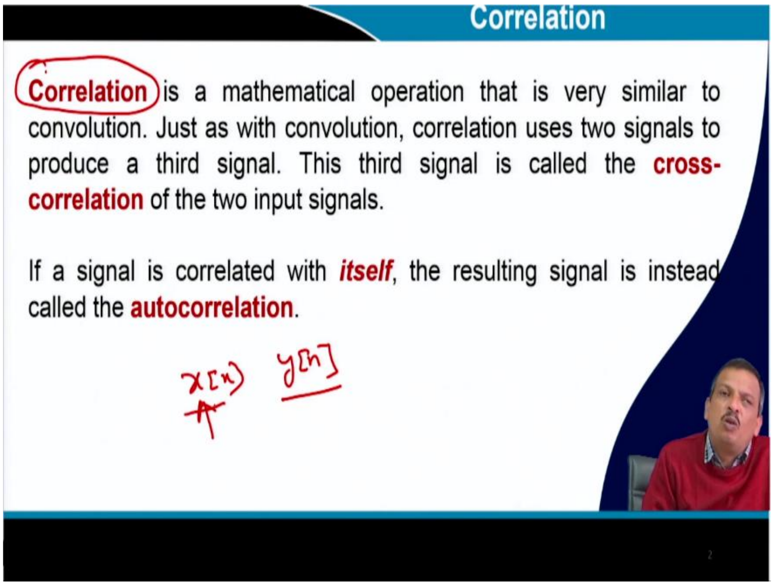
Signal Processing Techniques and its Applications
Prof. Shyamal Kumar Das Mandal
Advanced Technology Development Center
Indian Institute of Technology, Kharagpur

Lecture - 11
Correlation

So, today, we will talk about Correlation. We have already studied that LTI system and that convolution. Today, we talk about correlation, which is another kind of operation that looks like convolution, but instead of holding, we just slide the sequence. So, why is this correlation important? Why this? What is correlation, and why is it required?

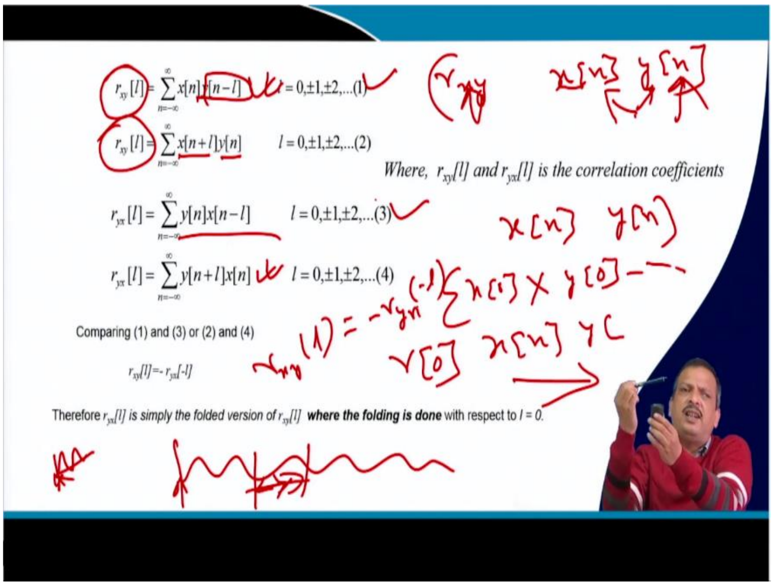
(Refer Slide Time: 00:47)



So, as you have heard, the word correlation means the relations between the two signals. So, suppose I have a signal, and I have another signal; if I want to find out the similarity between the two signals, that is called correlation, a mathematical operation very similar to convolution. But it is not that a signal is convolved with another signal. It is a measurement of the similarity.

So, if I have a signal two signal when I say the correlation, I want to determine the similarity measurement. So, correlation is another signal that is derived from those two signals, which indicates the similarity between the two signals. So, there is if the signal supposes one signal is my x[n] and another signal is my y[n] when I say the similarity between the two signals, then it is called cross-correlation. When I say the similarity between the, itself, then it is called autocorrelation.

