Real-Time Digital Signal processing
Prof. Rathna G N
Department of Electrical Engineering
Indian Institute of Science, Bengaluru

Module No # 07
Lecture No # 39
M3U24 - Adaptive Echo Cancellation

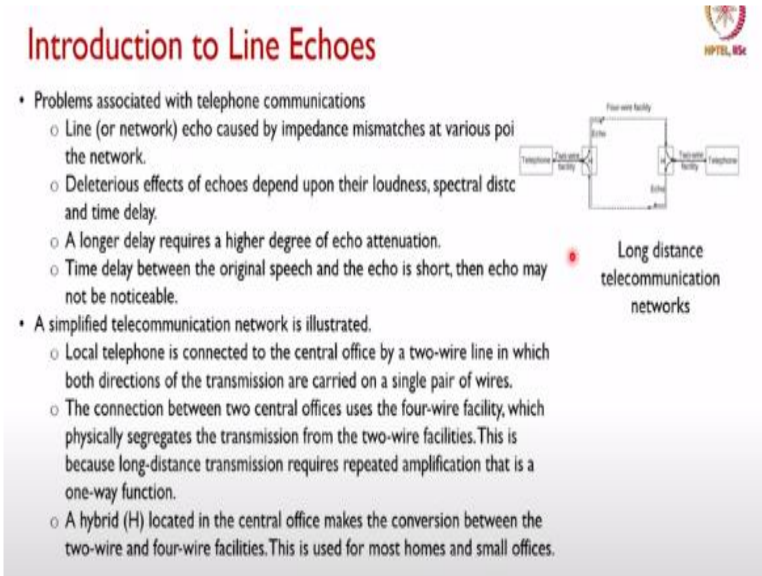
Welcome back to real time digital signal processing course so today we will be covering adaptive echo cancellation.

(Refer Slide Time: 00:33)

So in the last class we discussed about how do we do the prediction and we saw scrambling why do we need it to protect our own voice or audio files from a misuse of it. So we can do the scrambling and then send it to people who are only connected with our network. The other one is echo generation so most of the time we know that it is a hobby so how to generate the echo first we have seen the thing. Today we will be seeing how to do the cancellation of echo.

So most of the places what we visit which has the capability of echoing that is will have stones and other things which sometimes if it our voices reflect then we will be enjoying that when you go when all of us go for a picnic or other places. So in this case synthetic echo generation what we saw in the last class. So we will continue how to because this is one of the welcome for a obvious but in some of the cases we will see in a while that how it is going to hinder our output.

(Refer Slide Time: 01:53)



So well be looking at it how to do the cancellation, so first is we will see the line echoes basically. So in this case what we have is, this is the line indication that is most of you may ask that is, what you will be getting the cartoon network saying that it is what is it and take whatever phone lines or telephones what we showed to the kids, because they are used to mobile basically. But still some of the places only in the cities it may be not there but in what we will call it as villages still all of us know that in those people are also using most places mobile phones.

But still telephone communication is one of the parts of our past life and then some places it is still there. So what is it what we call it is line or network? So it, echo is caused by impedance mismatches at various point of interest what we will call it. So we are seeing this is a telephone where it is connected and then this is from the sender side and this is the receiver if we call it.

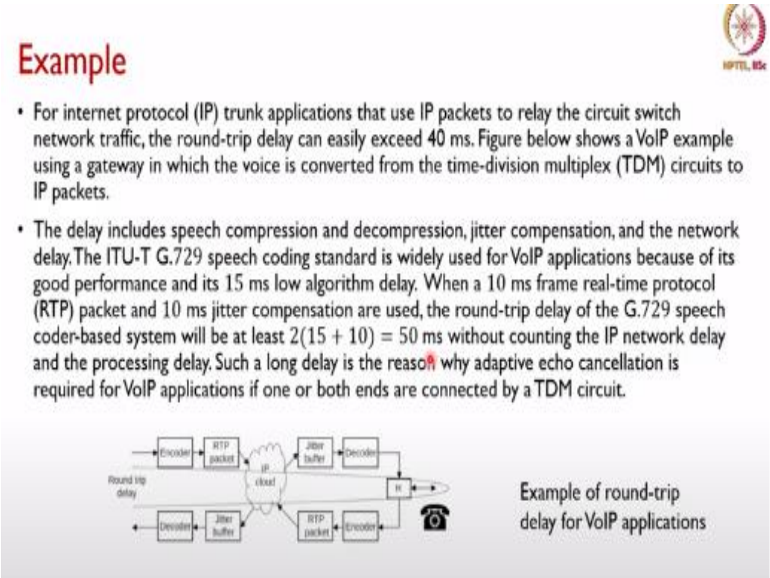
So these are the paths what it has to take it, so at different places you may have echo basically. So we say this, effects are they are going to depend on their loudness, spectral disc and time delay, so these are the ones which will be causing echo. So if there is no delay then we say that with not hearing any echo from the transmission part of it. What it says is longer delay requires a higher degree of echo attenuation, and time delay between the original speech and then the echo is short then, echo may not be noticeable.

