

Real-Time Digital Signal Processing

Prof. Rathna G N

Department of Electrical Engineering

Indian Institute of Science - Bengaluru

Lecture - 43

Speech Coding I

Welcome back to real time digital signal processing course. So, today we will discuss about a speech coding the first part of it. So, in the last class to give a recap of the class. So, we discussed about the graphic equalizer. Why do we need the graphic equalizer? So, we said that in different applications even varying your mobile phones when you are using it. So, you can make the audio equalize.

So, that whichever part of it you can hear. So, even in your music systems you can use the graphic equalization to get clear cut voice if you want it or if you want to suppress and only have only musical instruments you can have the thing and then use your voice for recording and other purposes as we said that is the application one can use. So, for that we will be using the graphic equalizer and then if you want to completely block the outside noise also noise cancellation what you can do that. So, today we will see little bit on speech coding, why do we need it and then what is the advantage of it.

So, what is it? We know that digital speech which convert analog speech to digital form and then transmit digitally. So, applications of digital transmission we know that noise is whatever channel noise is added we can do the cancellation what we have seen it in the previous classes. And, then even the storage as we know that we can compress the speech and then keep it and whenever we want to do the thing. So, we can decompress and then use it. So, the applications what we will be seeing in this case is.

as you are seeing the thing first one is in the telephony that is basically in cellular wired and then internet voice over IP we need speech conversion and in the speech the speech storage. So, that is automated call centers. So, you can record the voices and then you will be playing it. So, most of you may be receiving it in the automated through your mobile that these are the facilities available or why do not you register and other things voice messages will be coming in your mobiles. And in the high fidelity recording or voice what we need the thing speech coding.

And next one is all of us know that it is text to speech that is machine generated speech. So, we will be seeing that for the physically disabled people and then the challenged

people who are unable to hear the thing. Whether we can play them from the text whatever the materials we have most of it we have in the text form. Whether we can convert it to speech and then they can if their auditory system is good. So, they can hear the thing and then they can understand.

So, this is one more application what we will be seeing one can work on. So, what are the issues? So, one of the issue what we say when we are transmitting or whatever we need a efficient use of bandwidth that is we can compress to lower bit rate per user. So, that we can cover more users in this particular bandwidth. And then we have to look at the speech quality. So, want what we say is tall grade or better quality in a specific transmission environment.

So, environment is noise is bad, but still we want to have a better quality in the thing. So, the other one environment itself plays an important role. So, that is what we call it as bit error rate basically. So, how many bits we want to transmit or whether we can send lesser number of bits and then how much packet lost is going to be one of the challenges one has to look at it. And, then if they are out of order the packets are out of order and then what is the delay.

So, these are the issues when we are transmitting directly. So, by coding and then we can account for all this losses that is environment losses then we will be able to reconstruct our code at the receiving end properly. The other issue one has to look at it is hardware complexity that is the speed basically both coding at the sending end and decoding at the receiving end will have a delay. is going to be defined by what is the hardware you are going to use it. So, if you are using the software, so we will be seeing that delay will be much more.

So, if we have the codecs basically then we will be having much speed. So, computation requirement and power consumption always one has to keep it in mind. So, power consumption in most of the handheld devices which is an important criteria and we should not be foregoing the computation basically. So, we have to cater to whatever the coding techniques we are going to use it. So, one has to be supported in your hardware.

So, how we are going to do the processing of speech? So, we say that in speech coding in wireless systems all this thing 1G systems have analog speech transmission in that and we know that in 2G and 3G systems have digital speech and what is the type of source coding we are going to have it which is going to vary. So, we will see it in a while now. So, what is the motivation for digital speech? So, we want to increase the system capacity as we discussed in the previous slide and then whether is it possible to compress the speech. So, if it is possible we want to compress and then storage is one of the thing. The other one as we said even the transmission multiple users can be compressed speech of them can be transmitted in the single line.

and then what is the quality or bandwidth tradeoffs can be made depending on our this thing speech. So, how to improve the quality of speech that is usually will be having the error control That is ECC what we call it and possible and as we discussed in the last class even the equalization what we can do using different techniques and then many more what we can use it to improve the quality of speech. And, then how to improve security as encryption possible for privacy. So, one of the example we said scrambling, where I can scramble my speed signal and then this is the way one can do the encryption or you can do in multiple ways and then at the receiving end whether we can do the decrypted speed signal what we can do it. So, and then at the same time we have to keep it in mind that reduced cost and operations and maintenance what we have to be considering it because most of the units for majority usage we need this parameter to be considered.

