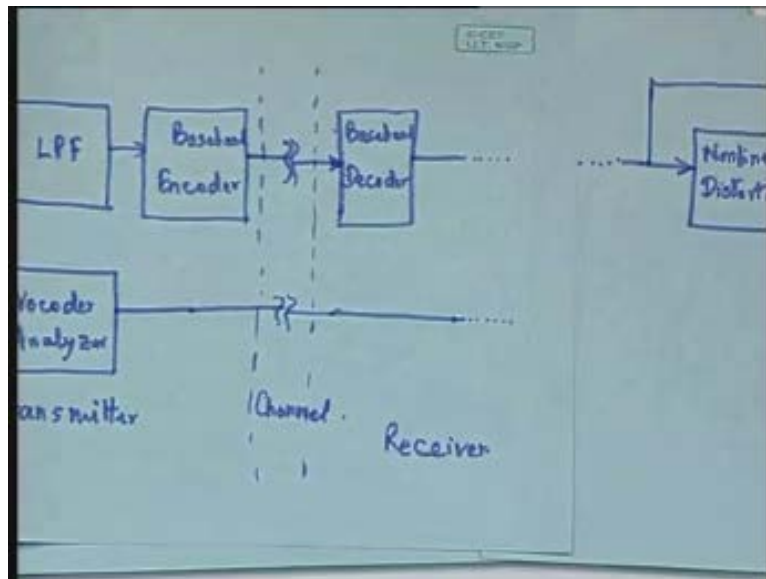
So the first block that we can expect in the receiver is the corresponding decoder of this. So there is going to be a baseband decoder. Then we are going to have..... I will continue in a different sheet because otherwise I do not have the space to draw everything..... so we will continue beyond this part and here the Vocoder analysis part also is coming to the receiver. So, at the receiver beyond this baseband decoder we are going to have a nonlinear distortion compensation, so there will be a nonlinear distortion compensation and then we are going to have a spectrum flattener, I will explain that why we need such a kind of a spectrum flattener and then we are going to have a high pass filter and then with this version we are going to add the decoded signals. So this is the decoded so this is from the decoder from the baseband decoder (Refer Slide Time: 33:03) and the decoded signal is added to this and then this added signal is going to the Vocoder synthesizer block and the Vocoder synthesizer block also receives the input from the Vocoder analysis.

(Refer Slide Time: 33:38)



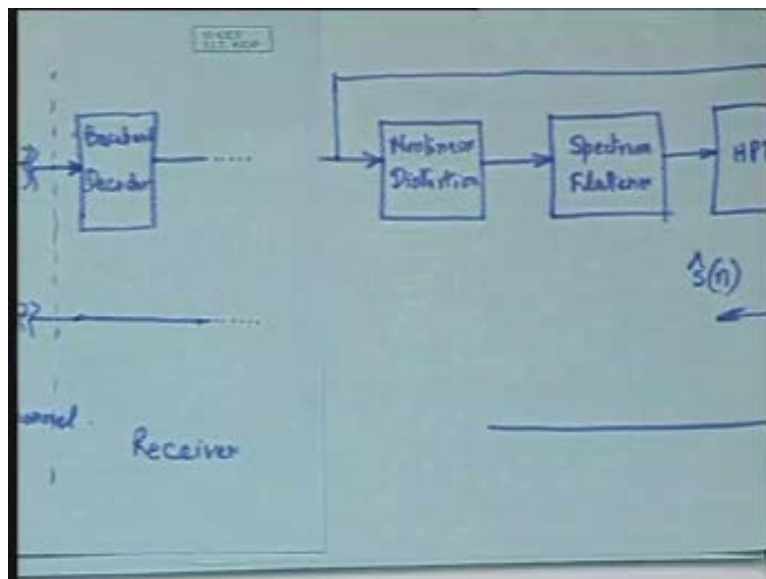
Therefore, all the parameters of the Vocoder are coming into the Vocoder synthesizer plus it is getting the signal in a processed form. What sort of processing is that we will be discussing shortly. And the output of the Vocoder synthesizer is going to be s cap n. So this whole thing what I have drawn in these two pages it is as if to say it is a continuation of this.

(Refer Slide Time: 34:11)



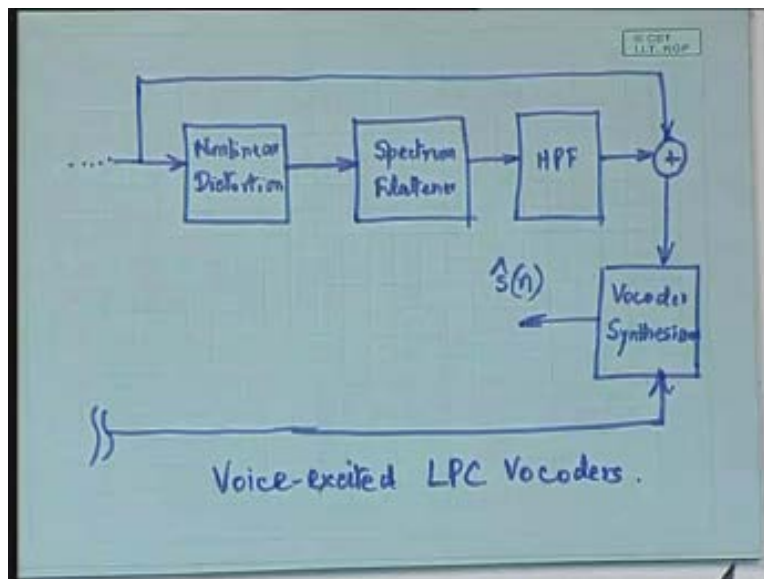
What we had drawn as the transmitter channel and then it is this receiver part that continues.