So we will be seeing in the lab, so how we are going to run the thing. What is it this? Is a simplified telecommunication network which is being illustrated, so what are it contains basically. So this is the local telephone is connected to the central office by a 2 wire line in which both directions of the transmission are carried on a single pair of wires. So as you can see here it is connected with 2 wire basically, so the connection between 2 central offices they use 4 wire facility, which physically segregates the transmission from the 2 wire facilities.

And this is because long distance transmission requires repeated amplification that is a one-way function. So you will be seeing that this is of office one you call it office 2 basically you will be seeing 2 lines going from here and then other 2 lines so 4 lines what is going. So this is how it is connected and when you are putting it to the receiver, so you receive a telephone, so you will be seeing again from the office it will be 2 wire connections what will be having it.

So what happens in the thing is a hybrid located in the central office makes the conversion between the 2 wire and 4 wire facilities. So this is used for most homes and then small offices, otherwise what we are going to have is basically a, exchanges which is through which you can be connected.

(Refer Slide Time: 05:49)



So as an example for this we know that internet protocol trunk applications that use IP packets to relay a circuit switch network traffic. The round trip delay can easily exceed 440 milliseconds, so the you can see this is how it is connected so what you have is we call from here to here is your round trip delay, so which can cause 40 millisecond. This is a voiceover IP example using a gateway in which the voice is converted from the time division multiplex circuits to your IP packets.

So you have an encoder so you will be having the RTP packets basically and use the IP cloud and then you will be either transmitted this way or you can pass it through the jitter buffer. And finally you will be decoding it and this is the hybrid circuit what you will have it as you are seeing the telephone is here, from there you can do the encoding and then you will be sending the rp packet here to the cloud. And then it this is going to have a jitter buffer and then you can decode it.

So this is how what happens although this is the receiver what you are intended to send the thing but there will be a round trip delay which is coming back to the person who is calling. So this is how the, what you will call it as echo comes back from here to here, whatever you have spoken sometimes as I have pointed out in the last class, so that you will be hearing yourself. So the delay includes speech compression and decompression that is why you have both encoder and decoder there and Jitter compensation.

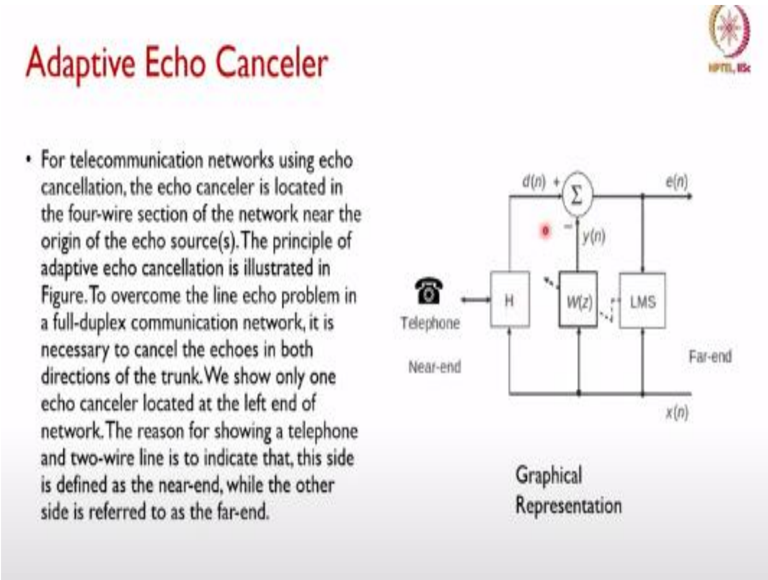
So what you call as jitter, it may be a little bit of noise or it may be a little delayed that is what we call it as jitter. And the network delay so we have to include the network delay also so the standard used for this voiceover IP is g.729, we call it as ITU-T standard for speech coding standard is widely used for voice over IP application. So why it is being used it because of it good performance and its 15 millisecond low algorithm delay what it has it that is round trip delay.

