Analysis Of Obstacle Detection

by

Megha Pandey(02D07006)

Jayaprakash(02D07021)

Introduction:

This project is based on the principle of echolocation that is used by bats and a few other species of animals for navigation without sight.The idea is being extended to design a theoretical model for a device that could help blind people "see".We have analysed a simplified system of the form transmitter — obstacle — receiver,with a few assumptions.And thereby we have proposed a method to map a given span of area in a form that could be used by blind people.

Underlying Principle:

Echolocation is a method of sensory perception by which certain animals orient themselves to their surroundings,detect obstacles,communicate with others and find food.The echolocaters include bats, tortoises , some whales,certain species of birds etc.In echolocation a series of short, high-pitched sounds are emitted by an animal. These sounds travel out away from the animal and then bounce off objects and surfaces in the animal's path creating an echo. The echo returns to the animal, giving it a sense about what is in its path.Once an obstacle is detected,the speed and intensity of signals is increased.A bat can determine an object's size, shape, direction, distance, and motion. This echolocation system is so accurate that bats can detect insects the size of gnats and objects as fine as a human hair.

System Specifications and Analysis

Description of the Instrument:

The device consists of a highly directional radiator,radiating short high energy pulses.Immediately after a pulse is emitted,it is made to act as a receiver.If a reflecting object is located in the path of the beam of the radiator,the energy in the pulse will strike it and will be scattered in many directions.Much of it is lost,but some of it definitely reaches the radiator.This is processed to obtain the distance,direction and velocity(if the obstacle is moving)of the obstacle.This information is displayed using an indicator.Here one cycle ends.

After the echoed pulse is received,the device functions as a radiator again;the direction of the beam being shifted vertically downward.And the whole procedure is repeated.This is done for a straight vertical line segment after which we shift to another vertical line,adjacent to the current i.e; we shift by a small angle.A successive repetition will span the required area.Finally the indicator develops a point-by-point image of the entire span covered.

Figure 1: *Block Diagram of the Instrument*

Assumptions:

1. $x(t) = e^{j\omega_0 t}$
   $y(t) = Ae^{j(\omega_0+\omega_d)(t-t_0)}$
   $= Ax(t-t_0)e^{j\omega_d(t-t_0)}$
   where
   A=Attenuation in the amplitude caused due to scattering of energy in various directions;
   ω_0 =frequency of the transmitted wave;
   ω_d =change in frequency of the signal due to doppler shift;
   t_0 =time delay of the received signal;

2. Any change in the reflected signal due to interference with noise is neglected.

3. The transmitter/receiver is assumed to be a point object i.e; it receives only that signal which is travelling along the direction of the transmitted wave.

4. The instrument is stationary.

Obstacle Analysis:

System : Obstacle

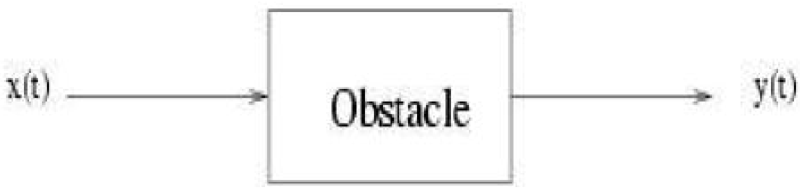


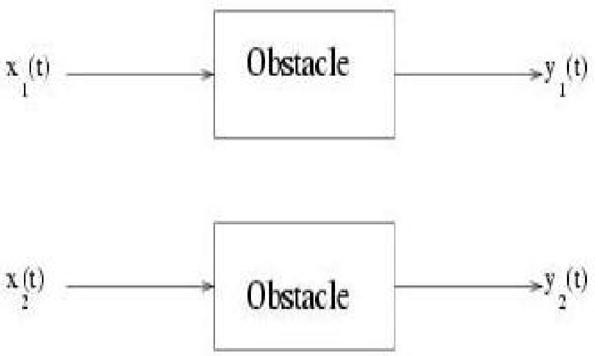
Figure 2: *Obstacle as a system*

Properties of the System

1. Impulse Response:

For $x(t) = \delta(t)$
$y(t) = h(t) = A\delta(t-t_0)e^{j\omega_d(t-t_0)}$
$= A\delta(t-t_0)$

2. Linearity:



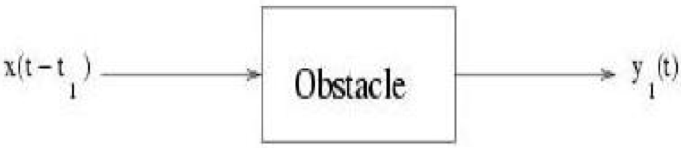
3. Additivity:

$$y(t) = A(x_1(t-t_0) + x_2(t-t_0))e^{j\omega_d(t-t_0)}$$
$$y_1(t) + y_2(t) = Ax_1(t-t_0)e^{j\omega_d(t-t_0)} + Ax_2(t-t_0)e^{j\omega_d(t-t_0)}$$
$$\Rightarrow y_1(t) + y_2(t) = y(t)$$
$\Rightarrow$ The system is additive. —**(1)**

4. Homogeneity:

$$y(t) = A(cx_1(t-t_o))$$
$$cy_{(}t) = c(Ax_1(t-t_0))$$
$$\Rightarrow y(t) = cy_1(t)$$
$\Rightarrow$ The system is homogenous. —**(2)**

5. Shift Invariance:



$$y_1(t) = Ax(t-t_1-t_0)e^{j\omega_d(t-t_0)}$$
$$y(t-t_1) = Ax(t-t_1-t_0)e^{j\omega_d(t-t_1-t_0)}$$
$$\Rightarrow y_1(t) \neq y(t-t_1)$$
$\Rightarrow$ The system is not shift invariant.

6. Causality:

As we have already seen h(t)=0 for t<0
This implies system is causal

7. Stability:

$$\int_{-\infty}^{\infty} h(t)\,dt = \int_{-\infty}^{\infty} A\delta(t-t_0)\,dt$$
$$= A$$
A being finite,the system is stable.

8. Memory:

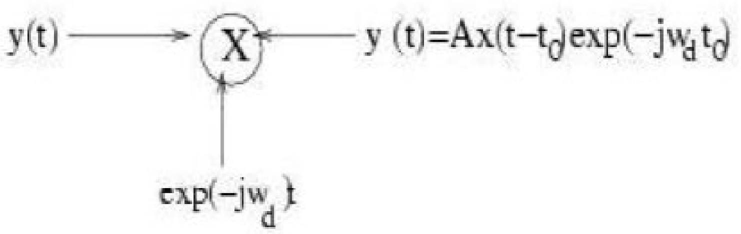
$y(t) = Ax(t - t_0)e^{j\omega_d(t-t_0)}$
$\Rightarrow$ The output at time t depends on input at time $(t - t_0)$.
$\Rightarrow$ The system has memory.

The reflected signal after being received is further processed in the following manner:

In the signal Processor :

1. Frequency$(\omega_0 + \omega_d)$ is detected,thus giving ω_d.
2. A positive ω_d indicates that the obstacle is moving towards the receiver,while a negative ω_d implies that the obstacle is moving away from the instrument.Its speed being given by// $v = c\left(\frac{|\omega_d|}{\omega_0}\right)$
   where, c=speed of the transmitted signal in air.

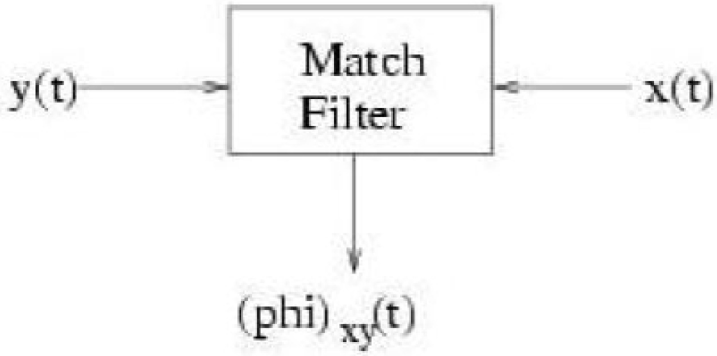
3. Now the signal is passed through a multiplier.



$$y(t)=Ax(t-t_0)\exp(jw_d t-t_0)$$

$y_1(t) = y(t) \times e^{-j\omega_d t}$
$= (Ax(t-t_0)e^{j\omega_d(t-t_0)}) \times e^{-j\omega_d t}$
$= Ax(t-t_0)e^{-j\omega_d t_0}$
$= Bx(t-t_0)$
where $B = Ae^{-j\omega_d t_0}$
This is the new signal obtained($y_1(t)$).

4. $y_1(t)$ is then passed through a match filter which performs cross correlation of $y_1(t)$ with $x(t)$ and maximises the obtained function.



$\phi_{xy_1}(\tau) = \int_{-\infty}^{\infty} x(t+\tau)y_1(t)\,dt$
$\phi_{xy_1}(\tau)$ would be maximum when $\tau = t_0$.
Thus we have obtained time delay t_0.

5. Distance of the obstacle can be calculated as :
$D = c \times \frac{t_0}{2}$

Conclusion

We have obtained a point-by-point mapping of a span of the required area.Each point gives a set of information including distance of the obstacle and its velocity.
This is a general method of detection.The input waveform can be chosen to suit case specific requirements.Once the mapping is obtained,it can be put to various uses.For example,for use by blind people there can be a board with movable pinheads corresponding to each of the points that are being mapped.A pin corresponding to an obstacle will be raised by a distance proportional to the distance of the obstacle from the receiver.If the obstacle is moving,the corresponding pinhead could be made to vibrate with a frequency proportional to the velocity.

Speech and Hearing

by

Piyush Nayar(02D07007)

Gaurav Boob(02D07008)

Abstract:

In this presentation, we have tried to understand how human vocal and hearing organs work. We have studied there input and output signals. We have then drawn the analogy of these organs with circuits and applied the concepts of signal and systems. We have tried to represent ear and vocal chords and vocal tract as a system and have defined approximate input and output signals to this system, thus drawing a parallel between mechanical system and biological system. We also have tried to find out various properties of these mechanical systems.

Introduction

The ability to use words as verbal means of communication is only with humans. Various vocal sounds can be produced using the vocal organs. The main organs which effect the production are the lungs, larynx, vocal folds, the vocal tract, the tongue and the lips. The control of these vocal organs is quite complex. The speech is analyzed using spectrogram. The speech production is explained using the source filter theory. The consonants and the vowels are quite different in their production while for vowels are made by fairly open vocal tract the consonants are made by constriction in the tract. In this presentation we have tried to explain the production of the vowels using the source filter theory. The larynx acts as the source its variation is quite complicated so we have used an artificial vocal cord model to represent the variation. We have tried to model the speech production in humans as a combination of two systems one the source other the filter. The inputs to the system are many but we have tried a simple model using the basic inputs for the cord and the tract. Using this we have studied the various properties of this system.

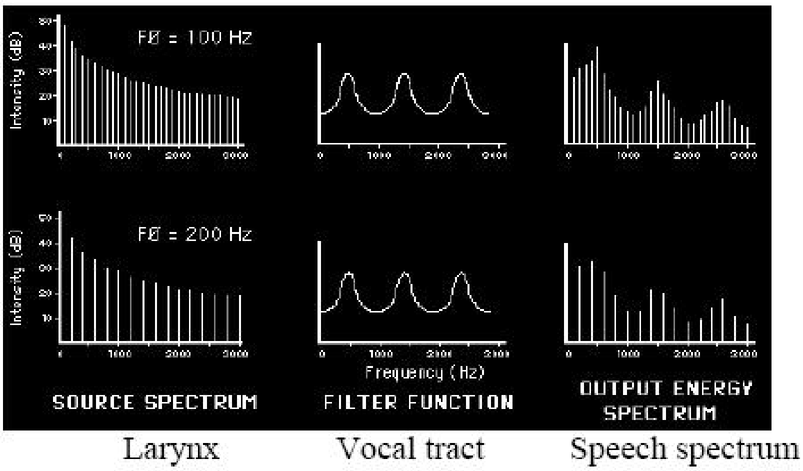
Ears are important sensory organ in a human body. The ears contain structures for both the sense of hearing and the sense of balance. They indicate direction of sound source better than eye and they are omni-directional. The human ear's purpose in the area of hearing is to convert sound waves into nerve impulses. These impulses are then perceived and interpreted by the brain as sound. The human ear analyzes sound into its frequency components not by its waveform. This is because almost all sounds are structured in frequency and it is easier to detect the sound in presence of other sounds. The human ear is made of three distinct areas: the outer ear, the middle ear, and the inner ear. The outer ear channels sound waves through the ear canal to the eardrum. The eardrum moves in response to the changes in pressure i.e. Waveform.The inner ear houses the "cochlea", a spiral-shaped structure that contains the organ of "Corti" - the most important component of hearing. The Corti sits in an extremely sensitive membrane called the "basilar membrane". Whenever the basilar membrane vibrates, small sensory hair cells inside the Corti are bent, which stimulates the sending of nerve impulses to the brain.

In this presentation, we have tried to explain the working of ear through its mathematical model. Ear is taken as a system which acts as converting input sound waves to nerve impulses.

System Representation

Speech Production

We have tried to represent the speech production using the source filter theory. According to it larynx acts as source and the vocal tract as a filter and the combination of the two gives the speech.



Thus we can represent the speech production as a combination of three systems.

Source system:-
The input is air given by the lungs.
The output is air pressure pulses.

Filter system(an LSI system):-
The input is air pressure generated from the source.
The output is speech produced as air pressure fluctuations.

System describing the effects of changes in vocal tract
The input is the constriction and the length of the tract.
The output is the frequency response of the filter.
Once the frequency response of the filter is decided by the constriction system the filter behaves as a LSI system. We could consider the filter system as a convolution of the two inputs the frequency response and the input obtained from source. In this presentation it is assumed, when talking about the filter system, that its frequency response is fixed and not considering it as an input.

Hearing System

We tried to represent Ear as a Bank of Band Pass Filters. According to it, the cochlea is represented as bank of Band Pass Filters.

The ear can be represented as a cascade of two systems:
a) The one formed by the outer and middle ear:-
The input is the sound wave
The output is also sound wave.
This system just amplifies the input.

b) The inner ear:-
The input is the output of the above system.
The output is neural signals.
This system acts as a bank of band pass filters.
These are a non linear filter which distills the various frequencies and send them to the brain.

Properties of the System

1.Linearity

We have tried to represent the speech production using the source filter theory. According to it larynx acts as source and the vocal tract as a filter and the combination of the two gives the speech.

2.Memory:-

The source system has memory as the pulse creation of the larynx is memory dependent the present air output depends on the crosssection which is decided by the previous output as a result of Bernoulli's theorem.