Coming to the distinct communication system in wireless mode what it shows in this diagram is we have the source place and we are going to do that encoding basically there and then after that we are going to encode the channel also and then use the modulator to generate the signal and pass it through channel. So, if we know the channel so, we will be encoding that also and then in the receiving place we are going to do the demodulation. So, it is corresponding to whatever modulated modulation technique what one has used it then we will go with the channel decoder and then we will be doing the source decoding and last is our destination. So, that is receiving this is from the sending end to this is at the receiving end how the wireless communication system looks like. So, what is the characteristics of speech basically? So, we call it as bandwidth most of energy between 20 hertz to about 7 kilohertz.

So, and then we know that human ear basically sensitive to energy between 50 hertz to 4 kilohertz. So, that is the reason why what earlier narrowband speech was 8 kilohertz sampling rate what we were using it. For all now mobile communications we know that the bandwidth what it is going to be covered is in this range. So, we will be going with the wideband speech which is 16 kilohertz basically. So, you will be seeing that what is the auditory which is given by this curve based on the thing.

So, narrowband and then wideband will be using based on it. So, the next one is based on the time signal. So, we say that it is highly correlated and then the speech is short term or stationary. So, for a while we can consider it as it is stationary and then we can process the signal. So, we classify the speech into 4 categories what we call it as one is voiced basically.

So, we say created by air pass through vocal cords. example like ha we you will be seeing that. So, you are seeing the speech how our mouth is going to be considered. So, we have the lips here and then this thing what is it teeth is here and our oral cavity and

this is the nasal cavity what we call it as. and then some of the other pharynx and then oesophagus what we will be having.

So, this is the ocular folds what we have it which will be vibrating. And the other one what we call it as unvoiced that is created by air through mouth and then lips. So, example S F. So, you can feel that when you are pronouncing it how the air is passing through your mouth as well as through the lips. The other one is mixed or transitional or in between what we call it.

And, the other part of it is silence. So, if nobody is speaking or whatever. So, we say that that is the silence zone. So, these are the four categories one can classify the speech in two. Now, what are the characteristics of speech? So, we said voice signals are produced when the vocal cords vibrate during the pronunciation of a phoneme what we call it.

and unvoiced signals by controls do not entail the use of the vocal cords. So, as an example so you can see that utterance of to what you do the thing and this you call it as the this thing. So, you will be seeing that consonant when we are pronouncing to t this is how the phoneme is going to be generated slash t and then uw. the consonant slash t is the unvoiced speech. So, this is the duration where you will be having it as unvoiced and then look like a noise as you can make out here.

So, then what happens while the vowel u w ok, voice speech is characterized by the strong fundamental frequency. So, you will be seeing that. which is coming with the thing. So, you can use MATLAB to generate this, this is a time versus amplitude what is given when you when we take up in the lab the example. So, we will see that how when a voice signal of 2 is given how the samples are generated.

So, you can see the same thing in the frequency domain this is what it is shown in the time domain. So, now, we will come to the digital speech. So, we say that speech coder why we need it device that converts speech to digital. So, we will be seeing that it can speech coding is bifurcated into it can be waveform coders or O coders. So, in this case you are seeing the span how it is varying.

So, you will be seeing the broadcast quality what you are seeing from 64 to 200 here. And, then the tall quality what you will see from 16 to 64 what it is going to be. So, this one comes in the waveform coding basically and communication what you can see is from 4 point sorry in between that 5 to 9.6 or little more than that what you have the thing. And this is the synthetic quality what we can generate the speech from here to here using O coders.

So, O coders are spanning between 1.0 to 16 in this case as whereas, our waveform coders are spanning between 4.8 to 16. And when you want to use both of them so you can generate the hybrid coders basically in this domain. So, what we say is that is

waveform coders they convert any analog signal to digital form. Whereas coders that we call it as parametric coders basically it is try to exploit special properties of speech signal to reduce bit rate.

and then it will be building the model of speech that is transmit parameters of model basically. And whereas, the hybrid coders combine features of both waveform and then vocoders and then speech coders are generated. So, what are the quality of speech with various coder is shown in this diagram. So, usually what we do is mean opinion score. So, it is a subjective measure of quality that is we put we generate the speech and then we will call the persons.

to evaluate whether they are able to identify the voice whether it is male voice or female voice and then what it has been sentence has been uttered if it is sentence generation and then we will say that what is the opinion of them then we will be calculating the mean opinion score from different characteristics of the people when they evaluate the thing. So, the other one is trade off in quality versus data rate versus complexity what it is going to be shown. So, you have seen here that general speech quality versus transmission rate has been plotted. This is the x axis whereas, of here it is MOS is our mean opinions core is put the thing. So, you will be seeing that synthetic quality is the data rate goes with 8 kilobit per second to 64 kilobit per second that is what you have seen in the previous one also.