Where we said it is 40 millisecond whereas in the j dot 79 it is 10 millisecond low logarithm algorithm delay what we have it to compute the thing. So when a 10 millisecond frame real-time protocol what we call it as RTP packet and 10 millisecond jitter compensation are used the round trip delay of this g.729 speech coder based system will be at least twice that of 15 + 10 which is equal to 50 millisecond.

So that is a 10 millisecond is our jitter and then we have algorithm delay of 15 millisecond because it has to go and then come back as you can see encoder and decoder is here, here also we have the encoder and decoder that is why it is twice what it is getting multiplied. So which comes to about 50 millisecond without counting the IP network delay, here IP cloud what you are using it you have not counted and as will the processing delay is ignored.

So if such a long delay is the reason why our adaptive echo cancellation is required for voice over IP applications if one or both ends are connected by our TDM that is time region multiplex circuit. So for this reason one has to use the echo canceller as you can see that, that is what it shows that round triple delay can exceed 40 millisecond this is with respect to G.729 shown that it is going to have a 50 millisecond round trip delay right.

(Refer Slide Time: 10:27)

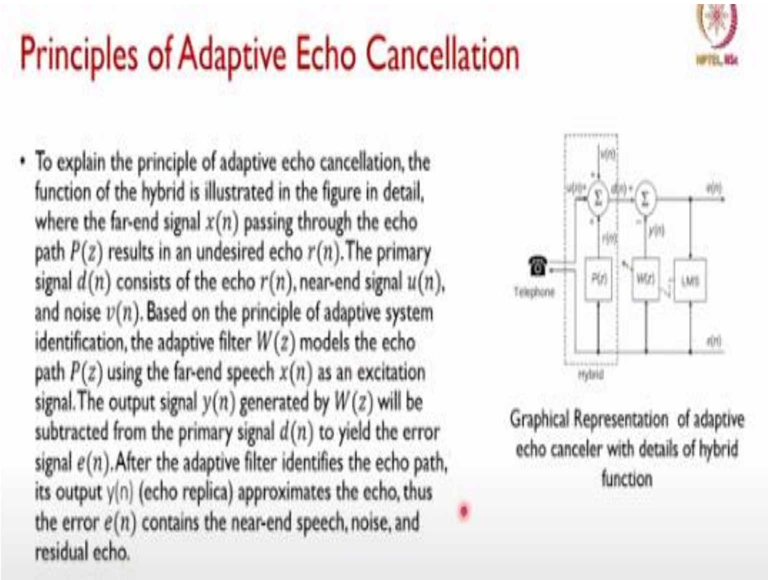


So how we are going to design the adaptive echo canceller what it is shown here, using echo cancellation the echo canceller is located in 4 wire section basically. That is the reason why all the stations they use the 4 wire section of the network, near the origin of echo sources. So you have the telephone this is near-end and then the principle of adaptive echo cancellation is what shown in the diagram here, so we will see how it is going to work.

To overcome the line echo problem in full duplex communication, I think you must be, knowing half duplex is going to be only one way communication full duo-play is you will be getting back your signal, that that is known as our full duplex. It is necessary to cancel the echoes in both directions of the trunk basically. So only one echo canceller located at the left end of network is it is what it is shown here, basically.

For showing a telephone and 2-wire line is indicate that this side is defined as the near end, while the other side is referred to as the far-end. So this you are calling it as far end and you will be having the near end the circuit what you are having is the hybrid circuit, so which you will be giving a compensation for the echo.

(Refer Slide Time: 12:08)



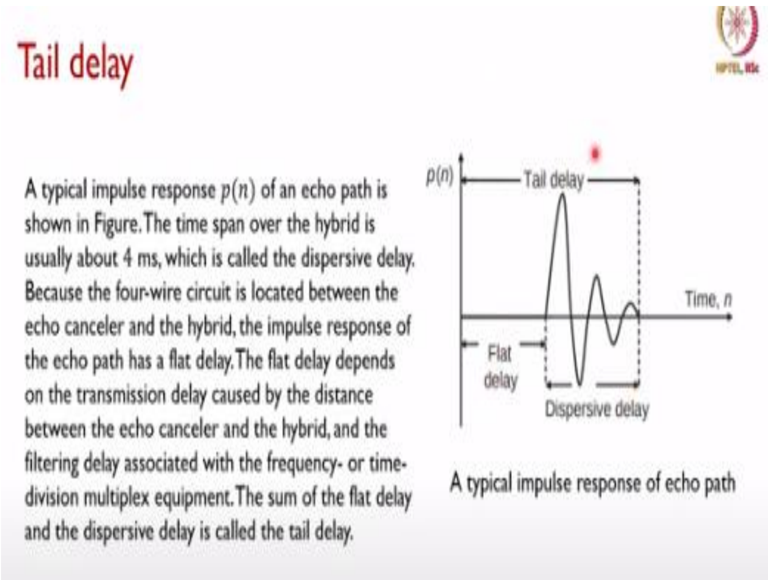
So we will see how principles of echo cancellation will be looking at it, so what is it, echo cancellation the function of the hybrid is whatever we have put the h here, the hybrid circuit is shown in this case which is expanded. So where the far-end signal this is what your far-end signal $x(n)$ passing through the, your echo path that is $P(z)$ is your echo path results in an undesired echo $r(n)$.