The filter is not memory less as the frequency response is not a constant which should be for a memory less system (impulse response of a memory less system is an impulse).
The system describing the effects of the length and constriction is memory less as the frequency response of the system will depend on the present input only.

3.Shift invariant:-

The source system is shift invariant as if the air from the lungs came after a time interval then also the pulses produced will remain the same.
The filter system is also shift invariant. The constriction system is also shift invariant.

4.Causality:-

The source system is casual. Since if the input is zero up to a certain time the output will also remain the same as without the air from the lungs the source cannot give an output air pulses. The filter system is also casual, since we can't expect to have speech without giving any input.
The constriction is casual as it is memory less.

5.Stability:-

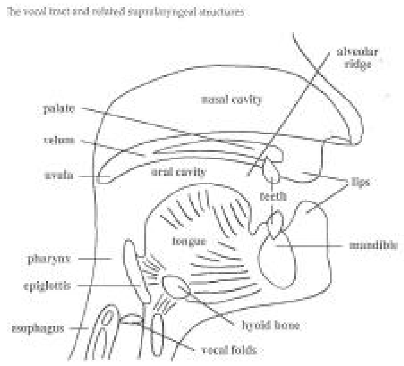
The source system is stable as the filter system is bounded and so x(t) is bounded. The filter system is bounded as it limits the frequencies and thus decreases the energy in the spectrum. Also the amplification of the filter is not unbounded.
The constriction system ideally is not bounded as for zero length the lowest formant frequency also becomes very high ,but in reality we never expect the length to decrease to zero and similarly for the other inputs. So within actual variable values of the system the system is stable.

6.Invertibility:-

The source system is invertible.
The filter system is not invertible as many inputs may have same output. The constriction system is not invertible as different set of inputs may lead to same output.

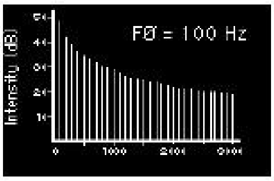
Working Of the Speech Production

Below is the diagram of the human vocal organs:



The larynx provide short pulses with a fundamental frequency the spectrum of the output are impulses with frequency multiples of the fundamental frequency and decreasing amplitude. The larynx takes air as the input from the lungs and gives pulses of air as output. This is done with the help of the vocal folds the sub glottal pressure forces the vocal folds to open and then move outward; when the glottis is open and the pressure between the folds drops, they reverse and then move inward the momentum of inward movement and, finally, the increased suction of the Bernoulli effect causes the folds to close abruptly. The sub glottal pressure and elastic restoring forces during closure causes the cycle to begin again

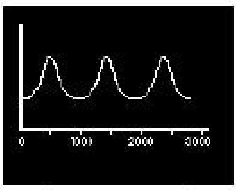
This is called the source. The variation of the frequency cannot be explained well with the changes in the larynx so we represent the larynx and vocal fold by an artificial vocal. The vibration of the fold in larynx makes the vocal sound. The tension in the fold may vary causing a change in the fundamental frequency The spectrum of the output of the larynx is as shown below. This is a typical output characteristic.



The amplitude decreases for higher frequencies.

The vocal tract is the most important part in the production of the sound. The output of the source is nearly the same for a person speaking at a constant pitch. The vocal tract acts as a resonating tube of varying cross-section. The resonances in the speech production are called the formants.

The frequency of resonant oscillation of the air in the vocal-tract is similar to in that in a tube like bottle. When we uncork a bottle a pulse starts it oscillates between the two ends of the bottle with a frequency inversely proportional to the length. In, vowel production, pulses of airflow are emitted by the glottis into the pharynx each glottal pulse is propagated upward to the open mouth as a pressure wave; at the mouth the pressure pulse produce an outward pulse of air flow, the escaping air particles of which now appear as a rarefaction (negative) pressure pulse for propagation back towards the vocal fold surfaces. The vocal fold acts like the bottom of a bottle and reflect the rarefactions pulse; it then is propagated upward again and so on. Thus the glottal pulse is repeatedly reflected back and froth between the vocal folds and the mouth. This round tip propagation is so fast that typically 10 such round between a glottal pulse. The resonances of the vocal tract are called the formants.
The frequency response of the tract (filter) is typically as below.

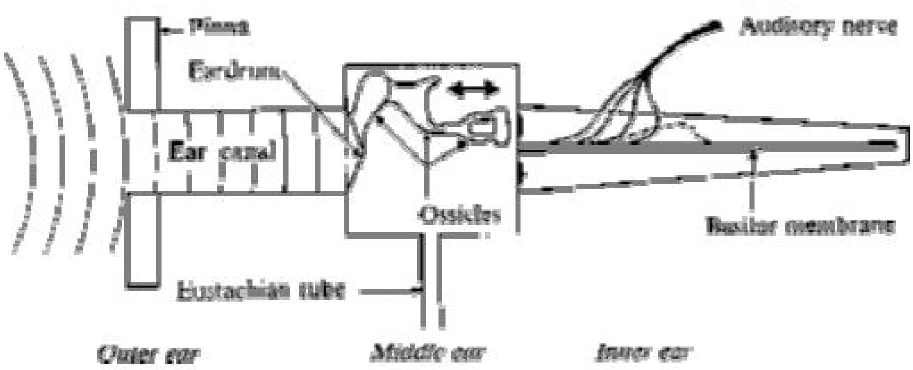


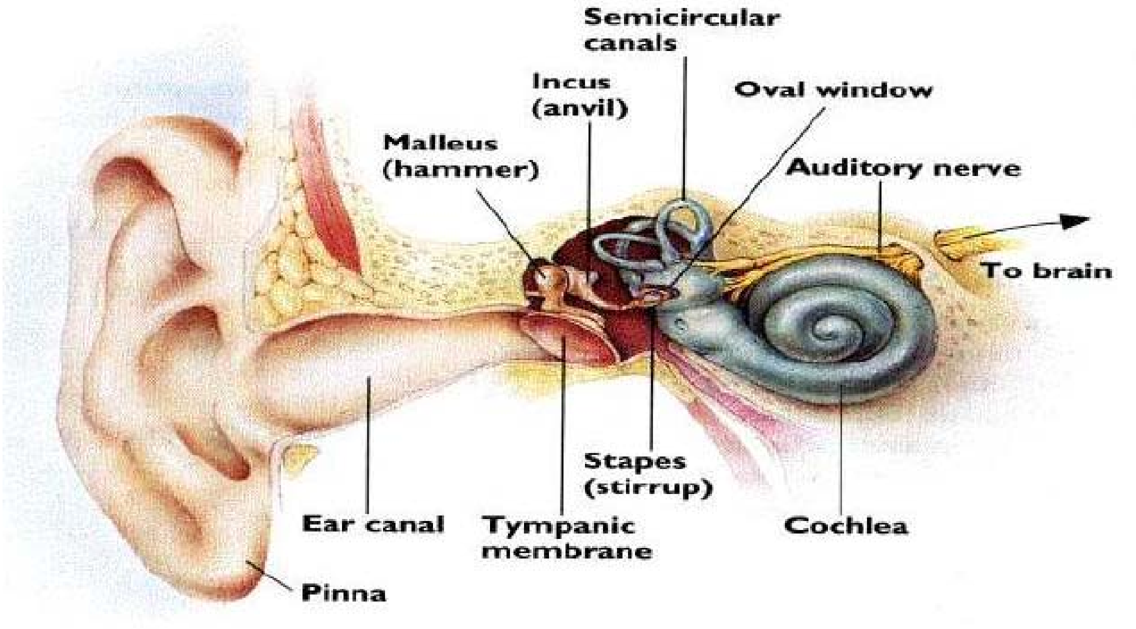
The typical response for no constriction tract

The no constriction tract response is also called the neutral position the vowel produced is [ ]
The formant for the above is 500Hz typical for a male sound. The formants are the acoustic properties of the vocal tract that produced the spectrum. The formant frequency locations for vowels are affected by three factors:-
· Length of the oral tract
· Location of the constriction in the tract
· Narrowness of the constrictions

Working of the Ear:

An ear is divided into three parts: outer, middle and inner ear.





The outer ear consists of "pinna". It has a quite significant effect on incoming sound, particularly at high frequencies, and contribute to out ability to localize sound- that is, where they are coming from. The incoming sound waves travel down the "auditory meatus "and cause the eardrum to vibrate. Eardrum is the interface between the outer and middle ear. The vibrations of the eardrum are transmitted through the middle ear by three small bones, called the "auditory ossicles", to a membrane covered opening in bony wall of inner ear.

The inner ear consists of a fluid-filled spiral-shaped structure called "cochlea". The middle ear ensures that the vibrations of eardrum are transferred efficiently to fluids inside cochlea. The middle ear also increases the sound pressure by large factor. This extra intensity ensures that much larger fraction of the sound energy is transmitted to the inner ear.

Cochlea is spiral shaped tube with bony walls filled with incompressible fluid. It is the most important part of ear. This is the part responsible for the signal processing. Main component of cochlea is "Basilar Membrane", which is not rigid. The displacement in basilar membrane caused by the inward motion of the stapes is transmitted along the basilar membrane just like the impulse is transmitted along a string which has given a sharp flick. Basilar Membrane does not have a homogeneous structure. Instead, it is rather narrow and stiff at basal end and gets wider and more flexible near apical end. So, when cochlea responds to a sinusoidal stimulus, the wavelength and amplitude of the wave in basilar membrane change as it travels along it. The wavelength gets shorter and the amplitude gradually increases until it reaches a certain point on the basilar membrane, after which it rapidly decreases. Whatever, the stimulus frequency, the disturbance on the basilar membrane does not travel much further after the amplitude has peaked. Since the point on the basilar membrane at which maximum displacement occurs varies with frequency, the basilar membrane effectively separates out the sinusoidal components in the stimulus. In other words, it performs a crude form of Fourier analysis.

Now we need to know what is the frequency resolution of basilar membrane? That is how close must two simple tones be in frequency before the basilar membrane is unable to distinguish them. we can measure this by assuming that it behaves as a band pass filter, with particular frequency, and band width and a sloping transition band at upper and lower cutoff frequencies.

The basilar membrane is Non-Linear. This means that one can't predict how it will respond to quite sounds by examining how it responds to loud ones. After processing in basilar membrane the spectral information is transferred to auditory nerve. This happens through organ of "Corti". It is combination of "tectorial membrane" and "hair cells". The hairs on outer hair cell actually make contact with the tectorial membrane but the inner hair cells probably do not. The tectorial membrane is effectively hinged at its edge so that when basilar membrane moves up and down, the tectorial membrane slides over it with a shearing motion, causing the hairs on the hair cells to be displaced. This causes the inner hair cells to fire and send signals up the auditory nerve to the brain.

Theory

Effects of constriction and length of vocal tract.

The effects of the constrictions and the length of the tract can be explained on the basis of the following principles:

· Length of the oral tract –
This depends on the physical size of the speaker. The length affects the frequency location of the formants. This helps us to predict the formants for child, men, and women.

The rule for length and frequency is: The average frequencies of the vowel formants are inversely proportional to the length of the oral tract. Thus we see that for children who have length half that of male the frequencies will be double.
· Location of the constriction in the tract
From the above rule we can get the position of the neutral formants but most vowels are produced by constriction first we consider the change in frequency due to the position of the constriction.
Oral constriction/F1 rule: The frequency of F1 is lowered by any constriction in the half of the oral part of the vocal tract, and greater the constriction the more F1 is lowered.
Pharyngeal constriction/F1 rule: The frequency of F1 is raised by constriction of the pharynx and the greater the constriction, the more is raised.
Back tongue constriction/F2 rule: The frequency of F2 tends to be lowered by a back tongue constriction and the greater the constriction the more F2 is lowered.
Front Tongue constriction/F2 rule: The frequency of F2 is raised by front tongue constriction and the greater the constriction, the more F2 is raised. Lip Rounding Rule: The frequencies of all formants are lowered by lip rounding and more the rounding more the formants are lowered.

Thus we see that these factors affect the output of the filter. It changes the impulse response of the filter. Thus due to this change of impulse response the output changes and we get different vowels.

Filters:-

The filter is a device which filters out the frequency components. It allows only certain frequencies to pass with amplification 1 and stops the other.
There are four types of filters:-
a) Low-pass: allows low frequencies to pass through.
b) High-pass: allows frequencies above a certain value to pass.
c) Band-pass: allows frequencies in a band to pass.
d) Band-stop: allows frequencies outside a band to pass.

An ideal filter is infinitely non-casual, while real filters are casual, so real filters allow some frequencies to pass but not necessarily with unity amplification.

Glossary and References

Glossary:

a) tract- the vocal tract in the human body.
b) formants- the resonant frequencies of the tract.
c) fundamental frequency- the frequency of the pulses by the source.

References:

THE ACOUSTICS OF SPEECH COMMUNICATION by J.M.PICKETT
An Introduction to psychology of hearing Moore, B.C.J.
www.phon.ucl.ac.uk/home/andyf/ISS/ishlect6.ppt
www.haskins.yale.edu/haskins/HEADS/MMSP/acoustic.html
citeseer.ist.psu.edu/context/85156/0
www.wfubmc.edu/neurology/rosenberg/audio1.html

Applications of Fourier Transforms

by

Mhaske Viabhav(02D07018)

Avdhoot Bane(02D07016)

Gaurav bharjatiya 02007029)

Introduction:

Who was Fourier?

Born in France in 1768, Jean Baptiste Joseph, Baron de Fourier came up with this piece of maths in his thirties after puzzling over problems with agriculture. Of great interest to him was the fact that the surface of the ground got quite hot during the day and cold during the night, but deeper down the temperatures were more nearly constant. He was seeking to understand heat flow.
Quarter of a millennium on, students of maths and engineering remain convinced he was seeking to impose misery and gloom on future generations. You will see that fourier analysis is used in almost every aspect of the subject ranging from solving linear differential equations to developing computer models, to the processing and analysis of data.

What is Fourier Transform?

Physical significance of Fourier Transform :
The Fourier Transform converts a set of time domain data vectors into a set of frequency (or per time) domain vectors.
Imagine you wanted to know about changes in soil temperature . Now suppose you to measure the temperature of the soil in your garden at dawn, midday, dusk and midnight, every day for a year. You would then have a list of real numbers representing the soil temperatures.

Now if we plot these readings on a graph the vertical y axis would be labelled 'temperature' and the horizontal x axis would be labelled ' time', we get a so called 'time domain' graph.
From the graph we infer the nights are cold and the days are warm and as it seems obvious, summer is warmer than winter!

The graph you've just imagined is (roughly) the sum of two sinusoids (sine waves). One with a frequency of one day as the temperature varies between day and night, the other with a frequency of one year as the temperature varies with the seasons.

The Fourier Transform provides a means of manipulating - or transforming - this raw data into an alternative set of data, the magnitude of which which can be plotted on a graph with differently labelled axis. Never mind the y axis label for now, but the x axis would now be labelled 'frequency' or 'per time' - this is in the 'frequency domain' graph.