(Refer Slide Time: 02:06)



So, let us define it mathematically. So, correlation, so r_{xy} is the measurement of the similarity between the two signals, x and y. So, I have two digital signals: one is x[n], and the other one is y[n],? I want to find out how similar y[n] is to x[n]. So, I want to find out how similar y[n] is to x[n]. So, how do we measure similarity? So, similarity measurement will be indicated by r_{xy}, so similarity of y with x r_{xy}.

So, how do I do that? So, if you forget about the signal, think that I told you the similarity between this object and this object. Or you can say that if I had to ask you what the similarity between you and your friend is, you might say, sir, my hairstyle and my friend's hairstyle are the same. My[n]ose tip and my friend's nose tip are the same. They may[n]ot be similar, but they are the same. So, how is it similar?

If I said the hairstyle is similar, is it exactly similar, or how similar is it? So, if I want to measure, then what will I do? If one thing is there, I will superimpose another thing, and if they are similar, they will be exactly matched. So, if I say mathematically, I want to do it. So, I have a signal x[n], and I have a signal another signal y[n].

If let us say the sample one x[0] is similar to y[0]. So, if both are similar, how can I do that? I can take the product and sum up all the sample similarities. So, that will be integrated that x[n] is similar to y[n]. So, the similarity index[n]umber is 0, so l is equal to 0. Now, you may say, suppose y[n] is a large signal and x[n] is a very small signal. So, let us say the length of the x[n] is m and the length of the y[n] is n. So, m is less than n or m is greater than n, x[n] is large, and y[n] is small.

So, now, if I say only r[0], then it may not be similar at the initial position, but it may be similar here; inside the signal, there exists some portion where it is exactly similar. So, how do you do that? So, for that purpose, what should I do? I suppose I want to measure the similarity of x[n] in the signal y[n]. So, what I said is how similar x[n] with y[n]. So, what will I do? I will make the x[n] is fixed, and I will slide the y[n]. So, what will I do? I will let shift x[n] y[n-l].

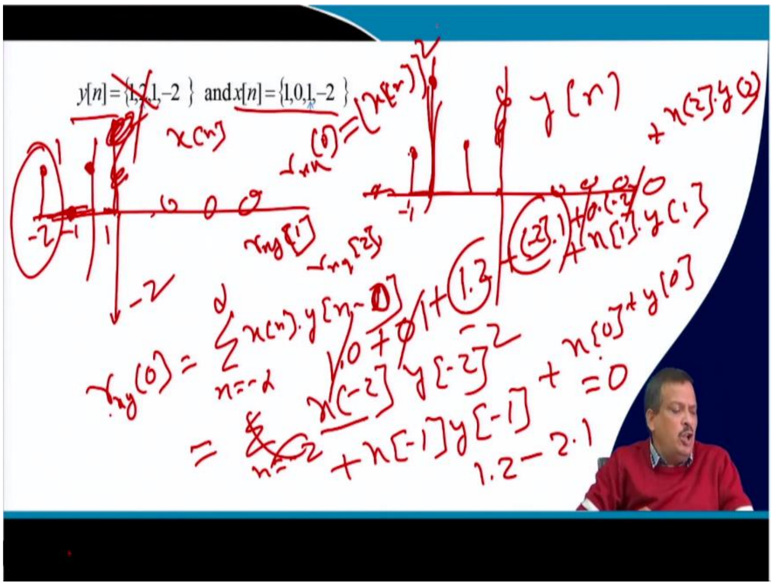
So, let us say this is one signal, and this is another signal. So, what will I do? I will first calculate the r[0], then I will shift this portion to this site and find out the correlation. So, I first calculate r[0], then r[1], then r[2], then r[3], then r[4], then r[5]. So, I slid this y[n] over this left side and found out the similarity. I can do the same thing: let this y[n] remain constant, and I can switch that x[n] on this side, so this is the right side shift. So, I can do the same things with a x[n+l] y[n].

So, how do I find the similarities? I have a two-signal. There are two signals. So, this signal is y[n], and this signal is x[n]. So, I will first put this y[n] with 0 0. So, l equal to 0, so both signals are here. Now, I can shift either y[n] on this side or x[n] on this side. So, what is the operation when I shift y[n] this side? Left shift operation. So, I can say y[n-l] while I am shifting this one, but while I am shifting this one, this one means x[n+l].

So, I am shifting it to the left side, and I am shifting this to the right side. So, both way I can calculate the correlation of x with y, r_{xy} correlation of x with y. Similarly, I can say the correlation of y with x, so I just replace the x with y and y with x; the equation is the same.