So when it is passing through so it will be $P(z)$will be causing this thing echo which is known as r n, so near-end signal $u(n)$ what you have considered and noise is defined with v of n. So; based on the principle of adaptive system in this case identification the adaptive filter w of z here models the echo path $P(z)$. What is the echo path is going to be modeled and what happens using the far end speech $x(n)$ as an excitation signal.

So for this it is modeling $P(z)$ based on r $x(n)$, and what happens the output signal y of n generated by this $W(z)$; we have y of n here will be subtracted from the primary signal d of n. So this becomes our desired signal and subtracting it will be calculating the error we will try to minimize that error. After the adaptive filter adding identifies our echo path its output y of an echo replica what we call it approximates the echo, thus the error a, of n contains the near-end speech, noise, and then residual echo.

We may not be able to completely suppress the thing but little bit of it is there most of it is going to be suppressed basically.

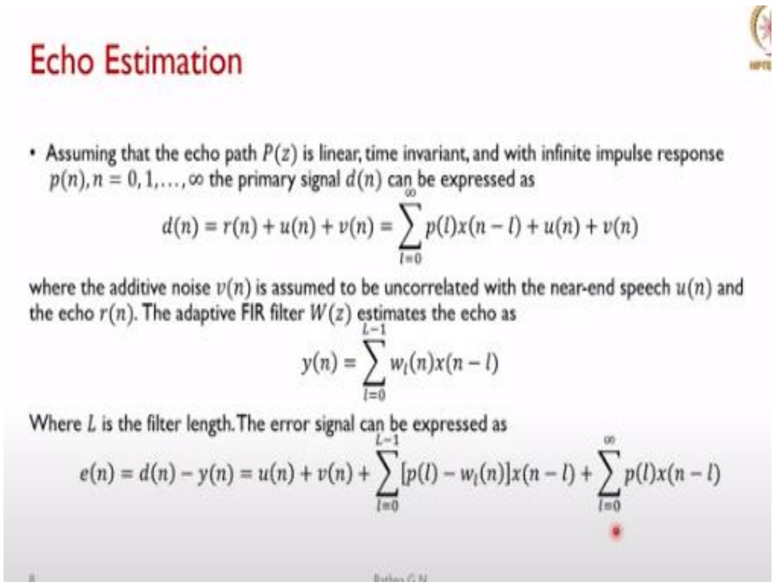
(Refer Slide Time: 14:29)



So what we call it as tail delay, what is that delay so we will be seeing the impulse response of this p of n of an echo path is what it is shown here the time span over the hybrid is usually about 4 millisecond, which; is called the dispersive delay. So you will be seeing that this is approximately, 4 millisecond which; is going to be dispersive delay. Because the 4 wire circuit is located between the echo canceller and the hybrid, the impulse response of the echo path has a flat delay.

As you can see it, here it there is a flat delay then you will be getting the dispersive delay, the impulse response of the echo path has a flat delay. The flat delay depends on the transmission delay caused by the distance between the echo canceller and then the hybrid, and also it depends on the filtering delay associated with the frequency or time division multiplex equipment. The sum of the flat delay and the dispersive delay is called tail delay, so the complete thing including flat delay plus dispersive delay is known as the time delay which is shown in this figure.

(Refer Slide Time: 15:50)