This graph looks very different! It will consist of two vertical lines rising from the frequency axis, one at a frequency (or period) of one day, the other at a frequency (or period) of one year. Thus we have 'analysed' a seemingly complicated (rather than complex in the mathematical sense) set of data and extracted the most interesting facts from it - days are warmer than nights, and summer is warmer than winter! There is still more information we can extract from the raw results of the Transform not shown on such a graph, we can look at the 'angle' information of the raw results which might tell us that the coldest day isn't January 1st, and midnight isn't the coldest part of the night.

We could use exactly the same Fourier Transform for more interesting purposes - we could for example sample some music, transform it, and plot the 'frequency spectrum' to reproduce those dancing bars of LEDs on the stereo system!

Mathematical explanation of Fourier Transform : The Fourier Transform is a generalization of the Fourier Series. Strictly speaking it only applies to continous and aperiodic functions, but the use of the impulse function allows the use of discrete signals. The fourier transform is defined as

$$F(j\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t}\, dt$$

The inverse transform is defined as

$$f(t) = 1/2\pi \int_{-\infty}^{+\infty} F(j\omega) e^{j\omega t}\, df$$

where

f(t) is the signal and

F(j ω) is the fourier transform of the signal.

Applications of Fourier Transform:

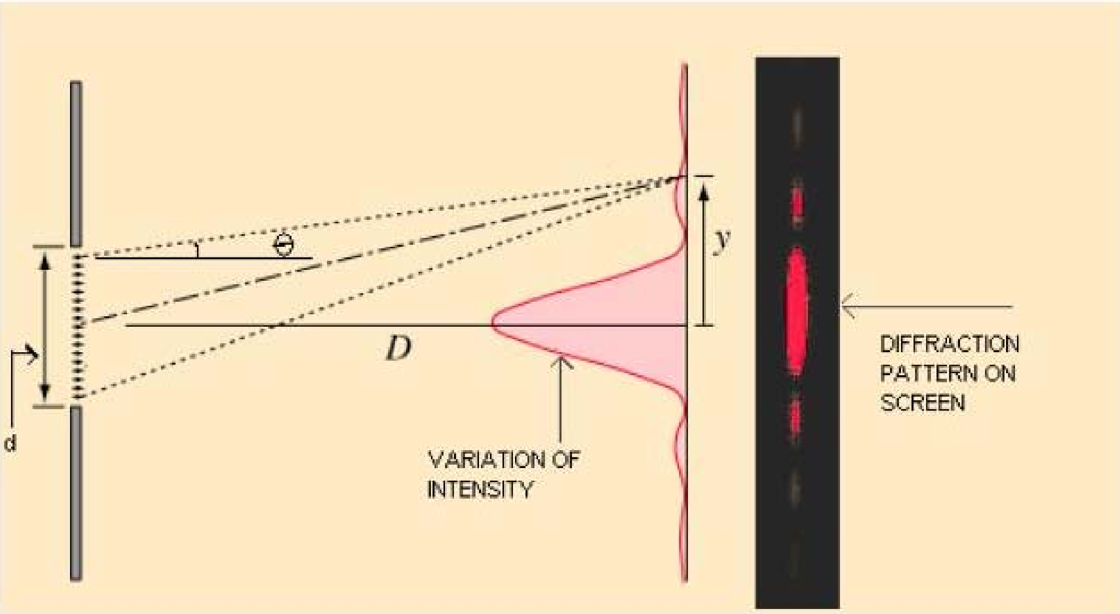
· Designing and using antennas
· Image Processing and filters
· Transformation, representation, and encoding
· Smoothing and sharpening
· Restoration, blur removal, and Wiener filter
· Data Processing and Analysis
· Seismic arrays and streamers
· Multibeam echo sounder and side scan sonar
· Interferometers — VLBI — GPS
· Synthetic Aperture Radar (SAR) and Interferometric SAR (InSAR)

· High-pass, low-pass, and band-pass filters
· Cross correlation — transfer functions — Coherence
· Signal and noise estimation — encoding time series.

Use of Fourier Transform in Optics

1.Single Slit Diffraction Pattern:

The spreading out of a wave when it passes through a narrow opening is known as the phenomenon of diffraction and the intensity distribution on the screen is called the diffraction pattern. In fresnel diffraction,the screen or the source (or both) are at finite distances from the aperture i.e slit ,then the diffraction pattern corresponds to the fresnel class.In fraunhofer diffraction both the source and the screen are at infinity.



SINGLE SLIT FRAUNHOFER DIFFRACTION PATTERN.

The minima on the screen is given by the equation : $d \sin\theta = m\lambda$
where

d : width of slit .
θ: angle shown in the above figure .
m: integer.
λ :wavelength of the signal.

The intensity distribution on the screen is given by :

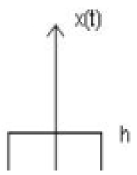
$$I = I_0 \sin^2 \beta / \beta^2$$

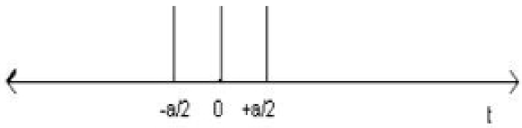
where

$\beta = \pi d \sin\theta / \lambda$
I_0=Intensity at θ=0 which is maximum.
λ is the wavelength of the light used.

A slit can be represented by a function where the value of the function represents the illumination. Let the slit width be a. Then the illumination will be constant on each point on the slit. Let constant illumination be h.

Then x(t) = h - a / 2 <= t <= +a / 2





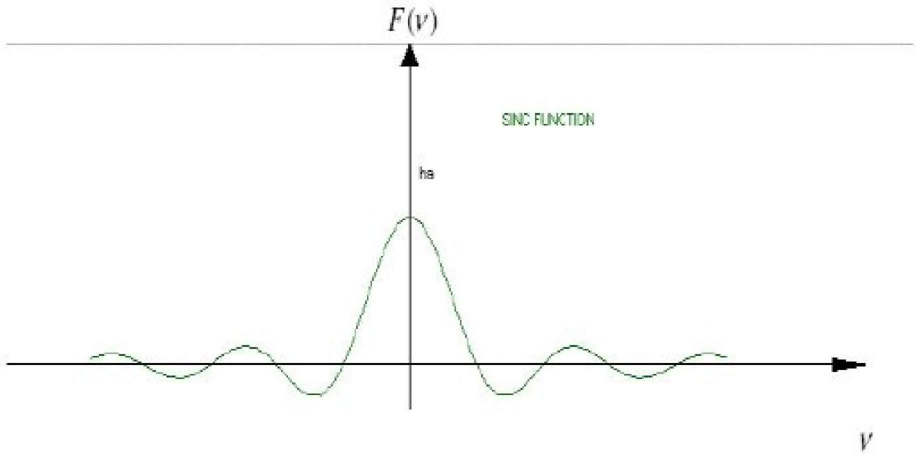
The Fourier transform is given by

$$F(\nu) = \int_{-a/2}^{+a/2} h e^{-2\pi j \nu t} dt$$

$$= h/2\pi j\nu \{ e^{\frac{-2\pi j \nu a}{2}} - e^{\frac{2\pi j \nu a}{2}} \}$$

$$= ha \frac{\sin\beta}{\beta}$$

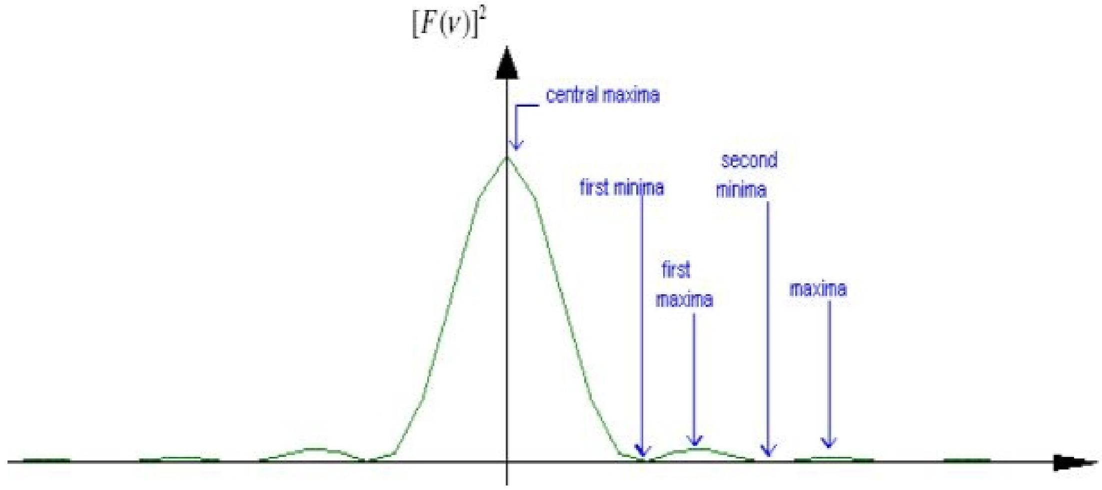
$$where \;\; \beta = \pi a \nu .$$



The intensity pattern is given by the square of the modulus of the fourier transform function .

$$[F(\nu)]^2 = h^2 a^2 [\frac{\sin\beta}{\beta}]^2 \quad [F(\nu)]^2 = h^2 a^2 [\frac{\sin\beta}{\beta}]^2$$

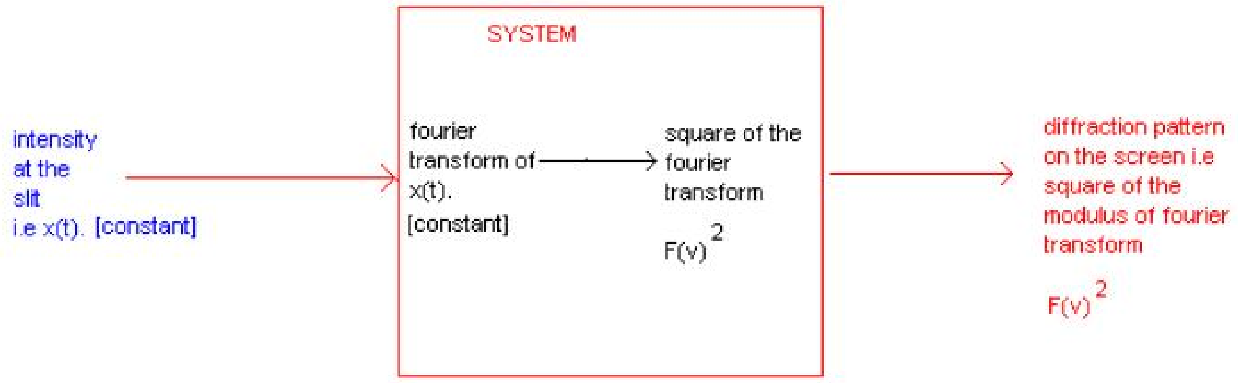
If we plot $[F(\nu)]^2$ *vs* ν *then we get as follows*



The properties of the above function(i.e the pattern of intensity maxima and minima) agree very well with that of the diffraction pattern that one observes due to the single slit.

SINGLE SLIT DIFFRACTION SYSTEM

$$x(t) \longrightarrow [F(v)]^2 = h^2 a^2 [\frac{\sin\beta}{\beta}]^2$$



Properties of the System

1.Linearity:

The system is not linear because if we increase the intensity at the slit then the intensity at the diffraction pattern does not increase linearly since intensity at the diffraction pattern is the square of the modulus of the fourier transform.

We prove this property by giving counter example:

$$x(t) = h$$

$$x(t) \xrightarrow{gives} y[v] = [F(v)]^2 = h^2 a^2 [\frac{\sin\beta}{\beta}]^2$$

$$F(v) = \int_{-a/2}^{+a/2} h e^{-2\pi j v t} dt = ha \frac{\sin\beta}{\beta}$$

$$now$$

$$bx(t) = bh \;\; then$$

$$If \;\; b = 0 \quad then \; output \; is \; zero.$$

$$If \;\; b \neq 0 \;\; then$$

$$[F(v)]^2 = b^2 h^2 a^2 [\frac{\sin\beta}{\beta}]^2 = b^2 y(v) \neq by(v).$$

Now since the system is non-homogeneous , hence it is not linear.

2. Causality:

Clearly if the intensity is zero at the slit then there is no diffraction pattern on the screen. The system is causal as the there is diffraction pattern on the screen only as long as slit is illuminated i.e the diffraction pattern on the screen depends only on the present value of the intensity on the slit.

The intensity throughout the slit is always constant.

$$At\ x(t)=b$$

$$y[v]=[F(v)]^2=[b]^2a^2[\frac{\sin\beta}{\beta}]^2$$

Therefore the output of the system depends only on the present input. Hence the system is causal

3. Memory:

The system is memory less as the diffraction pattern on the screen depends only on the present value of the intensity at the slit.

$$At\ x(t)=b$$

$$y[v]=[F(v)]^2=[b]^2a^2[\frac{\sin\beta}{\beta}]^2$$

Therefore the output of the system depends only on the present input .

Hence the system is memoryless.

4. Stability:

The system is stable since if the intensity at the slit is bounded then surely the intensity at the diffraction pattern on the screen is bounded i.e the intensity at the screen never becomes unbounded if the intensity at the slit is bounded.

For bounded input

$$x(t)=b\ \forall\ t$$

$$y[v]=[F(v)]^2=[b]^2a^2[\frac{\sin\beta}{\beta}]^2\leq b^2a^2$$

we get the is output is bounded.

Hence the system is stable.

5.Invertibility:

The system is invertible as for two distinct intensities on the screen there are two distinct diffraction patterns on the screen.

$$for\ x_1(t)=h_1\rightarrow y_1(v)=h_1^2a^2[\frac{\sin\beta}{\beta}]^2$$

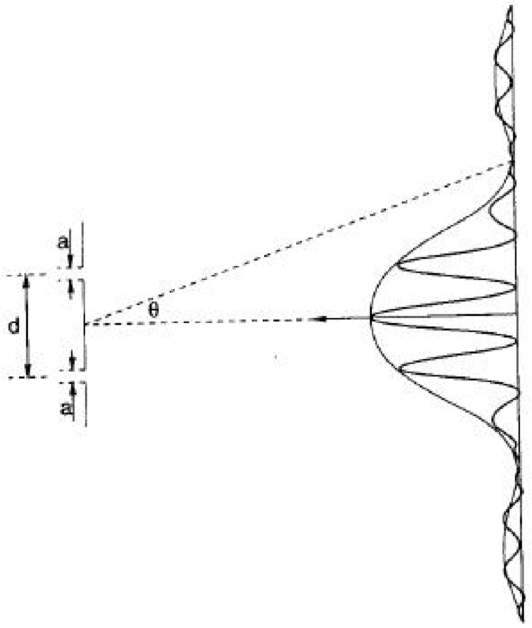
$$for\ x_2(t)=h_2\rightarrow y_1(v)=h_2^2a^2[\frac{\sin\beta}{\beta}]^2$$

Thus for two distinct inputs ,we get two distinct outputs.