So, let us say that I can compare this one with this one. So, if I compare this one and this one, both are shifted this side, I can compare. So, if I compare this one. So, it can be proved that r_{xy}[l] is nothing but a -r_{yx} [-l] ok. So, this is the relationship, but how do I calculate the correlation; the same procedure is the same mathematical procedure is same. This is y[n], this is x[n]. Either I can shift this one, or this one to this side.

(Refer Slide Time: 08:40)



So, let us do an example. Suppose I have a signal y[n] and I have a signal x[n], and the arrow indicates the 0th position. So, if I plot x[n], so this is the 0th position, the sample value is 1, then x[n] has a minus 1, 0, 1, minus 2 it has the value 1. Similarly, 0 is 1, and 1 1 is 2, so this is x[n]. Now, what is y[n]? y[n] minus 1 I have a value 1, 0 I have a value 2, 1 I have a value 1, 2 I have a value minus 2. I will take an eraser. This will be not here. This will be this side minus 2 here. also, this is not 2 plus 2. This is minus 2. This is ok.

So, now what I do? I want to find out r_{xy} 0, so 0th position, l equal to 0. What is the equation? My equation is n equal to minus infinity to infinity x[n] multiplied by y[n-l]. So, I am shifting the y signal with the x signal, so I will impose the y signal with the x signal and shift it and find out that r[0]. So, r[0] is no shift l equal to 0.

So, it is nothing but a minus infinity to infinity sum of the product of the sample and then sum. So, now I can see I have a signal x[n] with a value sample value minus 2; I have a value[1] and minus 1, I have a value 0. So, I can say this: I can see that x[n] minus, so maximum minus 2 I can say, so I can say[n] equal to minus 2 to n equal to 2, or you can also be infinite.

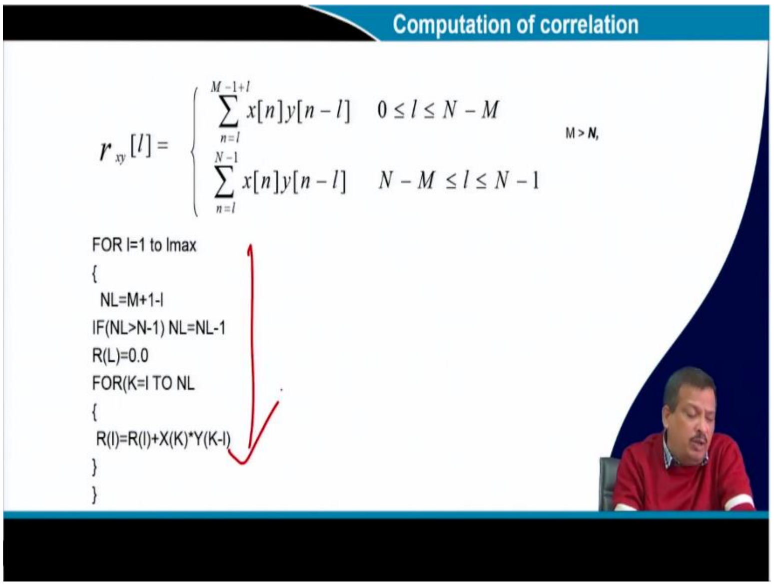
I am doing this because the signal is 0 outside this. So, I can easily say r x of y 0 means x[-2], then y of this is y[-2]. Because l equals 0 plus x[-1] y[-1], plus x[0] plus y[0], plus x[1] into y[1], plus I can say x[2] into y[2]. So, x[-2] I have a value 1 x y[-2] I have a value 0.

So, 1 into 0 plus x[-1] I have a value 0, y[-1] I have a value 1, plus x[0] I have a value 1, x[0] I have a value 1, y[0] I have a value 2, 1 into 2. Plus x[1] I have a value x[1] I have a value minus 2, y[1] I have a value 1, plus x[2] I have a value 0 into y[2] I have a value minus 2.

So, if I see all are 0, only these two terms will be there. So, it is 1 into 2 minus 2 into 1, so I can say it is 0. Similarly I can calculate r_{xy} 1 you can calculate r_{xy} 2 like that way. So, you do it with your paper pen and paper and see whether it matches or not. So, those are the measurements of the similarity between the two signals; if it is x and y, it is called cross-correlation. If I say r x with x slash, the y, y signal is not there; find out the similarity; r x with x, which is called autocorrelation.

