Real-Time Digital Signal Processing

Prof. Rathna G N

Department of Electrical Engineering

Indian Institute of Science - Bengaluru

Lecture - 44

Speech Coding II

Welcome back to real time digital signal processing course. So, last class we discussed about little on speech coding. Today we will continue on that regarding the O coders. So, this is the second lecture what we will have it. So, we saw that although I did not discuss the A law companding. So, which was in the analog domain.

So, you can go back and then check the things. So, we discussed little on mu law. And then little on the subband coding how we can bring down from PCM that is 64 kilobit per second to subband coding of 31.5 kilobits per second rate.

So, today we will see about the O coders. So, we say it is parametric coders basically and models the vocalization of speech. And, the speech sampled and broken into frames approximately we break it up into 25 millisecond and instead of transmitting digitized speech we are going to build the model of speech. and the transmit the parameters of the model and synthesize approximation of speech. So, we say that the first one is the linear predictive coders in this category LPC.

So, the basic coder model what we are going to have it. So, models the vocal tract as a filter and then filter excitation what it is going to be done. So, how it is going to be done? It is periodic pulse that is voice speech or noise we call it as unvoiced speech. So, using that we will do the filter excitation and then synthetically generate the speech at the receiving end. So, the transmitted parameters are gain, voiced or unvoiced decision So, if it is voice we have to send the pitch and the rest of it is LPC parameters what we have used in the coding techniques.

So, coming to the second part of our coders we as we said it is the linear predictive coders we are going to use it. So, the first is the excitation stage. So, what we are going to have is periodic pulse that is voice speech or noise we call it as unvoiced speech. So, this is how the pulse generator is going to be and we are going to provide a pitch period in this case. So, we call this is the voiced generation and here it is unvoiced we say it as noise generator.

So, this is our excitation model what we have it then after the excitation model ah. So, we have to select the excitation. So, that is going to happen here and we will be having the gain factor g. So, which is multiplied by this one ok. Then these are the transmitted parameters ah which are the ones we are going to what gain we have used it.

So, whether it is voiced or unvoiced decision will be putting it and then the pitch if it is a voice signal and the other ones are LPC parameters what will be transmitting it. So, here you will be seeing that this is the excitation model output is mu of t. and how we are going to see the LPC basically. These are the predictor coefficients. So, we will be putting the linear feedback and do the summation of the two of it and we will be sending y of t along with the parameters of LPC these are the parameters.

So, this is our vocal tract model what we are going to generate. And, then will be passing the LPC parameters to the through the channel to receiver place. So, coming back what is it example 10th order linear predictive coder what the example shows here. So, the samples voice at 8 kilohertz and we are going to have a buffer of 240 samples that is we are taking 30 millisecond of the data. And here the filter model is M is 10 is the order of the filter and gain considered is G and we know that z^{-1} is the unit delay we are going to have it and b_k are our filter coefficients what we will be using it in the system. So, then what happens to our impulse response $H(z) = \frac{G}{1+\sum_{k=1}^{M} b_k z^{-k}}$ then our G is going to be 5 bits. And then b_k is going to be are 8 bits each what we can use for the voiced and unvoiced decision we are going to have 1 bit and pitch is represented with 6 bits. So, this implies that we have 92 bits per 30 millisecond. So, that is band width per second what we are going to get the thing.

So, this is the model of the LPC coding the previous one we discussed about this. This is the model which is going to be used in the LPC. This is bits per second what we are going to have what we have to transmit totally. what we have is can achieve low bit rates that is from 1.2 to 4.8 kilo bit per second. So, the characteristics of LPC coding is quality is very low and complexity in this case is moderate and bit rate as we have seen that it is low and delay in this case is moderate and robustness is also low. So, quality of pure LPC vocoder to low for cellular telephony it try to improve quality by using hybrid coders basically. So, how to improve the quality by refining model of speech and improve accuracy of the model and improve input to speech coder. So, by using this we can modify our LPC vocoders.

So, how is it you can see that we can use the hybrid coders that is combined vocoder and waveform coder concept basically. And we will be using the LPC modified one we call it as residual LPC that is rel basically. And we can use the codebook excited LPC that is we call it as kelp we will see in a while how we are going to use it. So, you can see that this is again 2 which is deferred. So, you will be seeing that this is the voiced portion and this is the unvoiced region and then how the pitch is generated what you can see the thing this is the frame and then this is the pitch period what we call it on the y axis.

So, this is how we will be generating the pitch and then transmitting in our LPC. So, what how does the this thing RELP-OCODER looks like that is residual excited LPC to improve the quality of LPC by transmitting error that is we call it as residue along with the LPC parameters. So, you have the this is the input signal which is going to be buffered or windowed here and then you will be checking that it is voiced or unvoiced decision gain and then pitch which are getting transmitted. So, this is short time LP analysis what we are going to see that do that. And, then which is going into the LPC synthesis here and these are the this thing LP parameters which generated from here is going to be transmitted and from the synthesis you are going to subtract from this original and then we will be checking the residue.