Hence the system is invertible.

Double slit diffraction pattrern

The Fraunhofer diffraction pattern produced by two parallel slits (each of width a) separated by a distance d. We would find that the resultant intensity distribution is the product of single slit diffraction pattern and the interference pattern produced by two point sources separated by a distance d. theta is the angle made by the diffracted rays with the normal to the plane of the slits.



The intensity distribution on the screen is given by :

$$I = 4I_0 \frac{\sin^2 \beta}{\beta} \cos^2 \gamma$$

where

$$\beta = \pi a \sin\theta / \lambda$$

I_0=Intensity at θ=0 which is maximum.

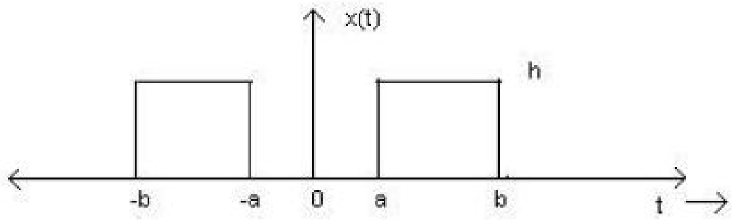
$$\gamma = \frac{\pi}{\lambda} d \sin\theta$$

λ is the wavelength of the light used.

Let both the slits have constant illumination **h** .

Then

$$x(t) = h \qquad -b \le t \le -a \;\; and \;\; a \le t \le b$$



The fourier transform is given by

$$F(\nu) = \int x(t) e^{-j2\pi\nu t} dt$$

$$= \int_{-b}^{-a} h e^{-j2\pi\nu t} dt + \int_{a}^{b} h e^{-j2\pi\nu t} dt$$

$$= \frac{h[e^{j2\pi\nu a} - e^{j2\pi\nu b}]}{-j2\pi\nu} + \frac{h[e^{-j2\pi\nu b} - e^{-j2\pi\nu a}]}{-j2\pi\nu}$$

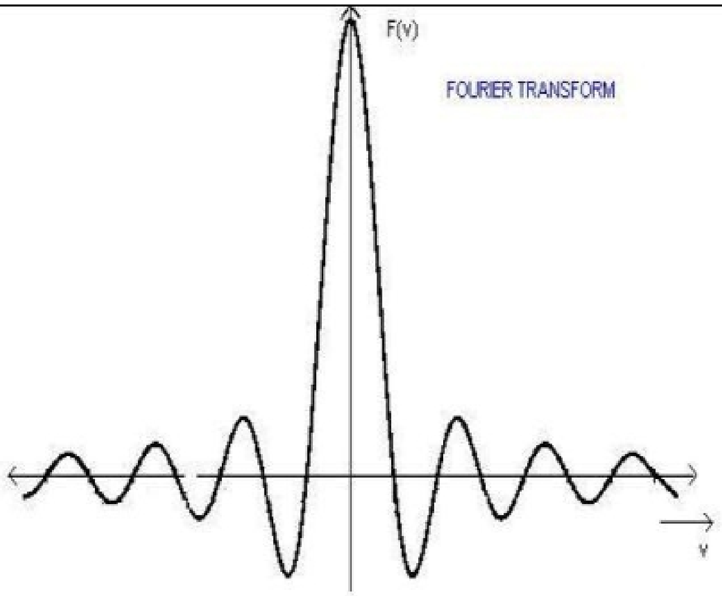
$$= 2h\{\sin\frac{b2\pi\nu}{2\pi\nu} - \sin\frac{a2\pi\nu}{2\pi\nu}\}$$

$$= (b-a)h \frac{\sin\pi\nu(b-a)}{\pi\nu(b-a)} \cos\pi\nu(b+a)$$

Now (b-a) is the slit width .
Let (b+a) is constant = c.

Therefore

$$F(\nu) \quad = ah \frac{\sin\pi\nu a}{\pi\nu a} \cos\pi\nu c$$

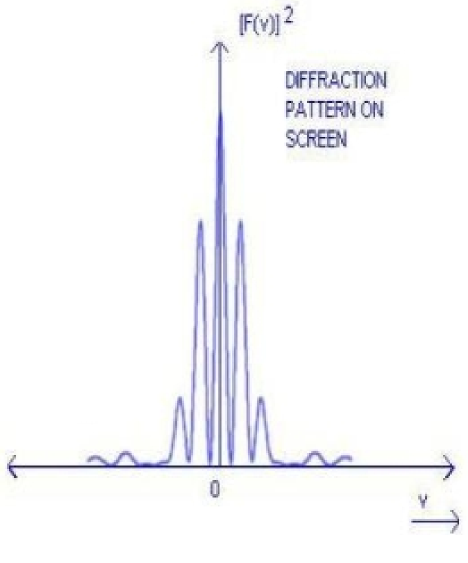


The intensity pattern on the screen is given by square of the modulus of the fourier transform.

Therefore

$$I(v) = [F(v)]^2 = a^2h^2[\frac{\sin \pi v a}{\pi v a}]^2[\cos \pi v c]^2$$

The graph of the intensity pattern is given below:



Similarly we can find the diffraction pattern on screen for **N-slits** by taking the fourier transform at all slits and then squaring the modulus of the fourier transform to get the intensity pattern on screen.

Properties of double slit system.

The properties of system of the double slit are similar to that of single slit system.

FAST FOURIER TRANSFORM ANALYZER

What is FFT?

One of the reasons for the tremendous growth in the use of discrete –time methods for the analysis and synthesis of signals and systems was the developments of exceedingly efficient tools for performance Fourier Analysis of discrete time sequences. At the heart of these methods is a technique that sis very closely allied with discrete-time Fourier Analysis and that is ideally suited for the use on a digital computer or for implementation in digital hardware. This technique is the ***Discrete Fourier Transform (DFT)*** for finite-duration signals.

Let $x[n]$ be a signal of finite duration; that is there is an integer N_1 so that

$$x[n] = 0\,, \qquad \text{outside the interval } 0 \le n \le N_1 - 1$$

Furthermore ,let $X(e^{jw})$ denote the Fourier transform of $x[n]$. We can construct a periodic signal $\tilde{x}[n]$ that is equal to $x[n]$ over one period. Specifically ,let $N \ge N_1$ be a given integer, and let $\tilde{x}[n]$ be periodic with period N and such that

$$\tilde{x}[n] = x[n], \qquad 0 \le n \le N_1 - 1$$

The Fourier series coefficients for $\tilde{x}[n]$ are given by

$$a_k = \frac{1}{N}\sum_{(N)} \tilde{x}[n] e^{-jk(2\pi/N)n}$$

Choosing the interval of summation to be that over which $\tilde{x}[n] = x[n]$, we obtain

$$a_k = \frac{1}{N}\sum_{n=0}^{N-1} x[n] e^{-jk(2\pi/N)n} \qquad (1)$$

The set of coefficients defined by eq. (1) comprise the DFT of $x[n]$. Specifically, the DFT of $x[n]$ is usually denoted by $\tilde{X}[k]$, and is defined as

$$\tilde{X}[k] = a_k = \frac{1}{N}\sum_{n=0}^{N-1} x[n] e^{-jk(2\pi/N)n} \qquad (2)$$

The importance of the DFT is visual from several facts. First note that the original finite duration signal can be recovered from its DFT. Specifically, we have

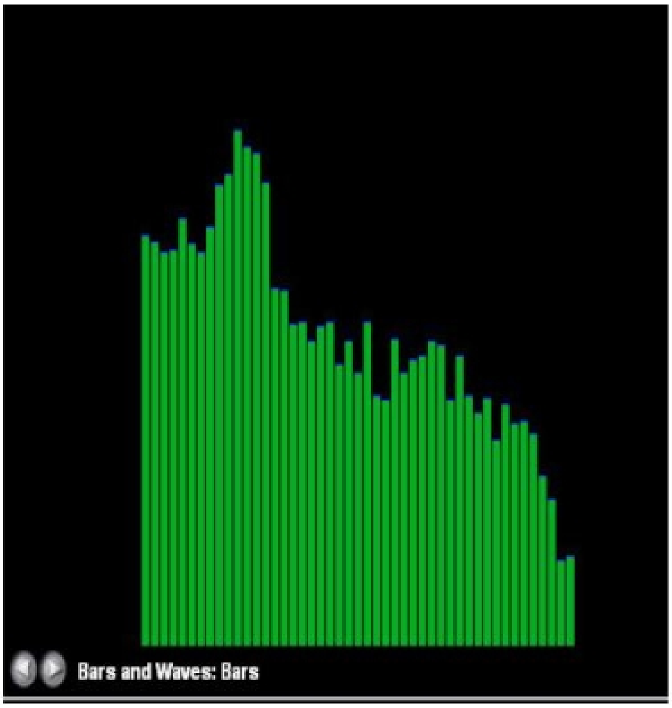
$$x[n] = \sum_{k=0}^{N-1} \tilde{X}[k] e^{jk(2\pi/N)n}\text{ , n = 0,1,2,.....,N-1} \qquad (3)$$

Thus, the finite-duration signal can either be thought of as being specified by the finite set of nonzero values it assumes or by finite set of values of $\tilde{X}[k]$ in its DFT. A second important feature of the DFT is that there is **an extremely fast algorithm,** called *fast Fourier transform(FFT)*, for its calculation.

Use of FFT in Windows Media Player:

All of us must have listened to music on computer in windows media player .On it ,we must have seen dancing bars as shown below. It is very interesting to know that these dancing bars are obtained by applying fast fourier transform on continuous audio signal.

Use of FFT in Windows Media Player



Steps involved in getting the visual bars in window media player :

1) DIGITIZE THE INPUT SIGNAL AT HIGH SAMPLING RATE.
2) FOR $x(0)$, FRQUENCY SPECTRUM (F.S.) IS PLOTTED.
3) IF TIME CONSTANT OF SAMPLING IS t_1 , STEP 2 IS REPEATED FOR $x(t_1), x(2t_1)$ & SO ON.

This step is repeated for all digitized input samples which leads to dancing bars on window.

Nyquist's Sampling Theorem:

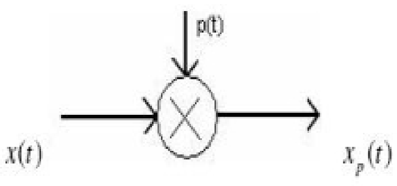
As long as the sampling rate is greater than twice the highest frequency component of the signal, then the sampled data will accurately represent the input signal.

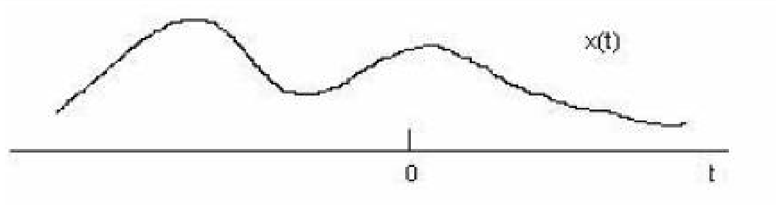
Audio range of frequency:

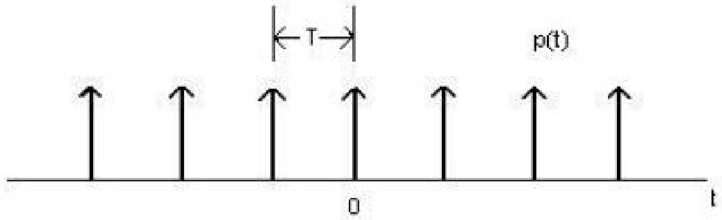
Human ear can hear frequencies ranging from 20Hz-20KHz.Hence,we have to sample the frequency ranging from say,0-20 KHz.

Sampling Rate & Time Constant:

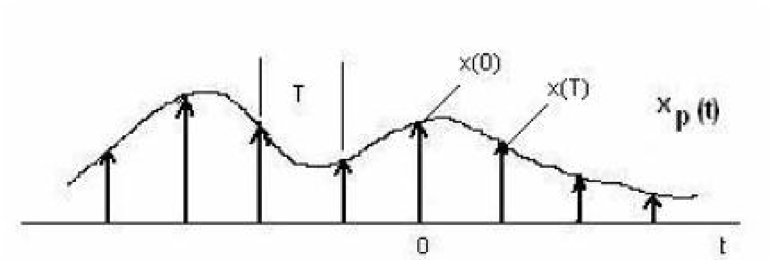
If sampling period is T , then







$x_p(t)$



$$x_p(t) = x(t)\,p(t)$$

$$p(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT)$$

$$x_p(t) = \sum_{n=-\infty}^{+\infty} x(nT)\,\delta(t - nT)$$

FREQUENCY RESOLUTION:

DEFINITION:
Minimum Frequency interval discernable.
i.e. Lesser the value of resolution , better it is. Frequency Resolution (F. R.) = 1/T
Reciprocating Bandwidth =1/ (time constant)

PHYSICAL SIGNIFICANCE OF F.R. & T :

Larger time window:

1. Better Resolution
2. Slower processing(Longer Time Window & more data to crunch) FR=WWWW1/TC

Smaller Time Window:

1. Lower Resolution
2. Faster processing
Frequency resolution is directly dependent on reciprocal bandwidth...Lesser the bandwidth Better is the resolution of the spectrum.

WINDOWS MEDIA PLAYER:

Say, if you want to have 50 dancing bars; then to get 50 bars (reciprocal Bandwidths) in the range 0-20 KHz., we should get reciprocal bandwidth of 400 Hz.
Time constant = sampling time period = (1/reciprocal bandwidth)

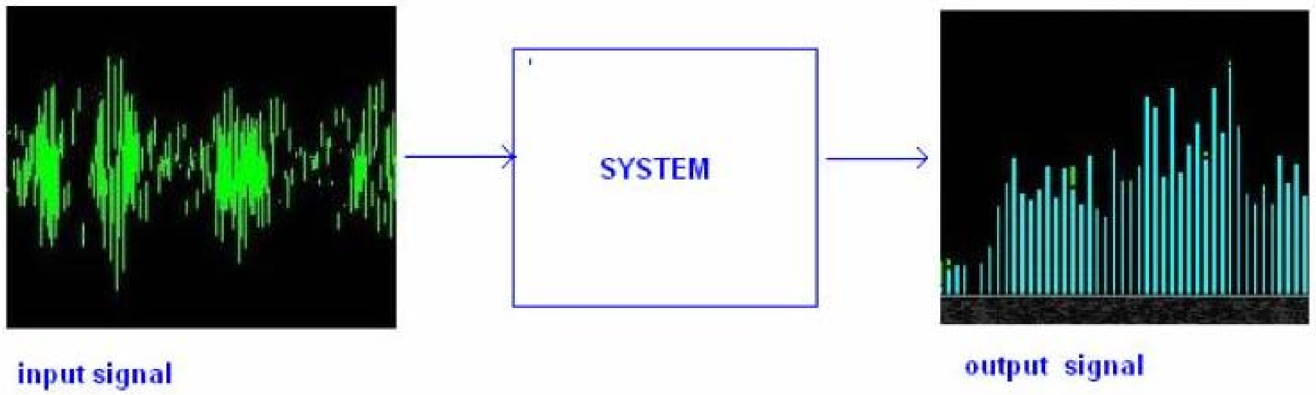
For 50 bars, we need 2.5 ms sampling time period.
It is not that we are not able to sample finer; but the finer sampling (i.e. lesser T Would lead to lesser resolution.)
If we take T large, bandwidth will be smaller .It will lead to more number of Bars; but this might be undesirable as large number of bars might look unpleasant.
It is our choice what time period to choose. According to our convenience , we should choose the sampling time .