So, when I say r x with x, r x with x 0 is nothing, but each of the samples will be multiplied and added up. So, I can say it is nothing but a x[n] whole square, which is called the energy of the signal. So, if it is autocorrelation, then r x x x 0 is nothing but the energy of the signal.

(Refer Slide Time: 14:13)



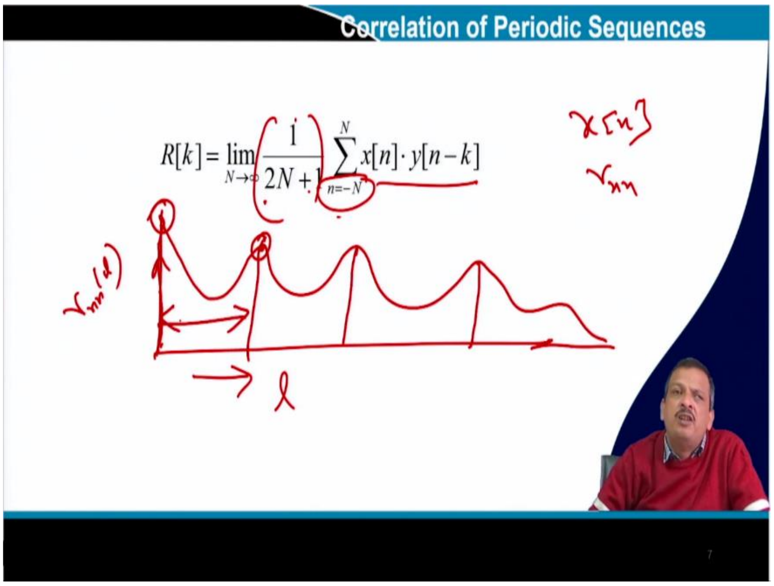
Now, you may say, why should I require this kind of autocorrelation, cross-correlation? Why should I find out that? What is the use of correlation? The main purpose of correlation was to find out the similarity. So, using that similarity to find out what we can do, suppose I give you a periodic signal and you determine the period of the signal.

So, what is the periodic? If the two signal if this signal has a signal, I have to determine the period of the signal. So, what can I say? If the signal is periodic, then our autocorrelation value will be maximum at 0 and at a periodic. That is why autocorrelation is used to find out the periodicity of the signal. Similarly, if you are reading about radar signal analysis, suppose how a radar signal finds a target.

So, you know that a periodic signal is transmitted by a carrier when it is reflected back due to the Doppler effect, there will be a delay in the periodic signal or phase change in the periodic signal. So, if the phase changes, I can find the similarity, and if there is a phase change, the similarity index will be changed if it is exactly the same signal I receive. So, I can see that if I find the correlation, it will be very high.

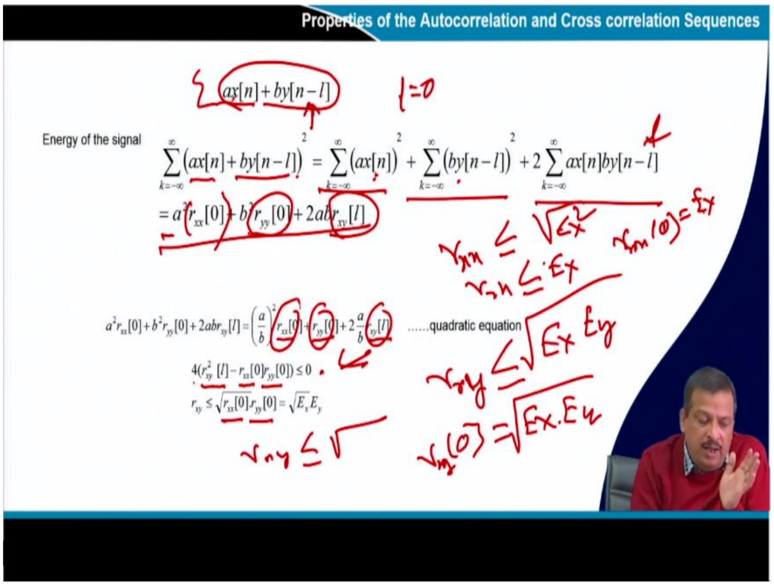
If the similarity is changed, then I can say there is a shift that happens, so that is why autocorrelation is used to find out the similarity. Now, autocorrelation is used in speech to find out the periodicity. So, now, you can use this algorithm or directly from this math, you can write down in C code or MATLAB code to find out the autocorrelation of two sequences.

