Advanced Digital Signal Processing – Wavelets and Multirate
Prof. V.M. Gadre
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture No. # 01

A very good morning, let me introduce myself first, my name is Vikram M. Gadre and I am a professor in the Department of Electrical Engineering at IIT, Bombay. I am going to deliver a series of lectures on the subject of wavelets and multirate digital signal processing. In this first, lecture or more appropriately exposition, I would like to bring before you the reason why we should be studying this subject or rather the reasons, the inspirations that have driven researches and scientists to understand the concepts of wavelets and multirate digital signal processing, the broad themes that we are going to address in this course and finally, a broad outline or a structure of this course.

So, before I begin a formal statement of structure it would be appropriate to put before you the thrill or the fascination of wavelets that I have or the inspiration that leads me to look deeper at this subject, which I would also like to convey to you in a few words and inspire you to study this subject in depth. Typically, this subject follows a basic exposition to signals system theory perhaps digital signal processing or discrete time signal processing and other basic courses in communication and signal processing.

So, I would pick this course at an advance level in signal processing, I do not intent to convey from this that the concepts are difficult to understand as you will see in some of the subsequence lectures the concept are actually very easy very simple perhaps even simpler than some of the concept that we learn in a basic course. In a basic course the whole subject is new to us, when we study signals and system the whole idea of abstraction is new abstraction of a signal, abstraction of a system, abstraction of transform, abstraction of a domain.

So, we assume that we have crossed that step, we have overcome that that obstacle, we have agreed that abstraction is good things in that sense this course is simple. We are moving one level higher where we are moving a little closer to reality. In a basic course on signals and systems or in a basic course on discrete time signal processing we tend to be idealistic, we tend to want to over simplify. In this course, we will refrain from doing

that at least in Toto. What do I mean by this? In a basic course on signals and systems, we assume that signals last forever. For example; when we talk about the Fourier transform, we think of basis functions or functions on which the transform is based, which start from time T equal to minus infinity and go up to T equal to plus infinity, T denotes time.

In this course, we shall recognize that this is a silly thing to do. No signal in the world at least no signal that we can deal with realistically is going to last forever, it is going to have its life time, it is going to begin somewhere end somewhere. After all the analyzer the person who analysis is also going to begin somewhere and end somewhere. So, the first thing we understand is we must deal with finite domains. We understand the finite domain as far as the natural domain goes very well, what I mean by the natural domain is, suppose for example; we are talking about an audio signal, the natural domain is time, so an audio signal begins somewhere in time ends somewhere in time, we do not listen to a piece of music forever.

But for that limited period of time in which we wish to listen to a piece of music we would like to understand better the content of the music to which we listen, we also wish to be able to develop the ability in a signal processing setup to enhance what we want out of the music and to suppress what we do not. We also wish to be able to characterize a system, which does so and all the while recognizing that we are not going to deal with the signal forever. We are going to deal with it over a limited range of the natural domain. So, here as I said an audio signal the natural domain is time.

Let us take another example of a natural domain, suppose I wish to deal with a picture, I wish to take a picture in which I have a phase and naturally many facial features, the eye brows, the forehead, the nose, the lips and other features that are associated with a typical face. When I wish to isolate a certain feature, I would imply being able to localize in the special domain. Now, this is an example of a two variable domain, unlike the other example of audio where it was one variable domain only time, here I have two special variables. Localization is the common theme let me spend a couple of minutes in explaining that a little better.

Suppose, I take a piece of audio, let us assume that in that piece of audio a number of notes are sung. If I take recourse to the Indian system of description you could have a

raga in which there are several different notes the components of the raga. Suppose, I wish to build a signal processing system that takes the rendition of this raga and identifies the notes that compose it, the different frequencies so to speak that come together and play to form that piece of audio. What do I need to do in such a circumstance for a signal processing perspective or from the point of few of analyzing signals? One thing is obvious; I need to segment in time. So, I need to be able to say that in this part of that time axis I had this note prominently played perhaps only that notes was played, in a different part of the time axis a different note was played.