So, what is the error we have got it or the residue which also is going to be subtracted. encoded along with the parameters and then the rest of the thing whether it is voiced or unvoiced, gain and then pitch. So, these are the three inputs along with in the LPC only these two are going. Now, even the residue is going to be put into the encoder and this encoded output is sent to the receiver. So, the other one is we call it as GSM speech coding basically, how does it look like basically GSM we know all of us use the thing uses a regular pulse excited linear predictive coder that is RPLPC for speech.

So, basically what does it have combined the DPCM we discussed in the last class concept with LPC. So, information from previous samples used to predict the current sample and LPC coefficients plus an encoded form of the residual that is predicted minus actual sample equal to error what it is going to be transmitted. and the represent the signal basically which are transmitted. So, as you can see here this is the analog speech. So, which is going to be low pass filtered and then you will be converting it into analog to digital form.

So, the sampling rate used is 8 kilohertz and in this case A to D converter has 13 bit bits per sample what it is going to be represented. So, what you will be getting at the output will be 104 kilo bit per second samples and then here it is going to be RPA LTP speech encoder is going to be put in and the output what we will be having is 13 kilo bit per second what it is represented in the GSM module. So, this is going with our to the input to the channel encoder. So, how the speech coding is going to happen we will see in a while. So, we said regular pulse excited long term prediction what we can have the thing.

So, that is speech encoder here our LPC is the speech coder and LPC we know that it is linear prediction coding filter what it is going to be present. and LTP is the long term prediction which is going to have pitch plus input parameter and RPE we said that residual prediction error which is going to be transmitted. So, what we have here is input is a 160. samples per 20 millisecond from our A to D converter. So, which we call it as 2080 bits what it is going to be transmitted and then this is our encoder here.

So, we will be converting it into different pattern as you can see that is 36 LPC bits per 20 millisecond and this is 9 LTP bits per second bits per 5 millisecond and this is 47 RPE bits per 5 millisecond. So, all the 3 which are getting generated from our GSM. So, total what it is going to be is 260 bits per 20 millisecond what we are going to have to channel encoder. So, what we are passing it through the system. So, how the coding is going to look like the first one we have to see the channel encoder also.

So, here it is going to be using class 1a. So, this is the error correction code what it is going to happen that is 3 bit error detection and convolutional coding that is error correction what we will be having it. So, we call it as CRC basically. Both the detection and then error correction what we will be having it and if it is a class 1B then convolutional coding is going to take place and class 2 no error protection is going to be provided in the thing. So, these are the three GSM this thing channel encoders are there.

And what we have is tail bits to periodically reset this convolutional coder when being used. So, what we pass is 260 bits per for 20 millisecond which is equivalent 13 kilobit per second. So, which class you are going to use is depends on the thing. So, here if it is 1a then you will be having 1a bits then 50 class what you are going to have and this is the error correction 3 bit CRC what we have it. and then for class 2 you will have 182 class 2 bits and 78 class 2 bits what is going to be generated out of it.

And this one goes as 53 bits because you are adding 3 bit correction codes. So, 50 + 3 = 53 bits. The other ones are one more whatever you have 182 class 2 bits are fed to your this thing convolution coder. So, which is has what you call it as 2, 1, 5. So, you are going to have 4 tail bits basically what it is getting added.

So, which comes down to 470 bits when it is coming out of the thing and this is the bit inter lever what you are going to have it which combines these bits and then the 78 class 2 bits which are coming. combined and you will be getting 456 bits per 20 millisecond as you can see from 260 you have increased to this and then whereas, in the case of kilobit per second it goes to 22.8 kilobit per second is the data rate what it is going from GSM speech coding from the channel encoder. So, you can have hybrid basically coders that is what we call it as code book excited linear predictive coding. So, that is problem with simple LPCs that is voiced and unvoiced decision and pitch estimation does not model transitional speech well basically.

and not always accurate. So, we said that accuracy was low in the LPC. So, whether we can use the code book excited model. So, what is that code book approach? So, that is pass speech through an analyzer to find closest match to a set of possible excitations that is known as our code book. and transmit the code book pointer plus LPC parameters and you can see that in NATDMA standard what they call it IS95 in 3G what they use it.

So, the standard for this is ITUG.729 standard. How is it going to be? It is going to have both the things that is here you have the vocoder basically seeing the thing voiced and then unvoiced what it will be going into the code book generation. So, from there you will be taking the code book generated signals and it passes through the synthesis filter and this is how the output is going to be represented. So, then if it is having a multiple excitation, so which is taken into the thing pass through the synthesis filter and this is our output. Then if it is a stochastic excitation is used, use the synthesis filter and then you take the output through that.

So, this way you can use the hybrid model to improve upon the standard of LPC. So, this covers are this thing a coders basically and then in the next class we will be discussing about code excited linear predictive model. Thank you.