Real-Time Digital Signal Processing

Prof. Rathna G N

Department of Electrical Engineering

Indian Institute of Science - Bengaluru

Lecture - 45

Speech Coding III

Welcome back to real time digital signal processing course. So, we are discussing about the speech coding why do we need it. So, today we are going to continue with the speech coding. So, in the last class we discussed about the OCODAS and then how the GSM model look like And then today we will look at our this thing code excited linear predictor encoder. So, how the graphical representation of kelp encoder is look like. So, what we have is a speech input and then here it is error weighting and then minimization.

So, that is whatever error which is going to be generated from this is going to be try to minimize this error from the thing and this minimized error is fed into our excitation code book. and the same error minimized error is going to be transmitted. So, in this case and here is our this thing speech input from the linear predictive filter. So, which is going to be subtracted from the thing and you will be generating the error.

So, which is going to be minimized ok. So, the same input is going to go into our coefficient determination basically and then even the gain determination is going to happen from the speech input. Then these are the LPC coefficients which will be going into our linear predictive filter and will be. For this the input is going to come from the gain and this filter output is fed as the subtractive for our this to generate minimize our error. Then our to the coding thing what we have this LPC coefficients goes as the input R mixer and even the gain factor is given to this mixer and then all these three get mixed that is error.

Here earlier only normal error was given here we are going to do the minimization of it. So, that is fed into our mixer mix it and then this is our kelp output. So, that is how code excited linear predictor is going to work. So, coming with the this thing code is that is general kelp architecture looks like this. So, what we will have is encoder is the one from the sender side.

So, we have the code book and we will be feeding in to this the perspective error is what is coming from the weighting filter and based on it you will be calculating your gain is your output. And then what we have is a fine structure that is the pitch what is calculated whether it is voiced or unvoiced. So, we will be taking in the unvoiced case pitch and

then we will be feeding it to the long delay correlation filter. So, this is the pitch lag what the parameter which is coming out and from the code book you will be having the optimum code word what you will be generating and from the gain you will be putting it to the gain and then the next one is our spectral envelope. how we are going to look at it that is short delay correlation filter what we are going to provide the thing.

So, to get our LPC parameters. So, these are the four that is even the code word and then the gain pitch lag and then LPC parameters are passed through this is our channel as you can see the thing transmission channel which goes through the channel. So, the other one is we have the reference of the original speech and what is calculated. So, we will be doing the subtraction of the thing. So, we will have the objective error which is going to be into our waiting block and then we will be calculating the perpetual error and then we will be feeding it into our code book.

So, what happens at the decoder? So, we have the optimum excitation which has come from the encoder. So, we will be providing the gain basically whatever decided at the encoder also. And we will be doing this pitch this one pitch synthesis filter what we are going to extract our pitch lag. And we have the LPC synthesis filter there it was LPC analysis filter we will use the synthesis filter here. and then combining the thing so we will be creating our output speech.

So, you may be wondering how these are the things is going to happen in the lab we will see that how these can be generated and then we will be as an application LPC coding we will take it up and then how it creates the synthetic speech in the lab. Now, what is it that is what it says is NATDMA IS 54 speech coder that is if we are using the sub band code book approach termed vector sum excited LPC that is we self see what it is going to be named. In that case what is the thing is going to happen this is our I that is code book 1, and then this is a γ_1 is the output. So, what we have is the γ_1 is goes as input and then comes out and then we are having a code book what is input to that is part of L 2 L actually and this is our γ_2 and then gain whatever you are putting into the think system these are the gains γ_1 γ_2 γ_3 are the gains. and then this is the code book 3.

So, you have the h parameter and beta is the gain for this. So, these are the filter coefficients what it are going to be generated. Then you will be seeing that all the from the You will be adding them up and then going to give it as LPC synthesis filter it is going to be passed and this are the output. So, the filter coefficients which are been calculated will be going into our output for combination. So, how we are going to evaluate the speech coders? So, that these are the basis one is qualitative comparison what we are going to do it that is based on subjective procedures in ITU what it says is RECP 830 that is in the page 830 it has been specified.

So, based on it one has to do the qualitative comparison. So, that is subjective is you will be putting various subjects basically to hear these voices and that as we said mean opinion score is going to be taken from each one and then say that how much has said that it is quality is good or bad or based on it. what we have to calculate as we can see that mean opinion score how it is going to be graded if somebody gives excellent which has the value 5 and 4 for the good and fair has 3 and 2 is poor and then 1 is bad this is the MOS one is going to have the thing based on it. So, you would be subjecting different subjects to assess the quality of the speech based on this parameters. So, the other one is major procedures that are going to be followed which are the ones that is absolute category rating.