Now, when I see different parts of the time axis they not only separated, but also might be different in lengths, so one note could be longer and another note could be shorter, it is not just a question of which notes had been played, but also for how long. Now, what exactly do we mean, when we talk about notes, if we take a signal processing perceptive I mean. After introduction to the fundamentals of signals and systems or even for that matter basic course on discrete time signal processing all of us are comfortable with the ideas of a frequency domain. So, we accept that we can think of continuous signals or discrete sequences has having embedded inside them a collection of sine waves. If it is a continuous signal that we are talking about we have continuous sine waves, if it is a discrete sequence that we are talking about we have sample sine waves.

And the whole philosophy of the Fourier transform is most reasonable signals that we deal with can be thought of as comprised of or composed of a collection of sine waves. In fact, in principle if the signal is not periodic, then we are talking about infinity of sine waves with frequencies ranging all the way from zero to infinity. If it is a periodic signal that we are talking about then we have a discrete set of sine waves possibly infinite possibly finite, which we call the Fourier series representation. Anyway, what I am trying to emphasize here is that there is a different domain to which we go when we wish to analyze a signal better.

So, in the language of signal processing, if I were to take that raga, that rendition of the audio piece the elementary audio segment as understood in Indian music and if I were to query what notes are been played I am equivalently asking what is the frequency domain content of this audio piece. If I visualize a frequency axis, what points on that frequency what points on that frequency axis are occupied, what are the locations where the transform is predominant; is prominent?

Now, this is where the whole story starts and perhaps is the most fundamental inspiration for this course on wavelets and multirate digital signal processing. I first talk about where wavelets come from; well Fourier transforms deal with waves, sine waves to be more precise. We recognize the merits of sine waves; sine waves have many nice properties, for one they occur naturally in many different circumstances. For example, an electrical engineer recognizes the sine wave as naturally emerging from an electricity generation system when there is an electromagnetic induction. If there is a perfectly circular rotating device in a magnetic field and if all is perfect in the generating system we would be generating a perfect sine wave from the brushes.

So, sine wave is good idealization to work with for an electrical engineer, but that is not the only point, sine waves are in some sense the most analytic the smoothest possible periodic functions. They also have the power of being able to express many other waveforms that means, they form a good basis from which many other waveforms can be generated. They have many other nice mathematical properties, if I take two sine waves with the same frequency possibly different amplitudes and phases, I would get back a sine wave of the same frequency of course, with a third amplitude in phase.

If I differentiate a sine wave I get back a sine wave in the same frequency and naturally if I make a combination of these operations namely differentiations or even integrations for that matter and linear combinations. And if I restrict myself to sine waves of a particular frequency I remain within the domain of sine wave of that particular frequency this is something beautiful about sine waves, it is not easy to find waveforms, which obey this.

And as I said before sine waves form a good basis, so they form good building blocks for being able to express a wide variety of signals. For all these reasons the sine wave has been very popular in a first course on signal systems discrete time signal processing and what have you communication. But as I said right in the beginning of this lecture one of the reasons why we are not so happy with sine waves is that they need to last forever, begin from minus infinity and go all the way to plus infinity. Otherwise, if you truncate a sine wave if it is one sided for example, and if you look at what happens to an electrical system when you apply one sided sine wave, by a one sided sine wave I mean, a sine wave which starts from some point its zero up to then and start from some point and continuous afterwards the sine wave.

The response is very different from what would be the response for a sine wave that started at the minus infinity in general, there would be transients which are not periodic and, then all these beautiful properties of sine waves and their responses go away. So, if I really wish to be able to apply the basic principles of signal systems and discrete time processing that I have learnt in a basic course, I need something unrealistic, I need a sine wave that last forever. How can I be more realistic in my demands by accepting that I cannot deal with waves, but it is more appropriate to deal with wavelets so that is where the word wavelet comes from, small waves.

Waves that do not last forever, function that are not predominant forever they are significant in a certain range of time perhaps only exist in a certain range of time and in significant outside. So, they have a certain support over which one might want to use them, one might want to consider them to exist and so on. Much more realistic assumption and that really is what we call a wavelet not a wave, but a wavelet. For example; you could if you wish think of truncating a sine wave to a rectangular region that means, suppose we take a sine wave to last from 0 to 1 millisecond as an example, it could be an example of a wavelet. Later we will see this is not a very good example, but in principle a wavelet, a wave that does not last forever. A simplistic explanation of what wavelets means, but that is not the whole story, our whole objective was to talk about the other domain too.