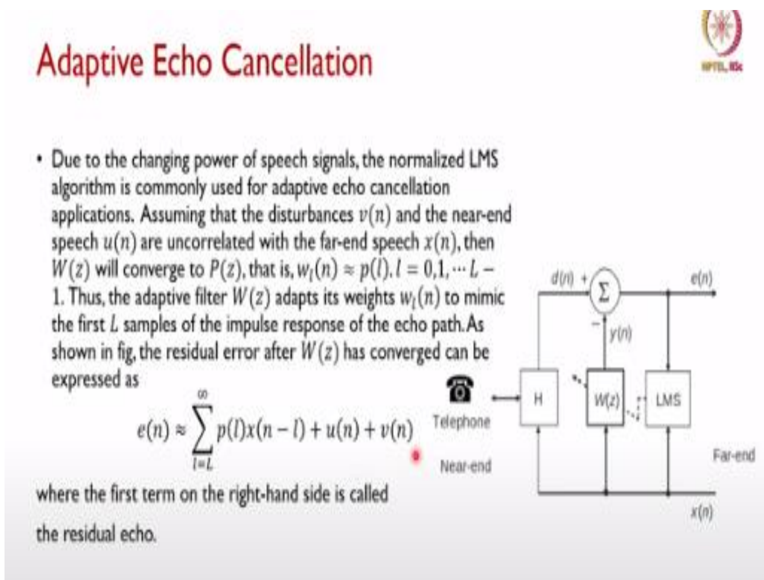


So how we are going to estimate the echo, so that is echo path what we have is a $P(z)$we say it is a linear time invariant. And with infinite impulse response that is IIR filter what will be defining it which is given by $p(n) = p(n),\ n = 0,\ 1,\ .\ .\ .\ ,\infty$. The primary signal $d(n)$ in that case can be expressed as which is given by $r(n) + u(n) + v(n) = \sum_{l=0}^{\infty} p(l)x(n-l) + u(n) + v(n)$.

So we call it as $v(n)$ as the additive noise which is uncorrelated with the near end speech $u(n)$ and the echo is $r(n)$. So this is the when you substitute the thing $p(l)x(n-l) + u(n) + v(n)$. So what will be the adaptive FIR filter $W(z)$ is going to be estimated as echo basically, y of n is given by we have seen that $l = 0$ to $L - 1$ is the length of the filter, w_l are the weights of the filter and then $x(n-l)$ is the input to the filter.

So that is what l is our length, and error signal in this case can be expressed as $e(n) = d(n) - y(n) = u(n) + v(n) + \sum_{l=0}^{L-1}[p(l) - w_l(n)]x(n-l) + \sum_{l=0}^{\infty} p(l)x(n-l)$. So this is how the substitution is going to happen $d(n)$ and then $y(n)$ so when you put the terms in proper way.

(Refer Slide Time: 18:08)

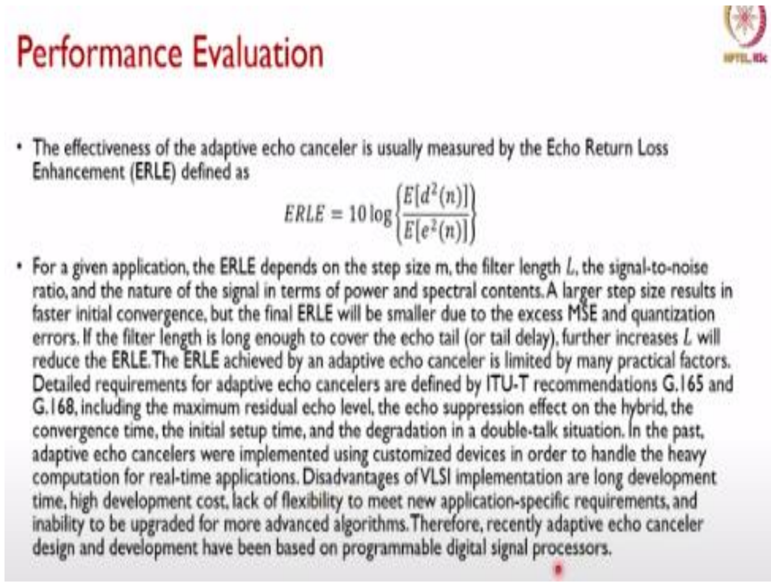


So, how we are going to do the echo cancellation, so due to the changing power of our speed signals the normalized LMS algorithm is commonly used for adaptive echo cancellation for all these applications. Assuming that the disturbances what we call it as v of n and the near end speech $u(n)$ are uncorrelated with the far-end speech $x(n)$ then $W(z)$ is going to converge to, $P(z)$; that is $w_l(n) \approx p(l)$. $l = 0,1, \cdots L - 1$.

So the adaptive filter $W(z)$ adapts its weights $w_l(n)$ to mimic the first L samples of the impulse response of the echo path. So as it is shown in the figure this is our hybrid, so that is what it is right this $W(z)$ is getting adapted. The residual after $W(z)$ has converged can be expressed as this is convert then you will be putting our $e(n) \approx \sum_{l=L}^{\infty} p(l)x(n-l) + u(n) + v(n)$.

So where the first term on the right hand side is called the residual echo, this what we call it as residual echo which is carried and then a signal and then the noise what we said that will be there at the far end.