In above case, as the bandwidth is 400 Hz.,the frequencies corresponding to multiples of 400 Hz would get the place in frequency spectrum ; but what about the frequencies which lie in between multiples of 400Hz?

DISTRIBUTION OF FREQUENCIES BETWEEN MULTIPLES OF 400 Hz:

1) If signal has frequency between say,400 Hz & 800 Hz, then that signal will be detected but Its part would be put in .400 Hz & part in 800 Hz.
2)One way to measure this signal accurately (say 500 Hz) is to take a time record that is (1/ 500) or 2 ms long.Then the signal will lie all in one frequency bin.
3) But this would require changing the sample rate based upon the signal (which you have not measured yet).
4)This not being the good solution, the way to measure the signal accurately is to lengthen the time record & change the span of the spectrum.

Thus, audible frequency range gets divided into bands of frequencies whose magnitude (height of column) depends upon the distribution of frequencies in audio signal & when at particular signal sample, frequencies corresponding to particular band appear, the height of column corresponding to that band varies with amplitude of the signal.



Properties:

1) Linearity:

Log Scale: The dancing bars are nothing but the plot of intensity versus that particular frequency band. Here, intensity is plotted in decibel i.e. log scale. Hence, if u add 2 signals with same frequency & intensity , it does not give you the length of the column doubled. But the length increases lesser than that due to log nature of the intensity column.. Hence, it is not linear in nature.

2) Causality & memory:

The frequency spectrum depends upon the instantaneous value of the input signal. Hence, it is causal & memoryless system.

3)Stability:

If maximum amplitude of the applied sampled signal is A, then the length of the column of bandwidth corresponding to that particular frequency is proportional to (kLog A). where k is finite constant.
If |amplitude | < A
| length of column of band | < kLog A.
Therefore, it is a stable system.

4) Shift Invariant System:

It is shift invariant system, as for input at time t1 , frequency spectrum is not going to change , as it depends upon the instantaneous value.

5) Non-Invertible system:

For system to be invertible, it should be uniquely determined .It happens only when sampling rate is higher than twice the highest frequency component. i.e. If sampling rate is minimum 40 KHz, then it is invertible but for our frequency spectrum where only 50 reciprocal bands are there this frequency is too large. Hence intermediate frequencies get distributed in adjacent bands. Therefore, it is not possible to have original acoustic signal from the dancing bars.

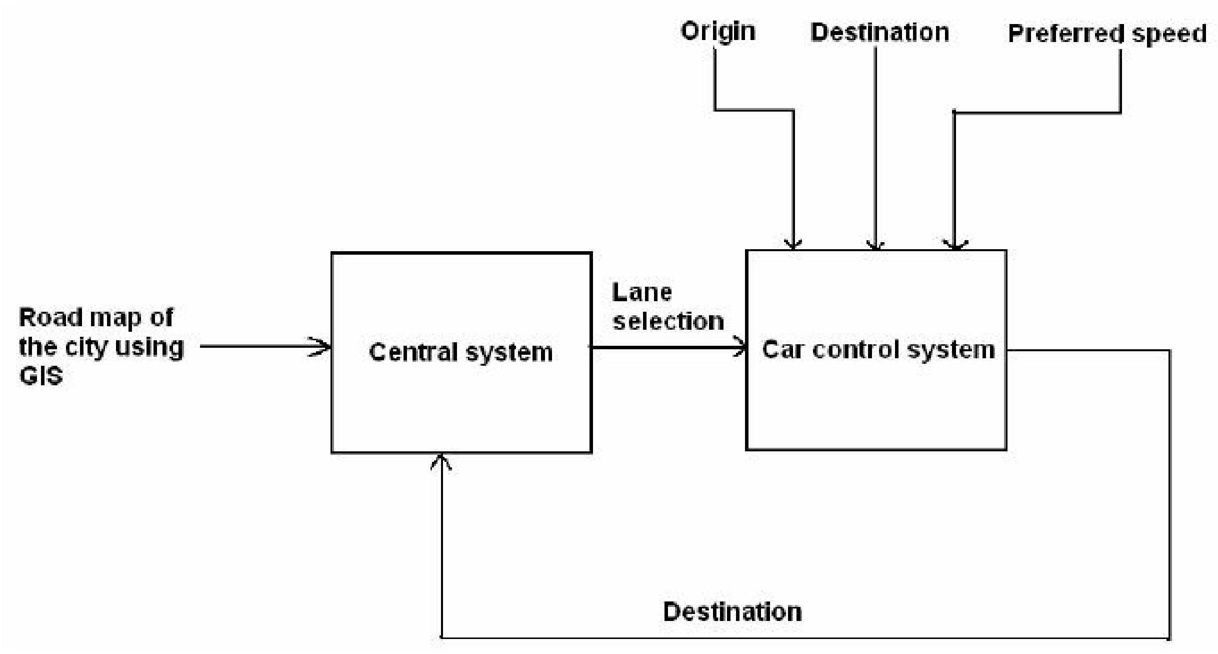
Knight Rider

by

02D07023

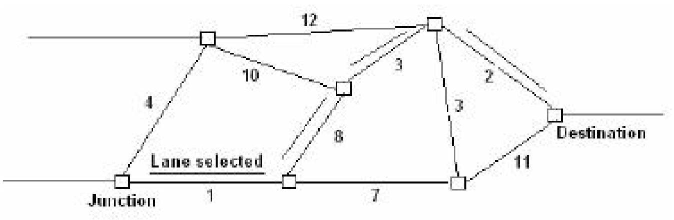
The concept of automation has been introduced for two reasons - travel time reduction and alternate employment of travel time. In the near future, as the population increases, no of vehicles will increase. Therefore traffic systems would be very congested, and lot of time would be wasted if travel path is not optimized properly. If we have autonomous control during travel, the time spent by the driver in operating the car can be spent for some other activities. So we can just enter the destination and sit back assured that we would complete our journey in shortest time possible.

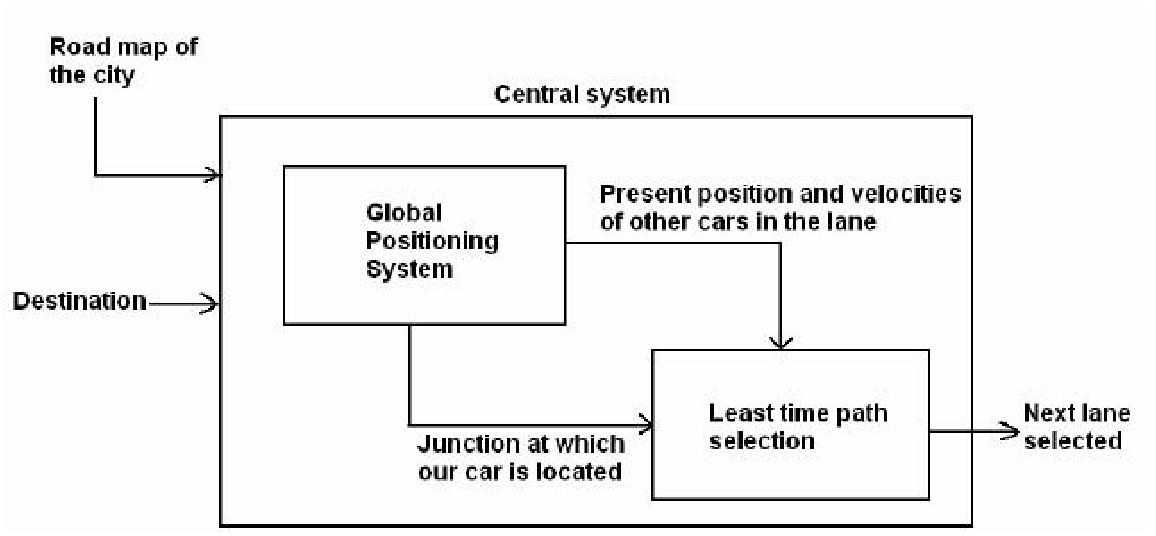
Introduction



For the travel time reduction, we get route decisions from central system. For autonomous control, we use car control system after route has been decided.

Central System





The Global Positioning System (GPS) is formed from an array of 24 satellites orbiting 24,000 kilometers (14,900 miles) above the Earth that orbit the earth twice a day. The basis of GPS is triangulation. Mathematically we need only three distances from satellites to get the coordinates of the GPS receiver (But here practically we use four for technical reasons). Distances can be calculated by time taken for radio waves from satellite to GPS receiver. GPS technology uses a clever way to calculate the time taken for travel. Each satellite generates its gpseudo random codeh (just a complex digital code which consists of on/off pulses) same code is generated at GPS receiver, so what reaches the receiver is delayed version of the signal. Radio waves take finite time to reach the GPS receiver until it syncs up with the satellite's code (correlation of the received code with receivers code reaches maximum). The amount we have to shift back the receiver's version is equal to the travel time of the satellite's version. Inverted Differential GPS (DGPS) is a cheap way for improving navigation accuracies in local area which is what we require in our application. Each car would be equipped with standard receiver and a transmitter to transmit back to tracking office their central system which ties all satellite measurements into a solid local reference.

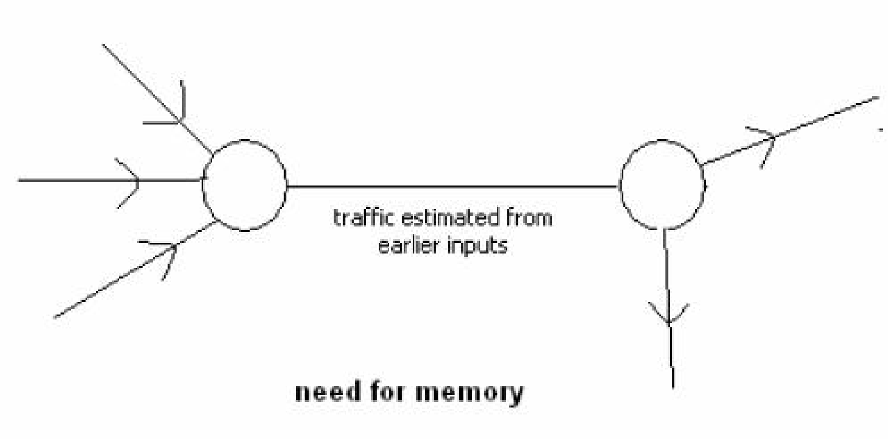
We give as inputs the origin and destination of our journey to the central system. The central system has the map of city and continuously monitors the positions and velocities of all the cars in the lanes through GPS. We want to reach our destination in shortest time possible. The central command system optimizes travel path to travel in shortest time possible. It basically has to solve "shortest-paths problem" to be more precise "single-pair shortest-path problem" at every junction when a decision is required.

Description of single-pair shortest-path problem
In this problem, we are given a weighted, directed graph G = (V, E), with weight function w: E¨(to) R mapping edges to real valued weights and two vertices u, v belonging to V. The weight of the path p is sum of weights of its constituent edges

w(p) = summation ofw(pi, pi-1) p0 = u pk = v
We have to find the shortest path from u to v where shortest path is defined as path with weight of the path being less than or equal to minimum of the weight of the all paths connecting the u and v.
No algorithm for the single-pair shortest-path problem is known to run faster the single-source algorithms (to find the shortest path from a single source to all vertices). We solve single source shortest path using Dijkstrafs algorithm (here the weights is time taken to travel the lane is always non negative)

Application to our problem with
V as the junctions of the city
E as the lane joining the junctions
w as the estimated time taken to travel the lane.

System properties of w
It takes as inputs, the positions and velocities of the all cars and the length of the road. It is the most important calculation for the system is dependent on various factors. It is not entirely dependent on the present input because when the time the car travels in the lane the traffic would have changed, but it would be dependent on the traffic in lanes connecting to the junction at some previous instant and turn ratios (we have to estimate the probability from data collected over period of time). So this system needs to have memory because of the fact that when the decision for shortest path is also being taken on basis of predicted from traffic at earlier instants and collected turn ratios over time.



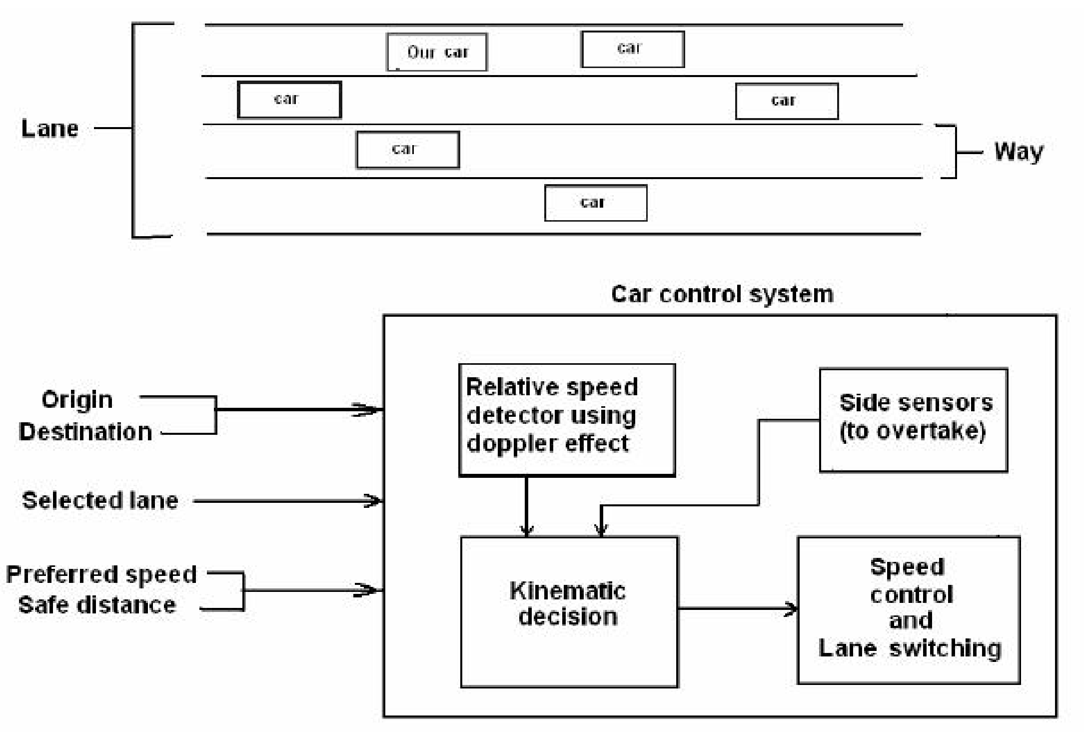
We consider other dependencies to model the real life situation like there is a road block due to some accident (in this case we have to consider the time taken to be infinite to cross the lane as the person would not want to wait for long time). All thought the input is bounded, output for weight function is considered to be infinite for calculation purpose. Suppose the road is under heavy traffic, then it would take us lot of time to reach our destination. In this case we have finite number of cars on the road (saturation) and all have very low speeds due to congestion, thus the input can be said to be bounded. But till it will increase the weight function infinitely, thus the output is unbounded.