So, going back to the example of the audio signal, if I thought of the audio signal as comprising of many sine waves to come together to form an audio piece, then I wish to be able to do something simultaneously in two domains and that is the key idea here. So for example; to put it in plain language I should be able to say well there was this two second audio clip out of which there were five notes being played, each note was played for different intervals of time may be the first node was played for 0.4 seconds, the second note was played for 0.7 seconds, the third note only for 0.2 seconds and so on. So, I need to be able to segment in time, but when I am talking about being able to identify notes I am also talking about being able to segment in frequency and low in the whole that is where the conflict arises.

A very basic principle in nature says, if I wish to be able to segment in time and frequency simultaneously I am going to run in to trouble. Nature does not allow it beyond a point and that is something very fundamental, it pops up in many different

manifestations in different subjects. In modern physics, they call it the uncertainty, the uncertainty of position and momentum, in signal processing, we call it uncertainty, the uncertainty of the time and the frequency domain. So, to put it simply though not very accurately, the shorter you play a note the more difficult it is to identify it, not very far from intuition. If you play a note for a long time and listen to it for a long time you are likely to be able to indentify better common sense tells us that.

But what common sense does not tell us is that you can never quite go down to identifying one particular frequency precisely. So, if I wish to be able to come down to a point on the time axis, then I need to spread all over the frequency axis and if I wish to be able to come down to a point on the frequency axis, I need to spread all over the time axis. That is of course, a strong version of this restriction, but there is a weaker and a little more certain version and that is as follows. Even if I am not quite interested in coming down to a point on the time axis, I am contend with being in a certain region as I said in the first 0.4 seconds out of the 2 second clip, I was playing note number one that means some frequency number 1. I would be able to say this at least to a certain degree of accuracy that is what I am trying to point out here, what the principle of uncertainty tells us is that this can be done to a certain degree of accuracy, you can identify that note to a certain degree of identity of accuracy.

Well what uncertainty also tells us in a more certain form is that; if I even chose to relax to a certain region of time, so I say well in this region of time tell me the region of frequencies which are predominant, even then there is a restriction on the simultaneous length or measure of the time and frequency regions. And of course, they have a tussle with one another, the smaller I make that time region the larger that frequency region becomes, that means, the more I want to focus in time the less I am able to do so in frequency. This is indeed something that arises a lot of thought, it may seem something far from our interests at first glance, but when we look at it carefully we realize it is something very fundamental to what we often desire and that is what now, I am going to explain to you with a couple of more examples.

We live in an age where we use mobile telephones, in fact more fundamentally digital communication. What are we asking for in digital communication when we look at it from a signal or system prospective or transform domain perspective or time and frequency perspective? Going right down to brass tacks, what we are asking for in digital

communication is I should be able to transmit a sequence of bits binary values 0 or 1. And how do I transmit that sequence of binary values, I chose may be one of two possible waveforms in the simplest scheme. So, corresponding to 0 I have one waveform, corresponding to the bit 1 I have a different waveform, to make life simple the two waveforms have the same time interval.

So, for example; we talk about I mean all of us here about computer networks and they talk about the speed of the network. So, they say well this network can operate at a speed of 1 mega bit per second, what does it mean? It means in 1 second I can transmit 10 raise the power of 6 bits. So, you have 1 milli; 1 microsecond to allow each bit to be transmitted. Give it a thought, here we are talking about time, what are we saying about frequency; now, let us go the mobile communication context. I have so many different mobile operators, obviously each operator will want its own privacy. So, what is been communicated on the network of operator one? Should not interfere with what is been communicated on the network of operator two.