(Refer Slide Time: 19:54)



So, how to estimate the performance basically, so effectiveness of our audio echo canceller is usually measured by the echo return loss enhancement. That is ERLE which is defined as that is based to the log what we will be taking it 10 log expected value of our desired signal square d squared of n divided by expected value of the error square function e squared of n. So for a given application how we are going to consider it ERLE depends on the step size m that is mu what we consider it the filter length l, signal to noise ratio and the nature of the signal in terms of power and spectral contents.

A larger step size results in faster initial convergence, but final ERLE will be smaller due to the excess minimum mean square error and quantization errors. And if the filter length is long enough to cover the echo tail or tail delay what it is known as so further increases l will reduce the ERLE. So ERLE achieved by an adaptive echo canceller is limited by many practical factors the detailed requirements for adaptive echo cancellers are defined by the standard ITU-T that is telecommunication standard recommendation for G.165 and then G.168.

Including the maximum residual echo level that is suppression effect on the hybrid the convergence time the initial setup time and the degradation in a double talk situation. So one has to we look at the, what is that double talk in the next slide. So in the past adaptive echo cancelers what was happening, implemented using customized devices in order to handle the heavy computation for real-time applications.

So we know that software delay is going to be much more. So hence there was a hardware which was designed to do this cancellation in real time applications to reduce the delay. So what was the thing disadvantages of VLSI implementation are long development time, high development cost, lack of flexibility to meet new application specific requirements and inability to be upgraded for, more advanced algorithms.

So when you design a hardware VLSI implementation it is very large system integration. So if the volume is too high then you can go for the VLSI implementation. So what are the drawbacks of it one is because development time one has to design your circuit from a transistor level, so which is going to take longer time that is what it says long development delay. And high development cost because once it becomes a long development, so you know that the person has to be paid and other things, so the cost is going to be very high.

And then what is the other disadvantage, lack of flexibility, so if I want to modify any circuit so then I have to go for redesigning so I would not be able to do that design within the time constraints. So these two will be going up again if I had to redo the design with little modification it is not possible to accommodate it, that is what it says any new application is coming, so we may not be able to use the same circuit it has to be completely modified.

And what is it the other thing is you cannot upgrade this system. So, most of you must be, knowing that you will be upgrading your memories if you know DDR3 and then now DDR4 which is coming in earlier DDR 1 , 2. So whenever those systems were there so if you want to upgrade it you had a limitation of memory for those systems; where for the other one you have to completely change your CPU so what you will be doing now also it is so fast or they become obsolete.

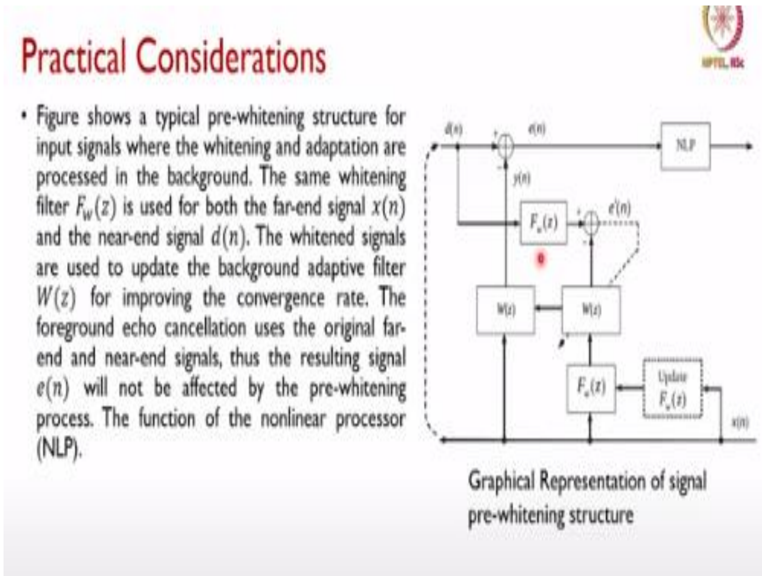
So for the advanced algorithms whatever nowadays being used so you will be not be upgrading the, whatever you have designed in hardware, so these are the drawbacks. So what is the solution what they gave is recently adaptive echo canceller design, and development have been based on programmable digital signal processors. So most of the even your mobile has a dsp, so which is programmable. So hence you will be avoiding you will be getting once in a while upgrade your software correct.

So whenever you upgrade the software even the hardware must be able to run those software. So if it is not going to do that then you may not be able to upgrade your mobiles or any of the hardware system for that matter. So that is how from the custom design it has gone to the programmable device what you can look at it, signal process.