(Refer Slide Time: 16:36)



Now, what are the correlations of periodic signals? Let us say cross-correlation or autocorrelation of a periodic signal if x[n] and y[n] both are periodic signals. Then it is nothing but a, a normalization factor limit n tends to infinite 1 by 2 N plus 1 n equal to minus N to N x[n] y[n] minus k, which is defined as Rk. So, this we use in many purposes for find out the periodicity of a signal.

(Refer Slide Time: 17:10)



Then there are properties of autocorrelation and cross-correlation sequence; suppose forget about this math's only thing. Suppose I have a signal a x[n] plus b y[n-l]. Now, what is the energy of this signal? This is a total signal. What is the energy? Energy is nothing but a sum and square and sum and square and sum, so square and infinite sum. Then I said this is a, this is b, so a plus b is the whole square; a square plus b square plus twice ab.

So, I can say this is nothing but a square, and this is square; that means r_{xx} 0 l equal to 0. This is nothing but a r_{yy} 0 and this is nothing but a r_{xy} l, because l is the shift. So, this is a quadratic equation if I divide this by 1. So, I can say a by b, the whole square plus r_{xx} 0 plus r_{yy} 0 plus 2 a by b r_{xy} l. So, this is a quadratic equation of let us know those are the coefficients. So, I know that I can from here I can find out 4 into r_{xy} square, minus r x 0 r_{xx} 0 and r_{yy} 0 must be less than equal to 0. That means, I can say r_{xxxy} is less than equal to root over of r_{xx}x and geometric mean.

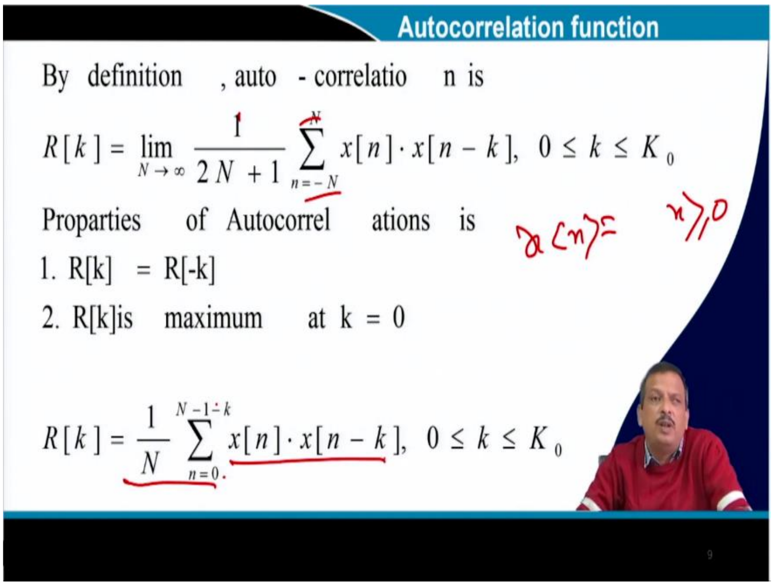
So, any time I can say the correlation coefficient r_{xx}, r_{xy} must be less than equal to the energy product or geometric mean of the two energy of the signal. E x is the energy of the x signal, and E y is the energy of the y signal.

When will it be equal? When I said r_{xy} 0 that time it will be equal to E x multiplied by E y root 2. But otherwise it will be less than this. So, if I say instead of$_{xy}$ it is autocorrelation then I can say r_{xx} must be equal greater than equal to root over of E x square y is replaced by x. So, r_{xx} must be greater less than equal to E x energy of the signal. So, when it is equal when r_{xx} 0 is equal to E x elsewhere, it will be less.

So, I can say if I compute autocorrelation r_{xx}, if I compute autocorrelation let us say I compute let us say I have a signal x[n]; I compute autocorrelation r_{xx}. Now if I plot this axis is the l and this axis is the r_{xx} l. So, I know if l is equal to 0, it will, and the value will be maximum; elsewhere, the value will drop down. So, if you see that the signal is periodic, then you can get this kind of plot. That means the next peak is not exactly equal to the initial peak, but the same peak will happen during the period.