Now, where is the separation going to occur not in time, after all there are many different people simultaneously using mobiles bought from both of the operators. So, the separation cannot be in time, we may argue the separation can be in space. So, in one region you may have mobiles from one operator and another region in space I mean mobiles from the other operator, so far so good. But that is also not always true; it is very common to see mobiles purchase from different operators, operating in the same room, there is no separation in time, no separation in space. So, where the separation then, the separation is to be in a domain, which is not so easy to see, but once we have done a course on signals and systems reasonably easy to understand and that domain is frequency. So, we say well operator one has this bandwidth allocated to him, operator two has the other bandwidth allocated to him.

Now, when we say this bandwidth may be a certain region of the frequency axis of size and let us say 2 mega hertz. When we say this region of 2 mega hertz is allocated to operator one and another region of 2 mega hertz is allocated to operator two, are we not talking about a segmentation in a different domain. In fact there we are talking about simultaneous segmentation; we have a segmentation in time, because you want to transmit different bits in different time segments. And you want to have a separation in frequency, because what is transmitted by operator one should not interfere with what is

transmitted by operator two. So, here is a very common, though not so obvious example of simultaneous desire for localization in time and frequency.

And, then the audio example which is of course, a little more obvious, a little easier to understand, this example is equally common at least in the scenario today, but perhaps not so easy to understand, but a little reflection makes it very clear to us, there is a desire to localize in two domains simultaneously.

Well let us go to a biomedical example, very often when one analysis an electrocardiographic waveform, what one wishes to identify are features in these ECG signal. Now, I do not intend to go into the medical details, but there are different segments in a typical ECG signal, they are often index by letters P Q and so on. Without meaning to focus on specific details of an ECG signal, let us try and understand the connection to time and frequency localization. When we talk about an ECG signal all features are not of the same length in time, some features are kind of shorter, some features are longer. In fact to go away from an ECG signal biomedical engineers often talk about what are called evoked potentials. So, we provide a stimulus to a biomedical system or to a biophysical system and we evoke a response and the waveform corresponding to that response is called evoke potential, it can be studied as an electrical signal.

Now, the evoke potential again typically has quicker parts in the response and slower parts in the response. Naturally, we expect the slower parts of the response would be predominantly located if you think of the frequency domain in the lower ranges of frequency and the quicker parts of the evoke potential waveform would be located in the higher ranges of frequency. Here is again an example of time frequency conflict, suppose I wish to be able to isolate the quicker parts, is it all right simply to isolate the higher frequency content in a certain signal and which comes from an evoke potential and suppress the low frequency part.

Well you see if we try and suppress the low frequency part, then we have already suppressed the slower parts of the response and if we try and suppress the higher frequencies in a bid to emphasize the slower part of the response we have suppressed the quicker parts of the response. So, if we think conventionally in terms of the frequency domain, may be high pass filtering or low pass filtering nothing works for us. If we do

high pass filtering, then we have effectively suppressed the slower parts of the response, if we do low pass filtering we have suppressed the quicker parts of the response. So, we need a different parading or a different perspective on filtering, we need to identify in different parts of the time axis, which regions of the frequency axis predominant.

And therefore, in a certain sense identify different parts of the frequency axis to be emphasized in different time ranges. This is another perspective again on the time frequency conflict and all this is going to lead us in the direction of building up of this course on wavelets. We shall of course, understand some of these concepts little better as we progress in the lectures, but for the time being I have given you these three examples with the intent of bring before you. Perhaps not completely, but at least in a way to inspire your imagination the whole idea of time frequency conflict or more generally the conflict between two domains, domains of analysis and representation of a signal and of course, then going further of even a system.

In a first course, we understand the domains very well, we understand there is a time domain, we understand there is a frequency domain, we do well because we keep them apart it makes life easier. But what we are trying to bring out through these three examples; the audio example, the digital communication example, and the biomedical waveform example, whether it is electrocardiographic waveforms or the evoke potential waveform. What we are trying to bring out is that one normally needs to consider the two domains together time and frequency, and when we try and do so there is a certain very fundamental conflict that we have to deal with that conflict call uncertainty appears as I said in different manifestation in different subjects.