(Refer Slide Time: 26:04)

So what are the practical considerations we call it as pre-whitening of signals in this case that is the convergence time of adaptive FIR filter using LMS algorithm we have discussed is proposed additional to the spectral ratio that is $lmax = lmin$. So since the speed signal is highly correlated with a non-flat spectrum the convergent speed is generally going to be slow. So the decorrelation that is what we call it as whitening of the input speed signal can be used to improve the convergence speed. So this is one of the application a practical consideration one has to look at it.

(Refer Slide Time: 26:53)

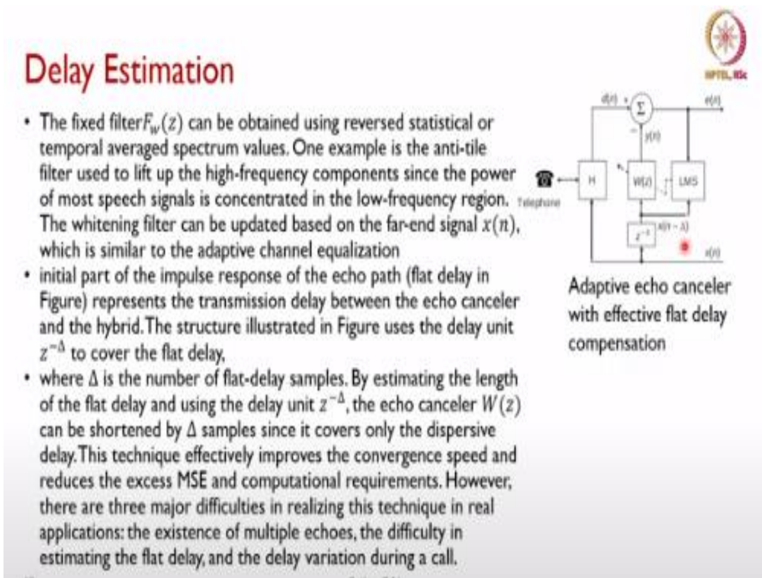


So the other practical consideration what it is shown in the figure here a typical pre-whitening structure for input signals where the whitening and then adaptation are processed in the background. The same whitening filter what you call it as $F_W(z)$ what it is given here is used for the both far-end signal $x(n)$ and the near-end signal you are $d(n)$. So you will be using for both far-end is $x(n)$ and then $d(n)$ is near-end; so you will be using the same this thing filter for both of it.

So the white and signals are used to update the background adaptive filter $W(z)$, so from here you will be based on it what you will be updating your weight functions of your filter for improving the convergence rate. The foreground echo cancellation uses the original far-end and near end signals thus the resulting signal $e(n)$ will not be affected by the pre whitening process. So the function of the non-linear processor what it calls will be used for this effectiveness.

So this is how you will be putting the thing although if it this thing what it says is it is original far-end and foreground echo cancellation is not going to be affected with your what you call it as whitening part of it that is pre-whitening.

(Refer Slide Time: 28:41)



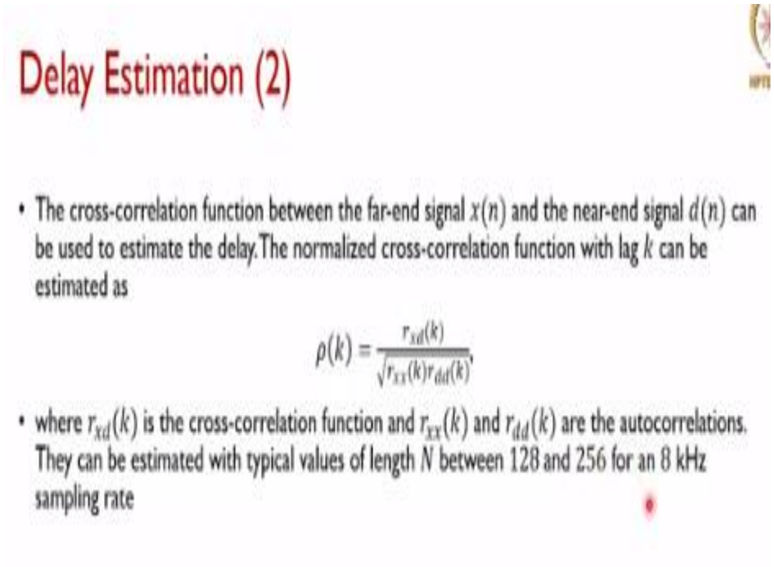
So how we are going to estimate the delay that is what the next slide shows. That is fixed filter what we have is $F_w(z)$ can be obtained using a reversed statistical or temporal average spectrum values. So the example is anti-tile filter used to lift up the high frequency component, since the power of most speed signals is concentrated in the low frequency region. So the whitening filter can be updated based on the far-end signal $x(n)$ which is similar to the adaptive channel equalization.