That is subjects listen to samples and rank them on an absolute scale result is a mean opinion score. That is in this case only they listen to the output and then qualitatively they compare the thing. But, here they have to do the absolute basically that is here the original and the synthetic speech what is been generated both of them they have to be able to absolutely compare and then give the mean opinion square sorry mean opinion score for that. So, that is what it says is comparison of mean opinion score that is CMOS what it says. So, it is much better has the value 3, better is 2, slightly better is 1, about the same is going to have 0 and slightly worse is minus 1 and worse you will be taking it as minus 2 and much worse is minus 3.

So, these are the scores basically. So, what does it how the comparison category rating is going to be generated based on this that is subject listen to coded samples and original uncoded sample PCM or analog basically any one of them because we know that PCM is very good that two are compared on a relative scale. Result is a comparison mean opinion score which is based on this what the subjects will be giving. So, these are the procedures one will be following in comparison or evaluating speech coders. So, further how it is going to be evaluated.

So, based on mean opinion score for clear channel environment that we call it as no errors in the channel result vary a little with language and speaker gender. So, as we know the accents from one person to the other one it is going to vary. So, based on it one has to do that analysis. So, you can see that the standard what it is given whether PCM standard or CT2 or DECT or TDMA or GSM or Q-KELP quantized KELP what we call it. and then this thing LPC coding.

So, quantized kelp you will be seeing that hybrid RALPC it can use or hybrid kelp what it can use and then LPC will be using the vocoder technology and we saw that the ITUG.729 is going to use hybrid kelp. So, what are the bit rates for these standards basically. So, we know that PCM is the waveform based. So, which is going to have 64 kilobits per second.

So, what was the that is. MO is value which is given is 4.3 for this whereas, our adaptive which is going to use differential PCM which we said it is 32 kilobit per second. So, the mean opinion score you can see is 4.1. Whereas, it says DECT is the ADPCM what it is going to use at 32 kilobits per second. So, this also has 4.1. Whereas, our TDMA technology that is hybrid V cell PCM you can see that the bit rate is 8 kilobits per second.

So which has MOS of 3. So, whether you can as you will be seeing that the transmission rate is very low in this. So, whether with this degradation you can accept this cases where in your application not much is required. So, you can use it so that you can save on the channel basically. Whereas, GSM as you can see which is the lowest what you can see 1.3 kilobits per second which is going to use our hybrid model that is REL PCM basically.

So, which gives the MOS score that mean opinion score of 3.54. So, that is the thing GSM is much more popular. And, then quantized kelp you will be seeing hybrid basically kelp what it is going to use which is at 14.4 kilobits per second. So, the mean opinion score you can see it is varying between 3.4 to 4.0. So, it has a better quality and then hybrid kelp which runs at 9.6 kilobits per second which has at the minimum that is what it is shown at 3.4. And, our LPC which uses a vocoder speech coding techniques which has 2.4 kilobit per second. So, the you can see very low mean opinion score what it has. And, G dot 729 which uses a hybrid kelp which is at 8 kilobits per second, so which has 9 score compared to our this thing what is it TDMA hybrid which has 3 at 8 kilobit per second. So, depending on the thing so, you can select that is as you can see that even the gender is going to matter.

So, if it is a male voice and female voice. So, we will see in the lab how it is going to vary. And, then you can decide on it what you want to select for your coding techniques. So, continuing on the evaluation types of environments recommended for testing coder quality. So, on what should be our environment on which you will be testing this quality.

It says clean channel, no background noise. So, you have to provide this so that you can evaluate it. The other one is vehicle that is emulate car background noise. So, if you are using the thing. So, you will be seeing that you have to emulate the car background noise and then you have to hear these speeches for giving mean opinion score.

The other one is if you are considering the street then emulate pedestrian environment. So, when the pedestrian is walking on the street what are the noises will be present. So, you will be using that to get the evaluation from the subjects. The other one what it says is hot that is emulate background noise in office environment that is voice band interference what you are going to do that. So, consider environment above for cases of what is it perfect channel, no transmission errors, random channel errors what you can consider and bursty channel errors.

So, when you are want to check the channel errors which you want to consider and then see what will be the mean opinions score given by the subjects. So, you may consider repeated encoding or decoding that is you will be seeing that mobile to mobile call one you can have it and then say that what is the because you are putting it in the clean environment and then that is emulating these things. And, then you are asking the person to give his mean opinion score you are not within the chamber. So, you are outside. So, then you will be having the finishing the thing you can ask them to give the score for that ok.