Calculations cant be made using linearity principle. This can be proved by counter example - suppose traffic doubles, the time taken to cross the lane does more than the double of the initial value, as the number of cars is doubled, the speed of each car decreases and thus the final weight function is more than the double of the initial value.
Weight function for present traffic can be copied from the similar past traffic (here presents and past refers to time period day). This is very helpful as in weekdays traffic is almost similar, we improve the present estimation because mostly past weight functions would be repeated. Weight of the path as defined above adds up the total time taken to travel between two junctions By solving this problem we get the path which takes the shortest time between the two junctions and thereby the next lane to be taken from the junction.

Our car starts from origin it receives from the central system the lane to be taken at the junction depending on the weight function estimation which is calculated from continuous GPS monitoring. It uses the internal control system to move through until it reaches the next junction. On reaching the next junction it again receives the lane to be taken from the central system depending on the then calculated weight function.

position = origin
while (position is not equal to destination)
ASK_LANE(position)
ASK_LANE(position)
FIRST_EDGE(SHORTEST_PATH(position, destination))
SHORTEST_PATH is calculated using the weight function derived as in the discussion above.
FIRST_EDGE is the starting edge in the shortest path thus calculated.

Car Control System

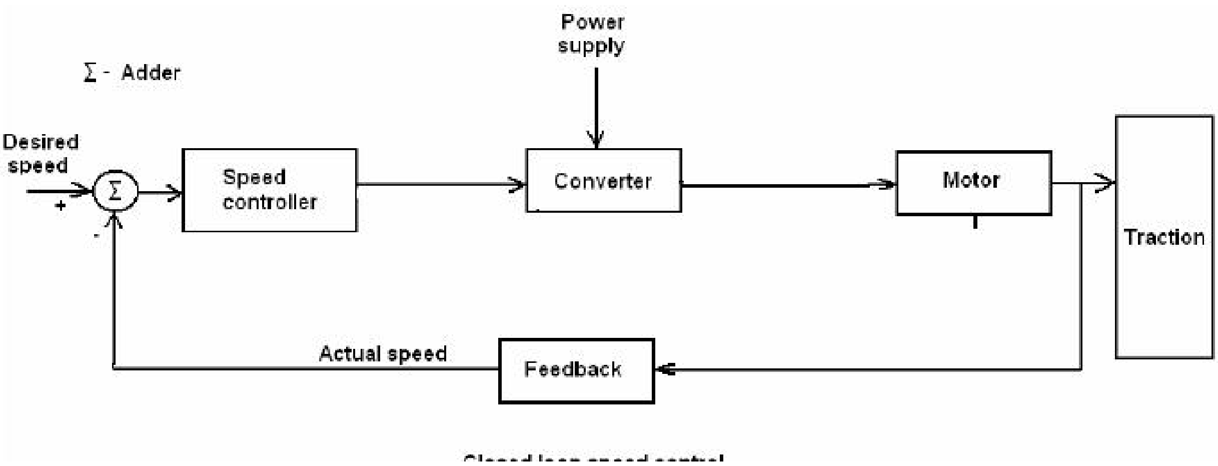


While moving in the lane, we use real-time control system to maneuver our vehicle through the traffic. Suppose we are traveling in multiple-way lane, we have to take decision whether to overtake by switching the lanes or not. Here we usually give as input our preferred speed of the vehicle and also a minimum distance to be kept from other vehicles. We need a sensor on the front side to know the relative speed of the car in front of us and sensors on the side to know whether there is a vehicle in the neighboring way in the lane or not. The sensor can be modeled on the principle of speed gun to get the relative velocity.

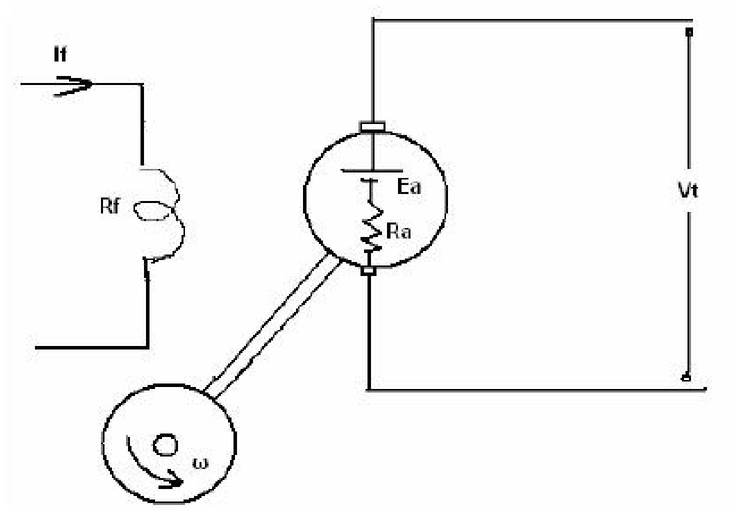
Speed Gun

To detect the relative speed of the car ahead of us we use laser gun technology. The basic principle behind this is that the speed gun emits a short burst of an infrared laser and receives the reflected signal (signal reflected from the car ahead).Now the gun counts the time taken for it to travel the round trip, this way we can estimate the distance (time calculated multiplied by the speed of light) between our car and the car ahead. To calculate the relative speed we just send many such pulses at some high rate and calculate the change in distance between the pulses and then divide it by the pulse rate. To attain greater accuracy one has to simply increase the rate at which infrared pulses are emitted.

Now if the speed of the vehicle ahead of us is less than our speed then we look out for the possibility to overtake through input from side sensors. When overtaking is not possible (may be due to continuous traffic in the side lane), then we go for braking so as to avoid crashing into the vehicle. For braking we use speed control system as discussed below.



Speed control in a DC motor can be achieved by
1) Armature voltage control (Vt)
2) Field Control (Q)
3) Armature resistance control (Ra)
But here we consider only armature voltage control.

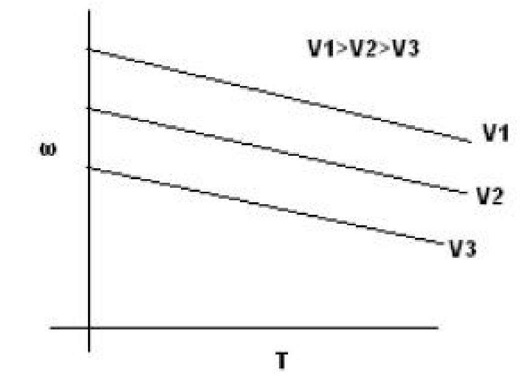


$E_a = K_1 Q\omega = V_t - I_a R_a$
$T = K_1 Q I_a$
Speed is
$\omega = (V_t - I_a R_a)/K_1 Q$
from the above two equations
$\omega = V_t/K_1 Q - R_a \cdot T/(K_1 Q)$^^2

Armature voltage control:

$\omega = K_1V_t - K_2T$
where $K_1 = 1/K_aQ$
$K_2 = R_a/(K_a.Q)$^^2

For a constant load torque speed will change linearly with V_t
Thus we can have a smooth control of speed from zero to base speed by varying V_t using choppers.



TORQUE OF MOTOR VS SPEED CURVE WITH CHANGIIG ARMATURE VOLTAGE

The speed is the intersection point between (motor torque vs speed) and (load torque vs speed). Load torque curve is got by solving a second order differential equation involving rotational inertia damping constant and spring constant.

Neuro Electronics

by

Saurabh Gupta

Ayush Mittal

Aneesh Nainani

Abstract:

We have undertaken brief study of two very interesting systems in this project.

1) The Brain
2) A system which takes the output from brain as input, processes it to make sense out of it and reacts accordingly.

Brain being a very complex system has been studied briefly in the light of concepts learnt in our course EE-210. We have studied how we read and comprehend the complex waves that brain generates.
Study of the other system involves how the output from brain (brain waves) is given as input to it, how it processes it and how it reacts to it. We have touched the hardware aspect of this system slightly. We have also seen the interesting applications of this system in practical life.

MOTIVATION

The topic that we have selected has been one of the most perplexing and bewildering subjects for scientists and engineers. The developments in this field are beneficial for development of artificial intelligence technology as well. The most important thing that brain can do but not a machine is pattern recognition and classification. Generating this ability in machines is something a “signals and systems” engineer would find interesting. More over, recent developments in the field of thought controlled devices were very amazing.

This field offers a lot of challenge for future engineers. A lot more than what has been done can be produced in this field. This motivated us to select this topic for our presentation.

The Brain

This system takes as it's input, electrical stimuli from our sense organs and generates as output what are known as "brain waves".

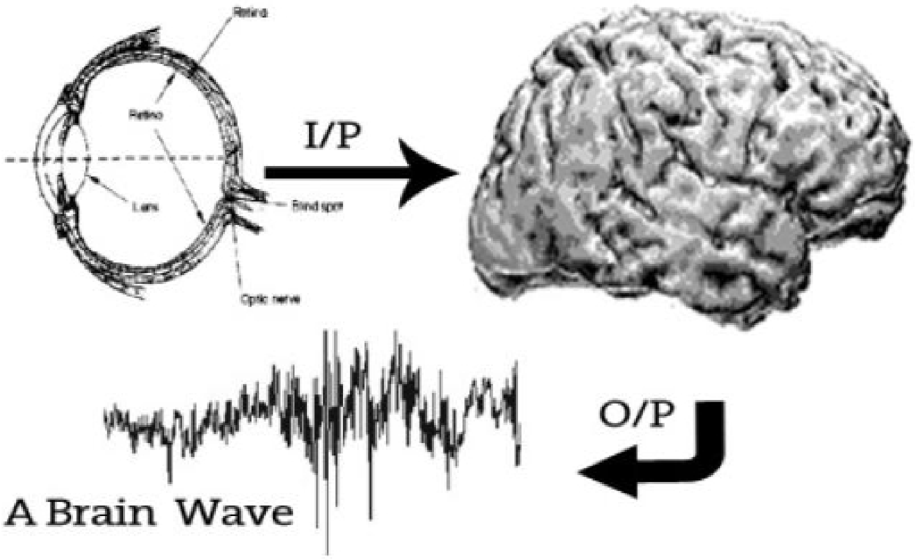


Fig.1 *The Brain System*

Brain has been one of the most complex systems for engineers to study. It is so because we can not characterize it under any of the properties that we have so far seen and many others which make studying a system easier. Brain is
1) not a linear system
2) not a stable system – stability can here be seen as property of the system to come back to same output after deviating from it due to small disturbance. Eg- a vertical pendulum is an unstable system.

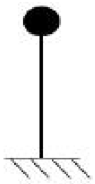


Fig.2 Vertical Pendulum- falls to ground as soon as disturbed slightly.

A small disturbance is enough to deviate brain from equilibrium.
On the whole, brain is a very chaotic system to study. For this system, we have seen how we can apply "signals and systems" to study some properties of it's
1) memory
2) response or brain waves
3) not, quite obviously, memory less.

MEMORY

We will see, under this sub-section, an unconventional but very convincing theory about the form in which brain stores the information. But before that we will look into an important property of FOURIER TRANSFORMATION.

Fourier transform of a function represents the magnitude of complex exponentials in it's Fourier Series.
Now consider a black and white image. We can express it as brightness as a function of (x,y). Let it be B(x,y). If we lose information of B(x,y) over a range of (x,y), we will lose the image from those points also.

Fig 3. A B&W image before and after losing information in spatial domain.

$$F(w_1, w_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(x,y)\, e^{-j2\pi(w_1x+w_2y)}\, dx\, dy$$

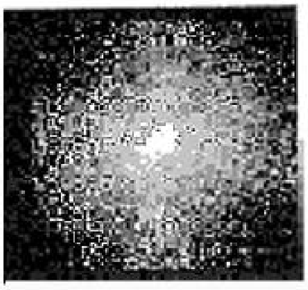


Fig: The graph of the transform is

Now if we inverse transform only a part of this graph and not the complete one, we will still get whole of the image, but only suffering in its resolution. This happens because we lose information in frequency domain and not in spatial domain i.e. we will still have a function (say B'(x,y)) on inverse transformation such that we have information of image over the entire area of original image but only varying a little from previous information.

Fig: Image after inverse transforming a part of the Fourier transform.

Coming back to brain now; there have been experiments which show that if a part of brain's memory is lost, it can still remember something completely but suffering only in finer details. The explanation to these observations is given on the basis of above mentioned property of Fourier transforms. It has been proposed that brain actually stores information, not as it is, but in form of Fourier-like-transform distributed over a certain region of brain.

Memory can now be defined as the experience generated when brain inverse transforms the stored information back to special domain. There are many experimental evidences beyond the scope of this presentation which establish this fact more firmly.