(Refer Slide Time: 34:18)



This whole thing, this entire block diagram we will be referring to as the voice-excited LPC Vocoders.

(Refer Slide Time: 34:43)



Now, essentially what major difference have we done?

You see that here we are having a low pass filter (Refer Slide Time: 34:54). Can anybody guess the purpose of the low pass filter? [Conversation between Student and Professor – Not audible ((00:35:02))] Correct, the higher formants are getting filtered out because we do not require the higher formants. Yesterday we had already discussed this aspect in our last class where the higher formants need to be filtered out that is why a low pass filter of the order of 1 kHz etc that really filters out the higher formants, it is generally used, because the whole idea should be that, from a low pass filtered version of the signal it should be possible that; or rather the low pass filter version of signal will give a better idea about the pitch period; although we are not explicitly computing the pitch period, there is no such autocorrelation and all these things which are obtained, it is simply the low pass filtered version of the signals where the formants are eliminated and it is preserving all the peaks that correspond to the pitch period and we are encoding that low pass filtered version of the signal directly. And what we are doing at the decoder end is that we are..... first of all the nonlinear distortion compensator is used just to take care of whatever nonlinear distortions are introduced in the speech generation process because that model we did not really discuss which is the radiation of speech that means to say that from the vocal tract ultimately when it goes to the resonator or rather the mouth cavity and

then it goes out, there are some nonlinear distortion effects which take place so that is compensated using this nonlinear distortion compensation and it is this spectrum flattening which basically eliminates the spectrum all the peaks which are there and then this signal the high pass filtered version of this signal this is added to the Vocoder's input.

So the Vocoder is getting two inputs basically: one is the Vocoder analyzer parameters or rather to say the Vocoder coefficients and then it is also getting the filtered version of the signal. In other words, the low pass filtered version of the signal is coming to the Vocoder synthesizer and then it is generating the s_{cap} of n or rather the synthesized signal.

Now in this case, of course, **this makes the pitch period estimation** this avoids the pitch period estimation no doubt and performance also will be better because it is always triggered by the speech waveform itself. The speech waveform itself is deciding the excitation; we are not having any separate impulse generator which is triggered by the estimated pitch period, rather than that it is the voiced signal itself the processed voice signal itself which is acting as the source of the excitation. So essentially this is having some kind of a better performance but **what we actually** we are also paying the price somewhere because what we are essentially doing is that here the low pass filtered version of the signal is completely encoded through the baseband so we require a speech encoder, so we require a kind of a waveform encoder over here. So you see that it is a combination of waveform encoder plus the Vocoder.

Now Vocoder will require certain bit rates; so we have already discussed that even if you leave aside the pitch period estimations and all which will consume less number of bits but **at least 10 to 12** at least 6 or 7 values of the predictor coefficients itself is ultimately leading to a bit rate which could be in the order of 2400 bits per second or higher and the encoding of the low pass filter will also require some extra bits so that at the minimum end you can expect that such kind of voice excited LPC Vocoder are going to have an increased bit rate of the order of 1 to 2 kilobits per second, 1 to 2000, 1 to 2 kilobits per second which means to say that it may be of the order of 4 kilobits per second, 4000 bits per second or I mean, between 4000 to 5000 bits per second as compared to 2400 bits per second.

Hence, naturally bit rate-wise it is really getting inefficient as compared to the normal L LPC Vocoder. But yes, performance-wise it is really picking up the variations that happens in the pitch signal and that really makes the speech synthesis process more realistic; rather to say that on the alternative side, whenever we are using the LPC Vocoder and the conventional pitch period estimation then such differences in the pitch period estimation or rather to say the errors in the pitch period estimation makes the speech generation process or the speech synthesis process quite artificial. So the voice that we hear **is having lot of** lacks the naturalness. But here it is going to be much better.

This is about the LPC Vocoder and I would like to answer some questions at this stage. If you have any doubts pertaining to the entire aspects of the LPC linear predictive coding and the realization of the Vocoders please feel free to ask me at this stage. Any questions? So **this is about** with this we come to the end of the linear predictive coding chapter. We have really gone through a major part of the speech processing because **we have** first of all we have done the vocal tract modeling and understood the basic speech generation and speech modeling process, understood that what is the formant and what are the speech parameters that one requires. Then we also went in for a detailed discussion about the waveform based coding and waveform based coding we started with the delta modulator, **we started** we also discussed about the differential modulator and also the adaptive coding, the adaptive prediction, then finally the ADPCM. Thus, the speech encoding philosophy is quite well developed and in fact at a later stage we are also going to discuss about the speech coding standards which are in existence of today. But for this lecture, thank you so much.