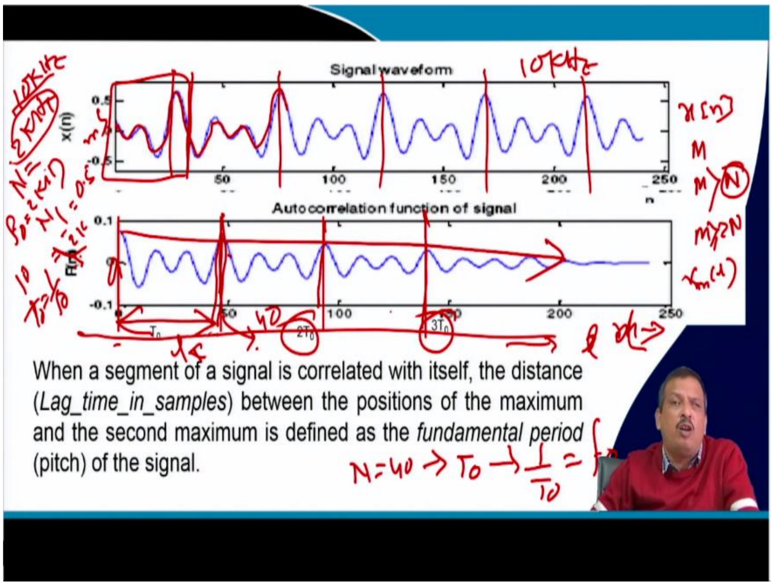
So, this is called the period of the signal details I will give you; I will show you the details here in speech signal, ok? So, this is the autocorrelation. So, I can use this property of the autocorrelation or correlation property to compute the periodicity of a signal.

(Refer Slide Time: 21:40)



So, this is if the signal is periodic, then it will be minus N to N 1 by[n]. So, I know that none of the signals if the signal x[n] exists only exist when the n is greater than equal to 0, then I can say that it is 1 by[n], n equal to 0 to n minus 1 k x of x[n] into x[n] minus k that is autocorrelation.

(Refer Slide Time: 22:10)



So, suppose an application, suppose I told you, can you find out the periodicity of a vocalic speech signal? That means, let us say this is a speech signal collected from humans. This is a vowel, which is periodic. You can say if you observe the signal, there is a periodicity in the signal; I can say this point to this point is a period, this point to this point is a period, this point to this point is a period. Now, I told you, can I find out the periodicity of the signal using correlation? Yes, if this is my x[n] and if I compute auto, this is the periodic signal.

So, I said take the M number of the sample, so M is greater than N where N is the length period, and N is the period, you know that. Suppose I have a signal. Yes, I will come to that practical scenario. Suppose I have told you that I have a signal whose sampling frequency is 10 kilohertz and let us say the signal has a frequency of 2 kilohertz. So, what is the length of the period? N, what is the value of N?

So, within 2 kilohertz, so 10 kilohertz is my sampling frequency. So, the signal period is f 0 is 2-kilo hertz f 0 is equal to 2-kilo hertz, so I can hatch two number sample numbers I can easily convert. Because I know 10 kilo hertz means 1 10 k sample in one second, there will be a 10 k sample. Then, in f 0, if the f 0 is 2 kilohertz, that means T is T 0 is equal to 1 by f 0, and One by f 0 means 1 by 2 kilohertz, which means 0.5 milliseconds.

So, in one second, if there are 10000 samples at 0.5 milliseconds, how many samples will be there? I can find out. Understand or not? So, that number of samples is nothing but a period of that signal. So, I have to find out the period of that signal, so n. So, let us say I compute this, so my M the I have taken the signal sample which is which contains more than one period.

Suppose, if I take this amount of signal, I cannot find out the period because it does not contain at least one complete period. So, that is why I take the M, which is much greater than N. Maybe 2 to 3 period is at least 2 to 3 periods will be there. M will be at least greater than equal to 2 times of at least 2 times of n.