BRAIN WAVES

The response of brain towards the stimuli is in form of electrical signals called brain waves. These waves are actually the voltage generated across the joints between neurons. Brain waves, on the basis of the stimuli that originate them, can be classified in four classes

1) Alpha waves- they originate when person is awake but in a very relaxed state. The frequency of these waves lies between 8 to 13 hz.
2) Beta waves- originate when a person is involved in some thinking process. Their frequency lies between 15 to 25 hz. These are of great significance, as these are the ones which, if analyzed finely, could reveal a great deal about the processes going in brain. And that

is exactly what we need for our “thought controlled devices”.
3) Theta waves –originate when person is sleeping.
4) Delta waves –originate while deep sleep

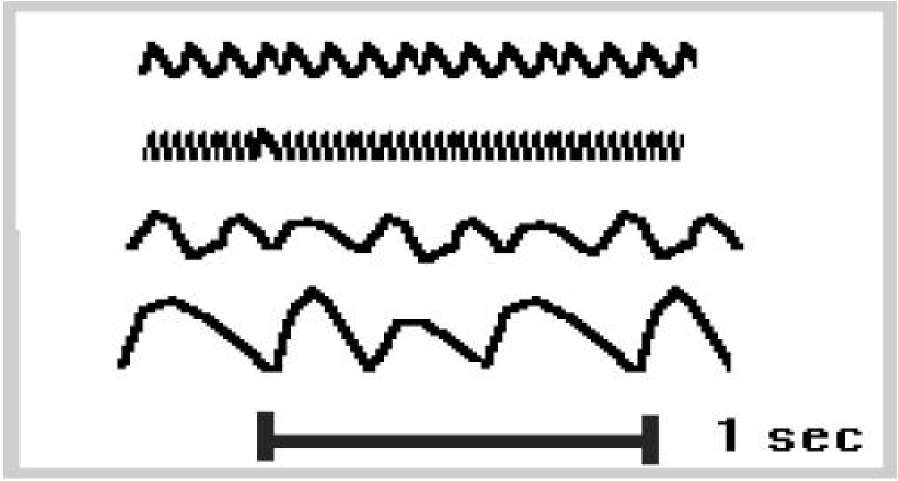


Fig: Four types of brain waves

We now proceed to our second system “THOUGHT CONTROLLED DEVICES”.

Thought Controlled Devices

What does the system do?

It is a system which acts as an interface between our brain and the task we want to accomplish. It reads our brain, tries to understand what it is thinking about and performs the task.

Under this section we will sequentially look at
• input to the system
• processing of the input information
• output of the system

INPUT:

As discussed above, the input to the system are brain waves. Now the question is how we read brain waves. We use a device called EEG (electro encephalo graph) machine to record the activity of brain. This was brought into use about 60 years back and we have developed much advanced devices, but still EEG continues to be the most useful technique. Electrodes are placed in a conventional way on the head. Their number can be as high as hundred. These sensitive electrodes record the minute voltage developed across a neuron and plot the brain waves. But these waves are overlap of several waves with different frequencies and as discussed above, we need to analyze them under different categories to get some meaning out of them.

Fig: EEG pattern of a person under observation

In the next sub section we see how our system can get something meaningful out of this crude input given to it.

PROCESSING

1) Amplification –The signal obtained are very low in magnitude. Small disturbance can cause a large deviation, so the brain waves are amplified.

2) Fourier transform –The signal obtained is transformed to frequency domain which helps in further filtering and analysis of the signal.

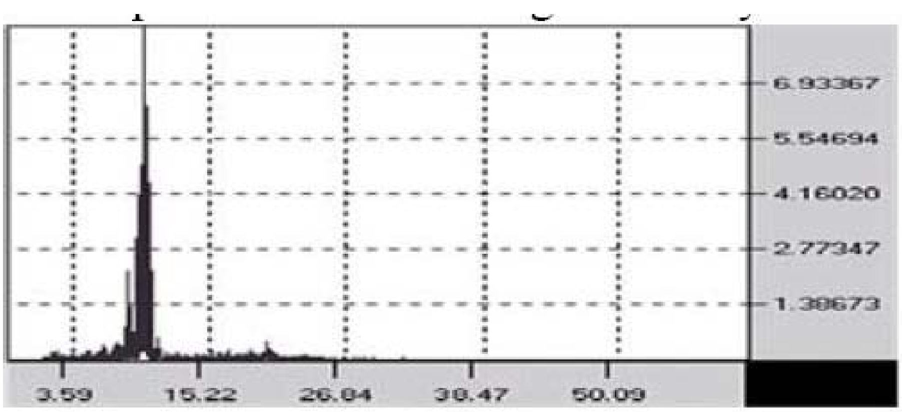


Fig: Fourier transform of a brain wave -frequencies between 8-13 hz can be seen to dominate, which correspond to alpha waves.

3)Filtering –We are interested only in waves of frequency less than 50 hz. So we filter the Fourier transform of the signal with 50 hz low pass filter.

4) Separation into different components – As seen before, brain waves comprise four categories of waves. We are basically interested only in alpha and beta waves. We now need to break the signal into its components which are waves with frequencies in a particular band. So we pass the signal through two band pass filters

- 8 to 13 hz band pass filter for alpha waves
- 15 to 25 hz band pass filter for beta waves

5) Further filtering –The beta wave Fourier transform is further filtered to get waves more specifically related to some stimuli. This is a difficult process, because it is difficult to identify the frequency related to the stimuli.

6) Inverse Fourier transform –The signals are brought back into time domain because two signals can be properly compared in time domain.
A data base of response of brain towards different stimuli is obtained. Special purpose computers are used to compare signals obtained after processing response towards an arbitrary stimulus with the signals in the data base.

OUTPUT

The system as an output executes the task according to the interpretation of brain waves.

Real Life Application

The system, as can be seen clearly, has got very interesting and widespread applications limited only by the extent of our imagination or in other words limited by nothing at all. We will, in this section, see some of the interesting applications of the system in our life that have been realized and many others which we are hopeful to accomplish in coming years.

A very common application that many people today are aware of is “lie detectors”. It works on detection of percentage of alpha waves in the brain waves, which as discussed earlier dominate when person is relaxed. Another application is detection of a word or a sentence being thought or heard by a person.

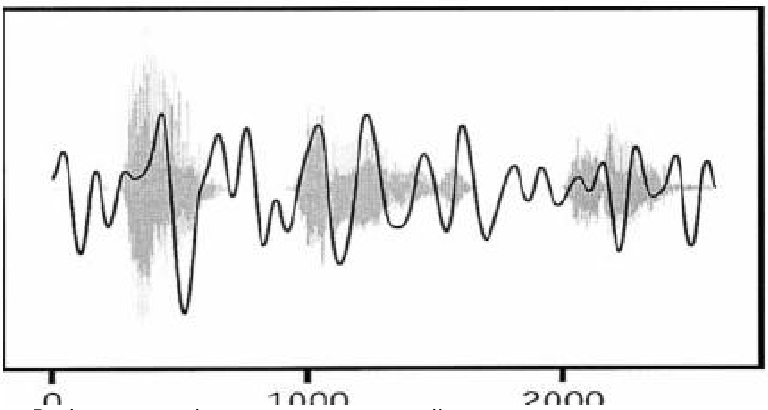


Fig: Brain wave and spectrum corresponding to a sentence

The system has also been put into use for some medical purposes. Using a specific model of this system, the state of brain is detected (relaxed or disturbed). It now produces its own electromagnetic waves which interfere with the brain waves to generate more of alpha waves and relax brain. Movement of paralyzed limbs has also been made possible now. Muscle stimulating electrodes are placed under skin which are controlled by signals generated by a specific form of our system which is implanted in brain. This application has been put into use over a considerable number of patients.

The concept of alpha waves has been put into use for making teaching and learning more effective.A very interesting and recent application of this system made news headlines around the world. A monkey was trained to play a simple computer game in which a dot moved in the direction in which monkey moved the joystick. A chip was implanted in monkey’s brain which served the purpose of reading the impulses in brain related to monkey’s hand movement. Now the joystick was disconnected from the computer and monkey was made to play the game again. Monkey’s brain not perceiving

Fig: Monkey's brain waves being used to play a game

the situation generated waves as in normal condition which were transmitted by the chip to a special purpose computer which processed them and made the dot on the screen move just as before. This significant progress in this field will have wide spread effects in coming years. It can be used, for example for enabling handicapped people to do simple computer operating like checking mails etc. The limitation of this system right now is that electrodes have to be placed over or inside the brain. In coming years engineers hope to develop technology which will enable reading brain waves from a distance and even in group of people. In coming years, scientists hope to be able to harness magnetic waves of brain to produce similar results. The advantage of using them is that they can be collected at a distance from a person. Today scientists also imagine to have made a satellite which goes around the earth reading the brains of people from such a distance. Sounds unimaginable but can be made true in coming years.

Yet another very interesting application of this system that can be thought of is programming the brain itself. We simply need to decode the language that brain uses. After that it would not be difficult to create electromagnetic waves which after interfering with brain waves would make a person think the desired way. Defense agencies of some countries are suspected to be supporting such a project with the ambition of controlling minds of entire countries in a situation like war.

A very interesting observation was made about so called haunted houses and palaces that most of them were situated near a source of high electromagnetic activity. Doctors and scientists use this observation to strengthen their claim that brain can be controlled to an extent of making people hallucinate.

This field offers a lot of challenges for future engineers, and has a great scope of advancements. We hope to get a chance in future to work more over the topic.

ACKNOWLEDEMENT

We are thankful to the instructor of the course EE-210 Prof. V.M. Gadre who gave us a chance and inspired us for this project.

Image Processing

by

02D07033

1. Basic Terminology

1.1 Definition:

- **Signal:** signal can be represented mathematically as a function of one or more **Independent** variable. Example: Speech signal can be represented by acoustic pressure as a function of time.

- **Types of signal :**

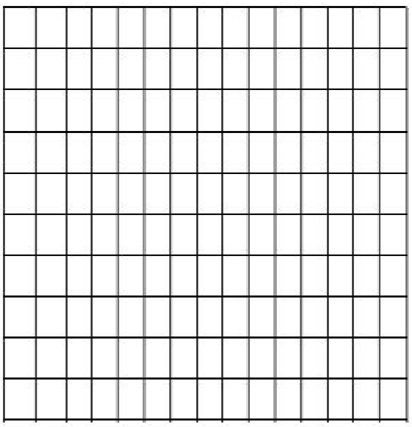
  Continuous: - Here independent variables take a set of continuous values.

  Discrete: - Independent variables take a set of discrete values.

Extending this idea of signals, **image** can be defined as "*Brightness f(x, y) as a function of two independent set of spatial variables x and y.*"

When the independent variables (x and y) and the amplitude values of f(x, y) are all discrete finite quantities, we call the image a ***digital image***.

Since x and y are discrete, <u>image can be represented as 2 dimensional matrix a[m, n]</u>. Every element of this matrix is called **pixel.** Brightness corresponding to a pixel is the magnitude of that pixel.



Pixel

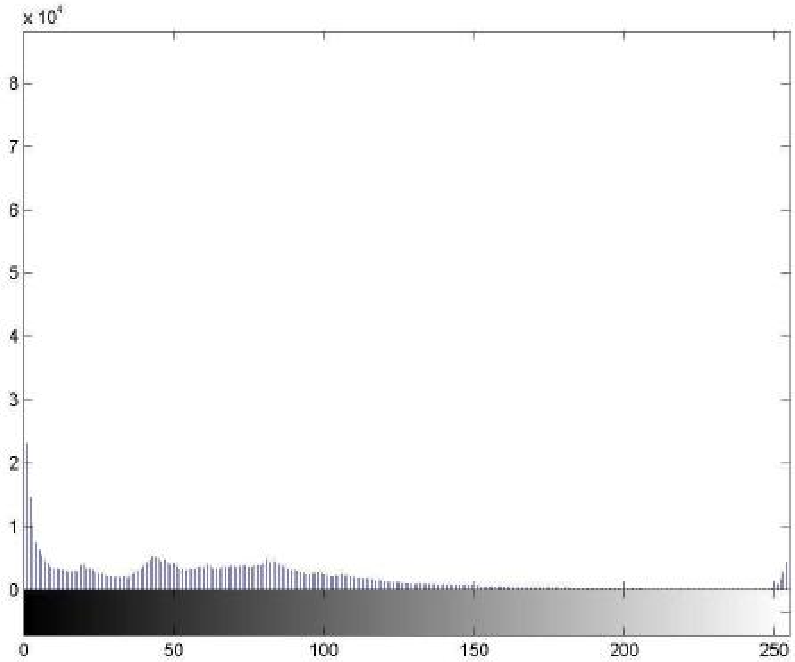
Gray level: Intensity of image is often called as gray level.

Interval between the minimum and maximum intensity that we can have is *Gray scale*. It's generally convenient to shift the interval numerically to interval 0 to 1, where ' 0'efers to black and ' 1'efers to white.

Gray level histogram: It's often useful to represent various intensities in the image statistically. This is done by gray level histogram. Histogram is a discrete function $h(r_k) = n_k$ where r_k is the kth gray level and n_k is the number of pixels in the image having gray level n_k.

For example consider the following image:

The following is it's gray level histogram:



1.2 Types of images:

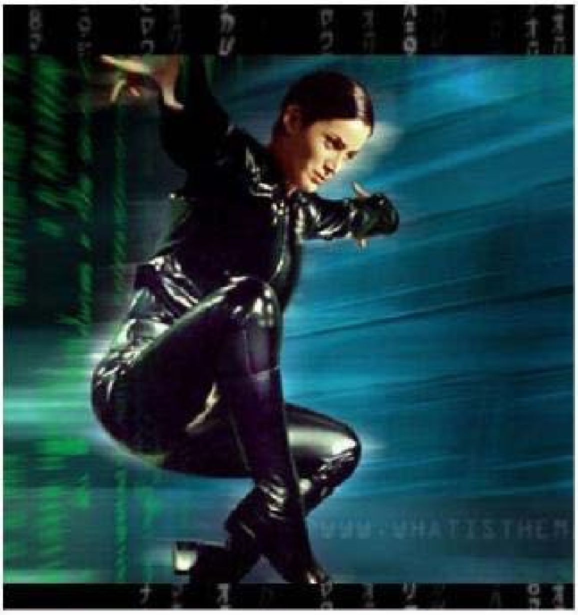
We often encounter 3 types of images in real life.

1. **Binary image**: Image containing only black (0) and white (1) pixels.
   It can be represented as matrix of 0's and 1's.

2. Intensity image: Image consisting of intensity values is called intensity image.

2. **True color (RGB) image** : Image in which each pixel is specified by 3 values, one each for *Red* , *Green* and *Blue* component. So it can be represented as a 3 dimensional matrix.

Example:



However there are other types of images also.

Basic Mathematical tools

Since we have defined image as a 2 dimensional signal almost all mathematical tools used in analysis of signal can be used almost as it is to images.

Let's define what we mean by convolution of images.

2.1Convolution: Convolution for a 1 dimensional signal can be defined as:-

$$f * g(x) = \int_{-\infty}^{\infty} f(x-\xi) g(\xi)\, d\xi$$
$$= \int_{-\infty}^{\infty} g(x-\xi) f(\xi)\, d\xi$$

Where f & g are two functions.

Before identifying the properties we first describe the mental process which leads to the graph of this integral:

(i) Take the graph of the function $f(\xi)$ and flip it around the vertical axis ξ.

The image above *when* passed(convolved) through a filter $h = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$ gives the following blurred image.

(ii) Slide that flipped graph to the right by an amount x to obtain the graph of $f(x-\xi)$.

(iii) Multiply this graph by the graph of $g(\xi)$ to obtain the graph of the product $f(x-\xi)g(\xi)$ function.