And we are going to look at that principle; the uncertainty principle as it applies to signal processing in great depth at a certain stage in this course. But before that we are going to consider one particular tool for analyzing signals, analyzing situations with a recognition that we need to be local and not global. So, we are going to build up the whole idea of wavelets by starting from what is called the Haar multi-resolution analysis. Haar incidentally is the name of a mathematician, call him a mathematician, call him a scientist what you will, but one of the beautiful things that this gentleman proposed was what is called a dual of the idea of Fourier analysis, What do we do in Fourier analysis? We allow even discontinuous waveforms, we allow non-smooth waveforms and we

convert them into a sum or a linear combination of extremely smooth functions, namely the sine waves.

Haar said can we do exactly the dual, can we in principle take smooth functions and convert them in to a linear combination of effectively jagged or discontinuous functions, why on earth would one like to do something like that? Again, let us deflect for a minute. A few years before this might have seen silly to do, but today it is not, what you are doing when you are doing digital communication, you are transmitting audio, you are transmitting pictures. And you are doing all this actually with a large level of discontinuity, how does one record digital audio? One firstly samples, so one takes values of the audio signal at different points in time, one digitizes them and one then records those digital values as a stream of bits. All of these are highly discontinuous operations, you are forcibly introducing discontinuity time around top of that you are introducing discontinuity in amplitude by quantization.

So, one thing to represent the beautiful smooth audio in terms of very discontinuous bit streams is very very beneficial to digital communication. And in fact, none of us complain when we have a good digital audio recording, sometimes we even say that digital audio recording is better than an analog recording as we had in the past. So, going from smooth to non-smooth has it place in modern communication and signal processing. And when Haar proposed that one should look at the whole philosophy and the whole principle of being able to go from smooth to non-smooth perhaps he was looking into the future, when this will be absolutely essential.

What we are going to do? In the very first few lectures immediately following this, is to look at one whole angle of wavelets and multirate digital signal processing based on the principles that Haar propounded. So, we are going to look at what is called the Haar multi-resolution analysis, and in fact if we understand the Haar multi-resolution analysis in depth. We actually land up understanding many principles of wavelets, many of the essentials of multirate processing specifically what is called two band processing very well. So, we shall draw upon the Haar multi-resolution analysis to understand some of the basic concepts that underlie the scores, and of course, build upon them further later on.

From the Haar we shall progress to slightly better multi-resolution analysis better in what sense we will understand and there are many different families of such better multi-resolution analysis, one of them being what is called Dobash family. Dobash is again the name of mathematician scientist, who proposed that family of multi-resolution analysis. As I said at a certain point in the course immediately following this, we shall then look uncertainty principle fundamentally and in terms of it is implications, from there we shall move to the continuous wavelet transform. So, in the Haar multi-resolution analysis, we have a certain discretisation in the variables associated with the wavelet transform.

Later on we shall go to what is called the continuous wavelet transform were the variables; the independent variables that are associated with the wavelet transform all become continuous. Following that we shall look at some of the generalizations of the ideas that we build up earlier in this course and towards the last phase of the course we shall look in depth at some of the important applications to which wavelets and multirate digital signal processing provide great advantages.

Now, I would like to spend a little while in this lecture on building up in parallel, some of the developments that took place to introduce the subject of multirate digital signal processing. What is multirate? What rate are we talking about here and why do we need to talk about multirate? Why is it connected with wavelets? Let us go back to the audio example or may be let us first go to the biomedical example. In the biomedical example; we said, we would have quicker parts in the response and slower parts in the response. The slower parts of the response are likely to last for a longer region in time, the quicker parts of the response are likely to last for smaller regions in time.

So, here other than the concept of being able localize on a certain region of time and of course, correspondingly on frequency, there is also distinction between what kind of localization is required for higher frequencies and lower frequencies. If we spend a little bit of thought and time in understanding these two kinds of components, we will realize that most of the time when we are talking about slower parts of the response or lower frequencies, we are talking about compromising on what is called time resolution. So, I bring in the idea of resolution here, resolution means the ability to resolve, the ability to be able to identify specific components.

So, for example; frequency resolution relates to be able to identify specific frequency components, and going further and being little a more pin pointed, when I am talking about frequency resolution, what I am saying in effect is. Suppose, I have two sine waves whose frequency start coming closer and closer together, over what region of time do I need to observe them. So, I can actually identify the two frequencies separately, how can I resolve the two frequencies? How much can I narrow down on the frequency axis.