So initial part of the impulse response of our echo path that is flat delay in the figure what you will be seeing it represents the transmission delay between the echo canceler and the hybrid. So this is what the delay which is getting represented thus that is what represented by $z^{-\Delta}$ and your input is getting delayed by $x(n-\Delta)$ in this case.

So delta is the number of flat delay samples and by estimating the length of the flat delay and using delay unit $z^{-\Delta}$ the echo canceler $W(z)$ can be shortened by delta samples since it covers only the dispersive delay part of it. So this technique effectively improves the convergence speed and reduces the excess mean square error and computational requirements.

So however there are 3 major difficulties in realizing this technique in real applications that is the existence of multiple echoes, the difficulty in estimating the flat delay. So if it is a single delay you can estimate it and then do the cancellation otherwise if like example we considered as a reverberation if we have multiple echoes then your flat delay and the delay variation during a call it is a little bit difficult to cancel.

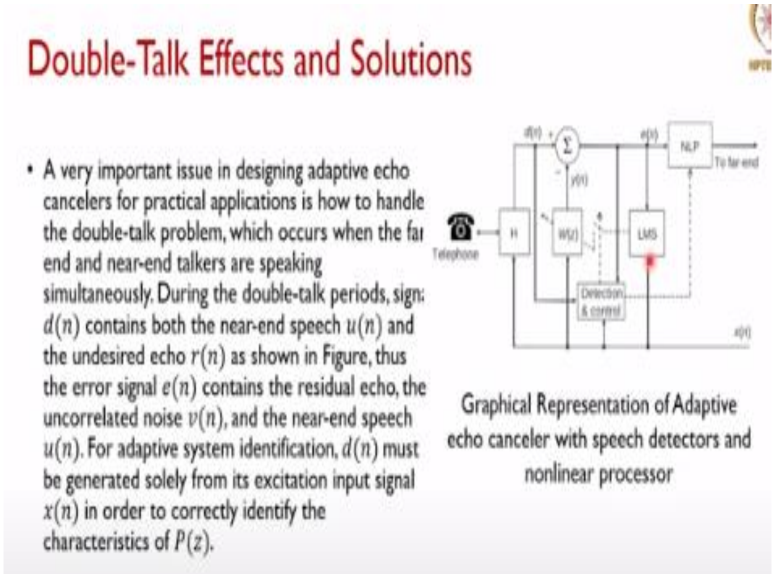
(Refer Slide Time: 30:58)



It so how you are going to estimate your delay, so one is using the cross correlation function between the far-end signal $x(n)$ and the near-end signal $d(n)$ can be used to estimate the delay. So what is the we have derived this normalized across correlation function is given by a $\rho(k) = \frac{r_{xd}(k)}{\sqrt{r_{xx}(k)r_{dd}(k)}}$ that is input signal and this is the desired signal autocorrelation value under the square root of it.

So what it defines is r_{xd} is the cross correlation function and $r_{xx}(k)$ and $r_{dd}(k)$ the autocorrelations. So it can be estimated with typical values of length n between; 128 to 256 for an 8 kilohertz sampling rate. So this is our sampling rate that is narrowband frequency what it has been considered in this case.

(Refer Slide Time: 32:09)



So what are the double talk effects so you would be seeing your own echo is coming and someone else also there will be a cross talk is going to happen, so how you are going to look into this. Designing adaptive echo cancelers for practical application is how to handle the double talk problem. So with occurs when the far-end and near-end talkers are speaking, so simultaneously it is going to happen.

So during the double talk periods what you will call that is sign basically $d(n)$ contains both the near-end speech $u(n)$ and then the undesired echo $r(n)$. So that is what, what it is shown so you have the hybrid as well as the, this is the far-end what you have the NLP in this case. So how you will be detecting it then control algorithm what you will be providing it. So does the error signal $e(n)$ contains; the residual echo, the uncorrelated noise of $e(n)$ and then a near-end speech $u(n)$.

So for adaptive system identification that is $d(n)$ must be generated solely from its excitation input signal $x(n)$. So in order to correctly identify the characteristics of our $P(z)$ for the hybrid thing, so you will be generating from the $x(n)$, so to do the thing.

(Refer Slide Time: 33:52)

So we will consider in the next class equalizer and then some of the speech coding techniques available in the literature. So, thank you for this hearing and then happy learning.