So, how you are going to do the codex selection? So, we know that for cellular need to consider basically quality, complexity, delay and compression rate. So, these are the standards which are available ITU that is international for telecommunication standards. G.711 is 64 kilobits per second. So, in this case coding delay is going to be 0. and even the decoding delay is 0, but we say the complexity of this is low.

Whereas, G.729 which has 8 kilobits per second and the delay is 15 millisecond and then decoding delay is 7.5 millisecond sorry coding delay is 15 milliseconds and decoding delay is 7.5 millisecond and the complexity of the code is medium in this case. So, if you use G.723 the versions are A and B what you have it. So, that is it can run at 6.4 this thing kilobits per second or 5.3 kilobits per second. Then the coding delay is 35.5 millisecond and decoding delay is 18.75 millisecond. And, the complexity is high. So, one has to accept which you want to have the this thing coding techniques you want to use in your work basically. So, if you do not want to have any delay, then you can use this G dot 7 11. So, we know that what are the 3G standards specified that is two competing 3G standards both standards use multimode kelp vocoders in them.

So, that is 3GPP that is CDMA 2000 and 3GPP UMTC what we have the thing. So, here it is multimode rate set in this case, here it is AMR NB multirate what it will be using it. So, here it is going to be variable bitrate O coder and source control of bitrate and channel coding treats all bits in this case equally. Whereas, you will be seeing that it is the fixed rate O coder what it uses. and voice activity detection is going to be discontinuous, transmission network control of coder rate what we are going to consider and then tailors channel coding to speech coder what the standard uses in the 3G.

So, what is it because as you know that wherever silence is required so we can use that and then you have we have to put that silence also in our speech basically. So, if the silence is not compressed we know that it will be taking more number of bits. So, much of a conversation is silence what it says is according to the standard approximately 40 percent is going to be the silence zone. So, we need not have to transmit that, but it has to be incorporated in the receiving end. So, what it says is it is going to do voice activity detector VAD that is hardware to detect silence period quickly.

So, you are building the thing in the hardware. So, we are going to use variable bit rate coders basically. to reduce bit rate when silence. So, I need not because I need not have to transmit this and say only that it is from this to this the silence zone. So, how we are going to take care of the discontinuous transmission which is called DTX ok. Stop transmitting frames and second minimal number of frames to keep connection up.

And, what the other one is comfort noise generator that is CNG that is synthesized background noise avoids did you hang up. So, that is how you will be having the noise generator and a random noise or reproduce speakers ambient background to cancel it out. So, for example, in the GSM codec and popular voice over IP which uses G.723 one codec has what is it all the three of them that is our voice activity detector and then discontinuous transmission and the comfort noise generator which have been combined and then supports the thing. So, in the CDMA 1 and then CDMA 2000 codec use variable bit rate approach.

So, that is what we will be using it. So, you can see that how the silence compression can be done. This is what you are seeing the A 1 this is the speech signal what we have it. So, you are seeing the here are voiced or unvoiced what we have it in between you will be seeing the silent zone. Here also it is a maximum silent zone what you are incorporating after that it is the silent zone. and then how this thing B 1 is going to look at it that is variable rate coding what you will be doing it.

So, you need not have to now you can transmit a 0 here wherever the silence zone is there and you can see here also and then the other places what you can say that it is now that is a variable rate basically that is b 1 which is number of bits are going to be very less for these places. And then when there is a signal you will be using the high bit rate. So, that you can get the value of the output. So, that is how you will be having the variable rate there.

So, that you can avoid the silence zone with minimal bits. and how the pulse is going to be generated that is kilobits per second. So, you will be seeing that this is a peak after that because this is a silence zone as you can see maximum silence what we have it. So, which goes down and then again you what you will be having it and then down. So, this is fixed rate with what you have is this thing VAD and then DTX incorporated in the sample speed signal how it is going to go about it.

So, the next one is what we have is the voice coding. that is basic voice coding approaches as we have seen that it can be waveform or vocoders or hybrid vocoders what you can use it. So, how we are going to do evaluation of vocoder quality. So, we have to see that code book based vocoders use in new technology and we know that in 3GBP and then ITU recently standardized a that is AMR wide band kelp basically. And, then input

in this case is 50 hertz to 7000 hertz. So, we are moving from narrow band to wide band signal as you can see rather than 300 to 3400.

This was the telephony standard and this is the new standard what we have it in 16 kilohertz. range of current systems. So, more natural quality speech that is slightly higher bit rate what we have to account for. So, this covers our speech coding. So, the application of speech coding what we will look in the next class. Happy learning and thank you for listening to this lecture.