Now, what we are talking about is not so much how much we can narrow down, but how much we need to. When we talk about higher frequency content or things that vary quickly, it is often though not always in that case, that we are willing there to compromise on frequency resolution, but we want time resolution. So, things that take place quickly and are transient short lived demand time resolution and things that occupy the lower frequency ranges, which last for a long time demand frequency resolution. So, very often it is true, that when one goes down on the frequency axis, one demands more frequency resolution, the ability to resolve frequencies more accurately, as opposed to time resolution, the ability to resolve, which time segment it occurs.

And when one goes to higher frequencies for normally demands more time resolution and lesser frequency resolution, one is asking for how closely one can identify two segments or two parts of the waveforms which vary quickly. That means one is trying to narrow down on the time axis and in doing so, then one must compromise on the frequency axis. So, this is what brings us to the idea of multirate processing, you see it means that when I talk about bands of higher frequencies, I must use smaller sampling rates in a discrete time processing system. When I am talking about lower frequency ranges, I must use larger time sampling points or sampling intervals. Why must I do so? To be most efficient in the processing operation. When I am talking about lower frequency so in evoke potential waveform, if I am trying to look at the slower components, I should not unnecessarily sample to frequency, it only increase my data burden and does not offer me anything special.

On the other hand, when I am analyzing the quicker components it is inadequate to use a low sampling rate, I would be doing injustice to the components, for those of us who might be exposed to the concepts of sampling and aliasing. If I am not faithful in my sampling rate of the quicker components I will introduce aliases, I would introduce spurious effects, which I do not want. So, on and all we recognize that it is not a good

idea to be using the same sampling rate for all frequency components. So, unlike a basic course on discrete time signal processing, where we assume all sequences are at the same sampling rate, here we need to deal with sequences that are effectively a different sampling rates in the same system. That means, we also need to deal with systems that operate with different sampling rates and that is why we talk about multirate discrete time signal processing.

Now, at a conceptual level, we understand very well why there is a close relationship between multirate discrete time signal processing, the idea of uncertainty, the requirement of resolution. And if we go further, then when we do multirate discrete time signal processing, we also bring in a new concept of filter banks versus filters. As I said if we go back to the biomedical example, there is the effect or there is the desire to separate components. So, when I wish to separate component, naturally I wish to have many different operators all at once. So, I need a system of filters, which not only have certain individual characteristic, but also have collective characteristics. So, I need to be able to analyze and then synthesize and all this with localization included this is what we mean by a filter bank.

So, a bank of filters as opposed to a single filter in discrete time signal processing refers to a set of filters, which either have a common input or a common point of output or summation output. This concept of a bank of filters and in fact two banks of filters, an analysis filter bank and a synthesis filter bank taken together is very central to multirate discrete time signal processing. We shall be looking at that concept in great depth, so we shall be building up the idea of a two band filter bank in reasonably great depth in this course. The concept of a two band filter bank is a great importance in being able to construct wavelets. In fact, we shall see even from the Haar multi-resolution analysis example, that there is an intimate relationship between the wavelet or the Haar wavelet and the two band Haar filter bank.

So, much so, that if I construct a properly designed two band filter bank, I also construct a multi-resolution analysis that goes with it so to speak. All this is very exciting and what we intend to do in the lectures that follow from here is to take these concepts one by one. So, in the next few lectures, we intend to talk about the Haar multi-resolution analysis to build up certain basic ideas from it, in the lectures that follow as I said before, we intend to talk about the uncertainty principle and its implications. Following that we move from

the discrete to the continuous wavelet transform and then generalize both the continuous and discrete wavelet transforms to a broader class of transforms.

Finally in the course, we would talk about what is called the wavelet packet transform, and some variants of the wavelet transform and variants of filter banks and to end we shall look at several different applications of wavelets taken from different fields. With that then, we come to the end of this first introductory lecture on the subject of wavelets and multirate digital signal processing and proceed there within the next lecture to talk about the Haar multi-resolution analysis. Thank you.