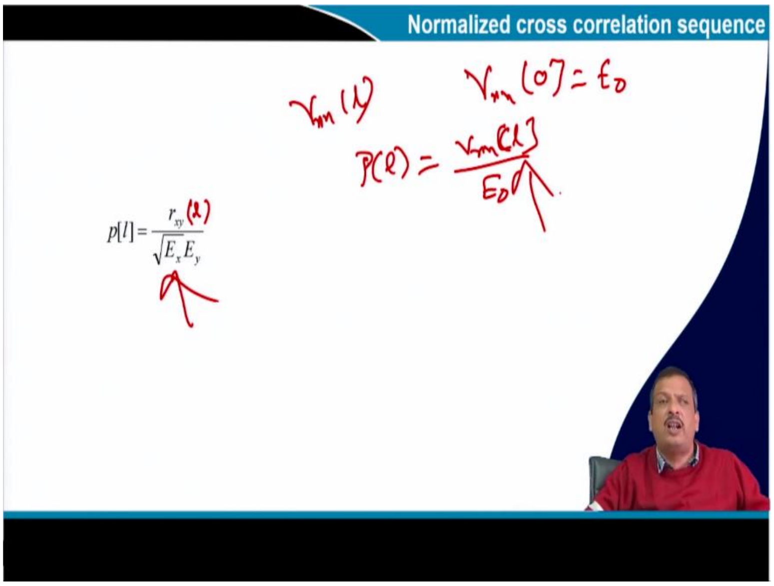
If it is there, then if I compute r_{xx} l. So, this is auto correlation r rm that is nothing but a r_{xx} l I compute it autocorrelation I know x[n] I multiply that things and compute it. And if I plot it again this axis is the l or I can say here it is m, this axis is m small m and this axis is the value of r_{xx}.

Let us say this is called l. So, this axis is l and this is a r_{xx} l, then you get this kind of plot; if you see it is going decay down. Now, I can say the maximum peak next peak can happen when it is a period. So, I can say this to this length this much of l is the period. So, if I know that l if I detect this peak and find out the index, let us know. The index is 40, let us say.

So, if this signal is sampled at 10 kilohertz I get the next peak at 40. So, what it means is that this 40 is the periodicity, so I can easily convert it to the hertz frequency. Because this is time N, which is equal to 40, that can be converted into T 0 once I get T 0, 1 by T 0 is equal to f 0; I get the fundamental frequency. So, I take the signal compute the autocorrelation and find out the next peak.

So, that peak gives me the period. I understand now the problem is the amplitude is decaying if I want to make the amplitude constant. So, this is f 0, so this is twice f 0. So, that is why 2 T 0 this is thrice f 0. So, now, if I want to make that amplitude constant, that is called normalization.

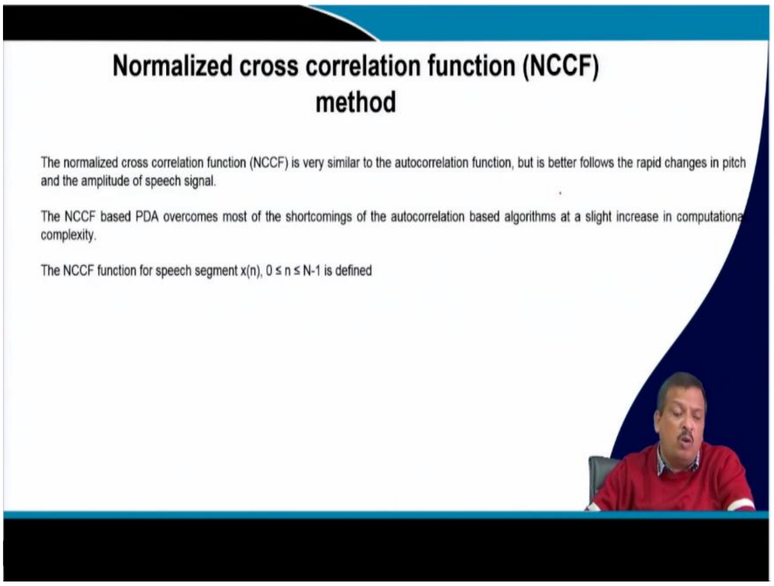
(Refer Slide Time: 27:29)