And then you will be putting the source coding is on this side of the line. So, and then the waveform coding goes beyond a 16 kilobit per second. So, you will be seeing that synthetic quality is at the bottom and then communication quality just spans between the two of them and the tall quality and broadcast quality has as we have seen in the previous slide also it is going into the waveform coding. So, you will be seeing that beyond what we can have is 16 kilobit per second whereas, the hybrid So, you will be between that synthetic and then communication. So, it lies in this domain and you will be seeing that mean opinion score is very high for our broadcast quality and then we will be seeing that it is low for our synthetic quality of speech what it is getting generated.

you will be seeing that how to select your now coders. So, we say that waveform coders what we call it as PCM pulse code modulation all of you would have heard the thing. So, this is the waveform coders what it does basically convert any analog signal to digital basically A to D converter one has to use to do the conversion of speech. and analog signal sampled as we know that it should be more than twice the highest frequency component. Then what we are going to do is we are going to do the quantization into n bit samples.

So, when we had taken the number system we have seen how many the number of bits what I am going to. ah represent my input signal even the speech coding is going to have

the same thing. So, you are seeing what we want to represent is in a linear way, but what we are going to depend on the number of bits. So, you will be seeing that it will be the nearest value what will be representing it which is the error what we have taken is $\pm \frac{\Delta}{2}$. So, that is what the error what we will be getting in representation.

So, n bit samples what we are going to represent it. Then uniform quantization what we have seen the thing and example is here what we will be taking is pulse code modulation. And the band limit speech for this is less than 4000 hertz. So, pass speech through either a mu-law or compander. So, for more details you can refer to the textbook for what are the compounding it is going to happen.

And here the sample is what we will be taking at 8 kilohertz and usually it is 8 bit samples, then what we have is a 64 kilobits per second what we can represent. And this is a digital rate that is digital sample output rate what we are going to get from this representation. So, as you will be seeing that it is 8 into 8 is 64 kilo bit per second what we will be getting it. And how is the characteristic of this? The quality is very high. And, then we know that complexity for pulse code modulation is low and bit rate as we can see that it is high and in this case its delay is going to be low and robustness of this is high.

So, coming to that is PCM speech coding system with analog compounding then digital conversion the standard the what we will be seeing is that is international telecommunication standard ITU which is G 700 standard basis for speech coding. In our PSTN in 60s at you can see in our digital lines whatever telephones what we have it in 60s is given by this. So, you will be seeing that this is an analog input what we have is a band pass filter then we will have the analog compressor here. and then sample and then hold circuit. So, which is going to use the mu-law compander and then what the output will be getting it is a pulse amplitude modulated PAM basically what we will get.

And then we will do the conversion analog to digital conversion which it is known as pulse code modulation here. And we will be putting it in the transmission medium digital to analog conversion what we have to do it. and then here again analog modulation you will have it this is the whole circuit. So, and then you have the analog expander and then use a band pass filter to get the analog output.

So, you the mu-law expander was used for this purpose. So, what are the companding techniques one has ok. What it says is analog compander emphasizes small values and deemphasizes large values in order to equalize a signal to noise ratio across the samples. reverse the mapping at the receiver with an expander. So, one is compander the other one is expander at the receiving end.

So, the thing is given as $F(s) = \sin(s) \frac{\ln(1+\mu|s|)}{\ln(1+\mu)}$. So, what the μ log companding how it is going to look like is shown in this figure. So, we have the input and this is the output. So, 0 is no compression is done.

So, you are seeing it is a linear what it is getting passed. And, then $F \mu$ is equal to 5. So, this is how the companding is going to happen and then $F \mu$ is 55. So, you will be seeing that it is the companding is going to be in this way. So, that is what the frequency what you will be generating and then sending it out. So, now, coming to pulse code modulation speech coding, this is digitally compounded PCM system.