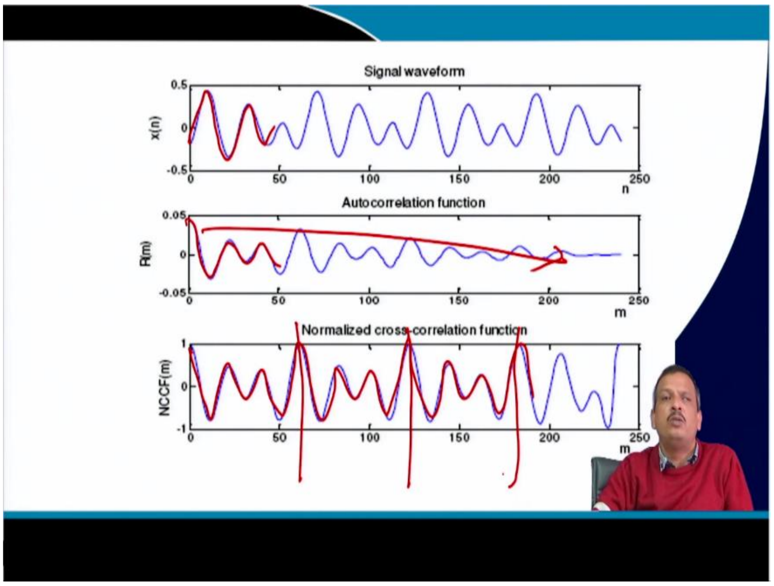
So, which is called normalized cross-correlation or normalized autocorrelation, how do you do the normalization? Forget about the theory that is written here. So, how do you do the normalization? So, I compute r_{xx} l, so I know r_{xx} 0 equals E 0. So, if it is autocorrelation, I can say normalize autocorrelation coefficient. Let us say P l is equal to r_{xx} l divided by E 0, if it is cross correlation_{xy} signal.

So, r x, so x y l divided by root over of E x into E y, the geometric mean of the two energy of the two signals. So, that is normalized cross-correlation, and this is normalized autocorrelation. So, that way, we can calculate the normalized cross-correlation and autocorrelation.

(Refer Slide Time: 28:35)

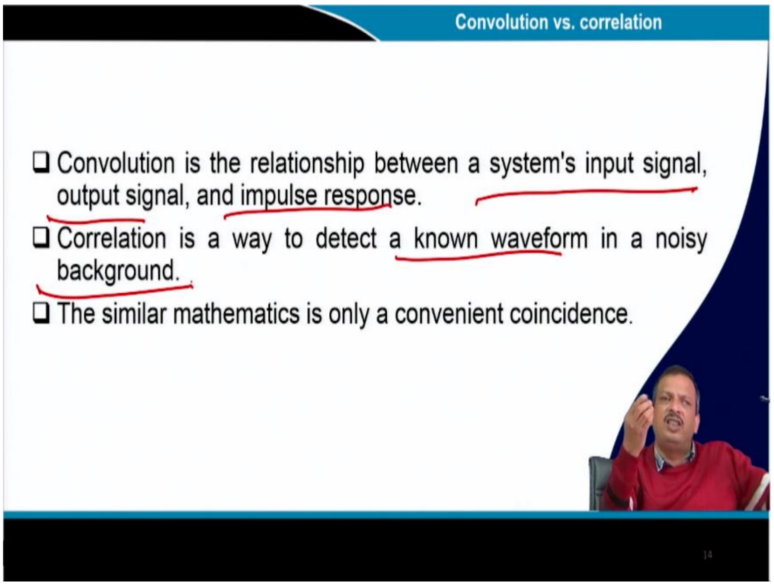


(Refer Slide Time: 28:36)



Once you normalize it, then if you see this is my signal, this is my autocorrelation without normalization; amplitude is decaying. Now if I normalize it, then if I plot it if you see the amplitude is not decaying. So, this is the f 0, this is twice f 0, this is thrice f 0. So, this is the requirement for correlation and this is a procedure to compute correlation. Is ok?

(Refer Slide Time: 29:15)



But you slide, you can say what the differences are. Suppose somebody asks you what the difference is between convolution and correlation. So, convolution is the relationship between a system input signal output signal and impulse response. I know the output signal is a convolution of the input signal and impulse response.

Correlation is an I have to detect a known waveform in a noisy background correlation only operation we do not perform, that is folding. In convolution, we fold it and then shift it here without folding, there is just one signal, and there is another signal; either we slide this one, or we slide this one. This is the difference between correlation and convolution.

Thank you.