The standard for this in ITU is G dot 7 1 1 standard. So, what it has? So, better quality speech than analog companding as in the previous case. So, we will see how it is going to be the thing. And, you can modify this PCM coding to differential PCM which is known as DPCM to reduce bit rate from 64 kilobit per second which was in PCM to 32 kilobit per second. So, since the change is small between sample transmit one sample in this case, then on transmit difference between samples use 4 bits to quantize and adaptively adjust range of quantizer. That is it is going to improve the quality. So, what you will be when you are doing adaptation. So, DPCM is going to be named as ADPCM which has the standard G dot 726. So, you can see from G 700, 711 and then 726. for the ADA PCM.

So, what we have here for the this thing PCM transmitter. So, you know that it is analog input you will be band pass filter what you will be using it to eliminate what is the kilohertz what we need it. So, if it is in the 8 kilohertz we know that it is between 50 hertz to 4 kilohertz that is what the samples what we are interested in. So, that is the reason why what we have the band pass filter here and then do the sample and hold a circuit basically we call it as a PCM transmitter. So, again a PAM and then you will be converting analog to digital form and what we have is the linear PCM what will be taking the output and we will use the digital compression here. So, this compressed a PCM is going to be transmitted through the transmission medium basically.

Then this is going to be our receiver part which is going to be in the reverse direction as you can see. So, we will be expanding the thing from compressor will be expander is going to be used here and then this again it is a linear PCM what it is going to be used and then we will be converting into digital to analog converter and then Thus will be analog modulation whatever we had the thing and then we will be having the whole circuit and then convert it into or using the band pass filter. So, we will get back our analog output. So, this is the transmission as you can see PCM transmitter and then PCM receiver. So, these are the blocks which are present in the transmitter and then these are the receiver blocks which is going to work on it.

So, coming to what is this thing are differential PCM speech coding how it is going to be because what we said is from 64 kilobits per second we are reducing it to 32 kilobit per second how it is going to have. So, we said that it is going to be what is it transmit is going to be one sample that is since the change is small between the sample. So, we will use transmit one sample. and then transmit difference between the samples what it is going to be.

So, we will be using only 4 bits to do the quantization. So, how does the figure look like? Here it is the analog input, here it is going to be a low pass filter and then what you are going to have is here there is a differentiator basically. So, that is a summer basically this is positive and negative. So, you will be taking the difference between that two and then you will be doing the analog to digital converter and encoded different samples what you will be sending it out for transmission. And the same thing what you are going to do is you can convert it into digital to analog converter. And, then put the integrator and then accumulated signal level what you will be subtracting from this and then the rest of the signals are passed through this.

So, what happens at the receiver actually? So, it has to be same replica in the reverse way what it is going to work. So, you have a low pass filter which receives the thing. sorry this is the DPCM signal what it is coming and then you are going to take it to the digital to analog converter and then you are passing it through the integrator in this case and then you will be adding the signals from the hold and circuit to check that the difference between the signals and then reconstruct it and then which is getting passed through the low pass filter and then we will be getting the analog output. So, this is how our differential pulse code modulated speech transmission and receiver looks like.

The other way of doing is we can do the sub band speech coding. So, what is it? We are going to partition signal into non overlapping frequency bands and then use different A to D quantizer for each band. So, we will be seeing that as an example. So, if we have 3 sub bands.

So, we know that this is 5.6 kilohertz and 12 kilohertz and 13.6 kilohertz. So, sorry it is in the 3 bands what you will be having it. So, then you can reduce from 32 kilobit per second in the DPCM to 31.2 kilobit per second. So, what is it? The range is band there are these are the 3 bands what you have it. So, you will be having 50 to 700, you will be using only 4 encoding bits because most of the signals are present in the thing.

So, you will be giving more number of bits in this range and lesser 1 bit less in 700 to 2000 range. and we know that not much information is here. So, only it is going to be done with 2 bits. So, that this is how total what it is going to be generated.

So, what happens to this? So, here it is going to be this thing. So, many kilobits per second what we have to transmit in this frequent range basically and this is the kilobit per

second and this is in the third band what we have to do it total generating 31.2 kilobits per second. So, how does it this is going to be looking like? So, as you can see the name itself says that we are going to have 3 bands. So, we will be using band pass filter in the first stage. to pass only 50 to 700 hertz and then the second one is band pass filter in the range 700 to 2000 hertz and the third range is going to be 2000 to 3400 hertz.

These are the three bands and we need A to D converters in all the three bands, then later on mix them and send it through the channel encoder. So, this is how subband speech coding is going to happen and in the decoder part of it you will be reconstructing. So, we will see little more on the vocoders and speech coders in the next class. Thank you.