(iv) Find the area under this product function. As one slides the flipped graph to the right, this area generates the graph of $f*g(x)$.

Ex: *Periodic train of Gaussians via convolution*

Consider the graph of the Gaussian

$$f(\xi) = e^{-(\xi-c)^2/2b^2}$$

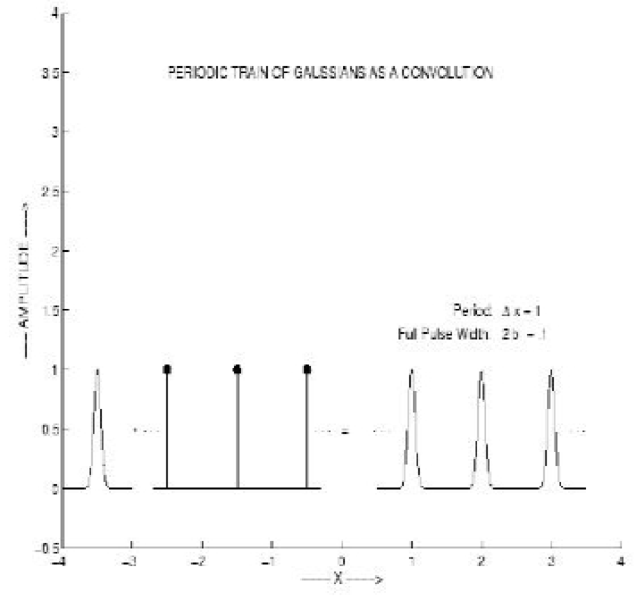
having full width **2b** centered around $\xi = c$,

$$g(\xi) = \sum_{n=-\infty}^{\infty} \delta(\xi - n)$$

Convolution of function $f(\ )$ with $g(\ )$ is found out to be $= \sum_{n=-\infty}^{\infty} f(\xi - n)$

$$= \sum_{n=-\infty}^{\infty} e^{-(x-n-c)^2/2b^2} \ .$$

This is a periodic train of Gaussians, and the period is. $\Delta x = 2\ \Delta x = 2b$



2.2 Fourier transforms:

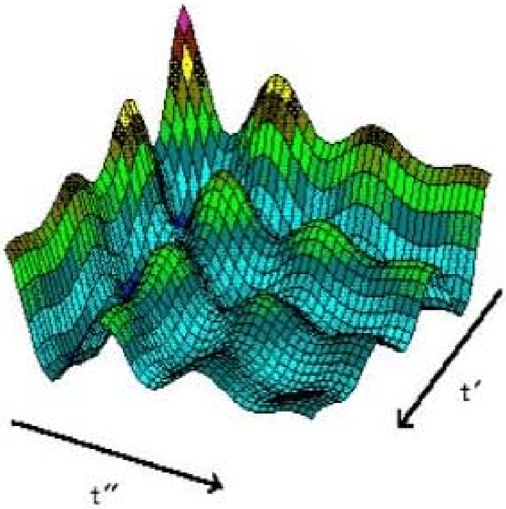
Any function or signal can be thought of as a combination of sinusoidal function with different frequencies and if the signal is aperiodic , we will get all possible frequencies .

The same idea can be extended to any 2 dimensional spatial signal .

Let's see what we actually mean by Fourier transform of 2 dimensional signals.

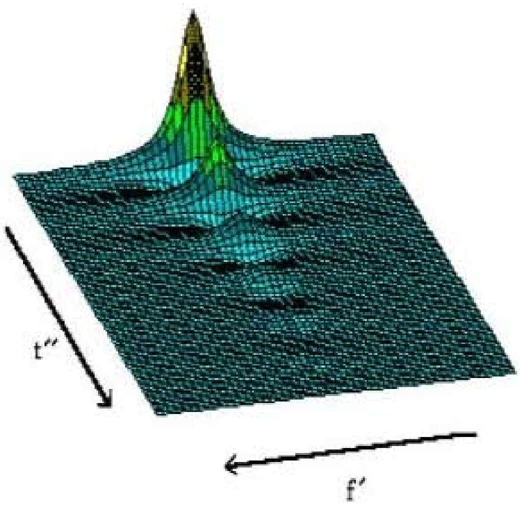
The two-dimensional Fourier transform (2-DFT) is an FT performed on a two dimensional array of data.

Consider the two-dimensional array of data depicted in the figure.



This data has a t' and a t" dimension.

A FT is first performed on the data in one dimension and then in the second. The first set of Fourier transforms are performed in the t' dimension to yield an f' by t" set of data.



On the same line of thought, In 2 dimension the convolution of two functions $f(x,y), g(x,y)$ is given by

$$h(x,y) = f(x,y) * g(x,y) = \iint f(x',y')g(x-x',y-y',)dx'dy'$$

In case of digital images the convolution can be visualized as follows

In convolution, the value of an output pixel is computed as a weighted sum of neighboring pixels. The matrix of weights is called the *convolution kernel*, also known as the *filter*.

For example, suppose the image is

```
A = [17  24   1   8  15
     23   5   7  14  16
      4   6  13  20  22
     10  12  19  21   3
     11  18  25   2   9]
```

and the convolution kernel is

```
h = [8   1   6
     3   5   7
     4   9   2]
```

The following figure shows how to compute the (2, 4) output pixel using these steps:

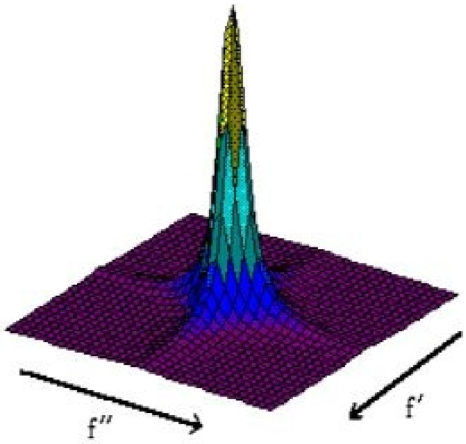
1. Rotate the convolution kernel 180 degrees about its center element.

2. Slide the center element of the convolution kernel so that it lies on top of the (2,4) element of `A`.

3. Multiply each weight in the rotated convolution kernel by the pixel of `A` underneath.

4. Sum the individual products from step 3.

The second set of Fourier transforms is performed in the t" dimension to yield an f' by f"set data.



DEFINITION: **-** Given a function $f(x, y): R^2 \rightarrow R$ modeling for example the image intensity at a point (x, y) its two dimensional Fourier transform is defined by

$$F(u, v) = \int_{\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i2\pi(ux+vy)} dx dy \qquad (1)$$

Here (u,v) stand for the spatial (x and y) frequency contents of the signal $f(x, y)$.

The inverse Fourier transform is given by the formula

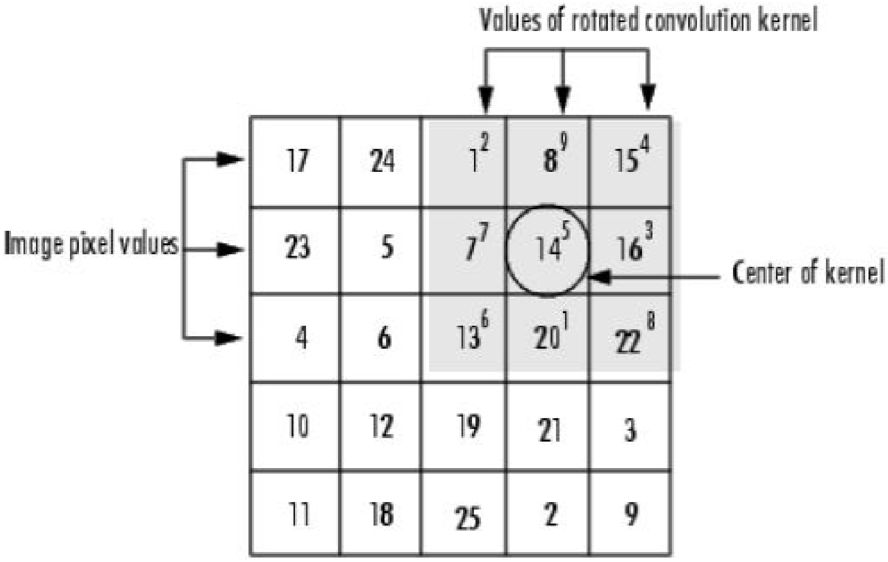
$$f(x, y) = \int_{-\infty}^{\infty} \int_{\infty}^{\infty} F(u, v) e^{i2\pi(ux+vy)} du dv \qquad (2)$$

Discrete Fourier transform of a function (image) f (x, y) of size $M \times N$, equation is given by

$$F(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp(-j2\pi (ux/M + vy/N))$$

Hence the (2,4) output pixel is

$$1 \cdot 2 + 8 \cdot 9 + 15 \cdot 4 + 7 \cdot 7 + 14 \cdot 5 + 16 \cdot 3 + 13 \cdot 6 + 20 \cdot 1 + 22 \cdot 8 = 575$$



In the frequency domain, we have $H(u,v) = F(u,v)G(u,v)$

This expression must be computed for the values of **u** = 0,1,2,. . ,M -1 , and also for **v** = 0,1,2,….,N -1.

Similarly, given F(u, v) ,we obtain f(x, y) via the inverse Fourier transform given by the expression

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v)\exp(j2\pi(ux/M + vy/N))$$

for **x** = 0,1,2,..,M-1 and y = 0,1,2,…,N- 1

Above two equations comprise the two-dimensional, discrete Fourier transform(DFT) pair.

Sampling

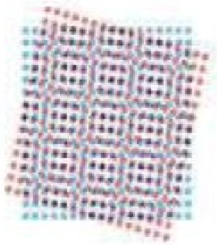
A methodology of representing a signal with less than the signal itself.

Why do we need sampling?

As obvious from definition, *sampling is needed to represent signal in more economic forms.* For instance, in terms of space needed to store the image.

However, sampling can have adverse effects on quality of the image .

Ex : Moiré's pattern

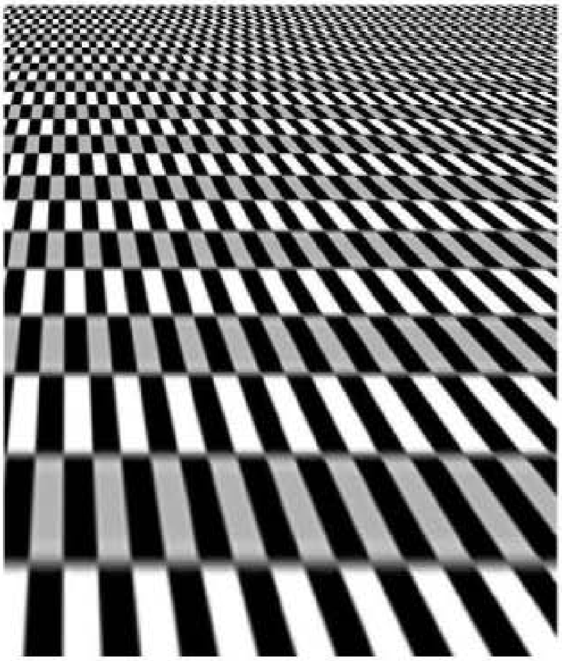


Above effect is "ALIASING ".

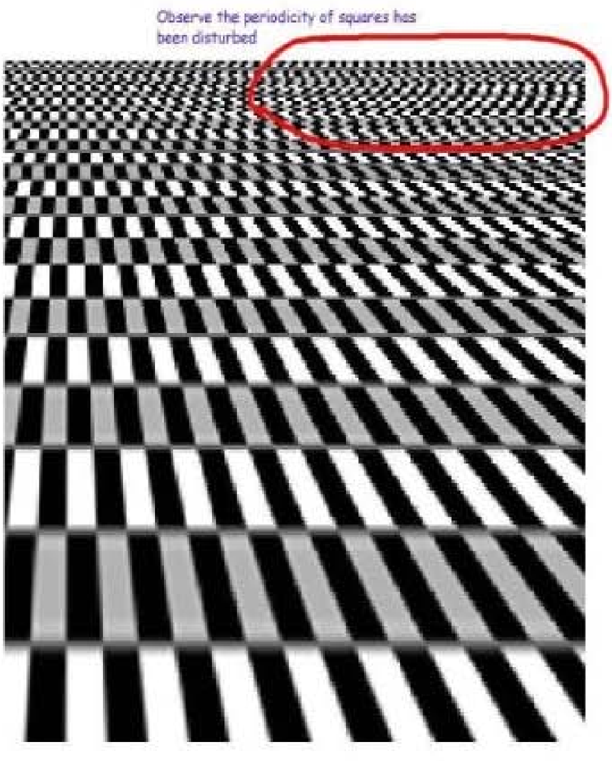
Passing a signal through an ideal low pass filter is same as convolving the input signal with the impulse response of the ideal low pass filter.

when the above photo is passed through a low pass filter, the resultant image obtained is as shown below.

Aliasing can be easily shown in by the following checkerboard image



Now if we sample this image with inappropriate sampling frequency we will get the following

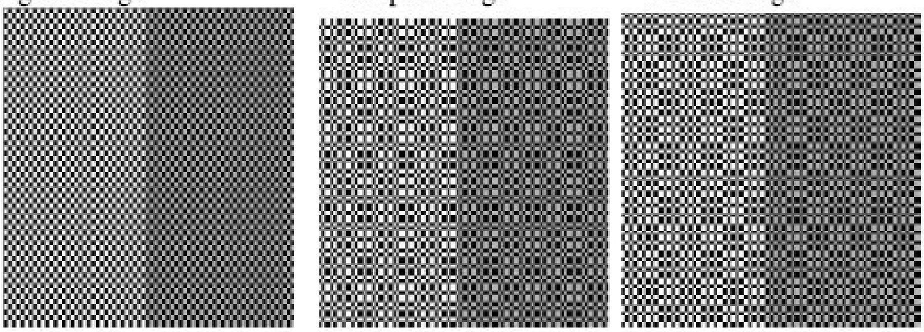


Effect of low pass filter :

Although we can avoid aliasing using low pass filter , the originality of the image is reduced considerably i.e. we find a considerable degradation in the image in terms of its sharpness. This is clear seeing the above 2 images

Therefore the more we band limit the filter, the more we have to compromise on images quality (sharpness).

Original image Sampled image Further image



Hence to avoid 'Aliasing' w e need sampling theorem

3.1 Sampling theorem:

In layman' term, the theorem may be rephrased to say that to measure the frequency of any repetitive event by sampling it , then you have to sample the event at a frequency at least twice the frequency of the event itself.

Sampling of images:

As in the case of 1-D signals, band limited images can be recovered from periodic samples. Thus for example if the image $f(x,y)$ is band limited, that is to say that if

$$F(u,v) = 0 \quad \text{for } |u| > \frac{\pi}{\omega} \quad |v| > \frac{\pi}{h}$$

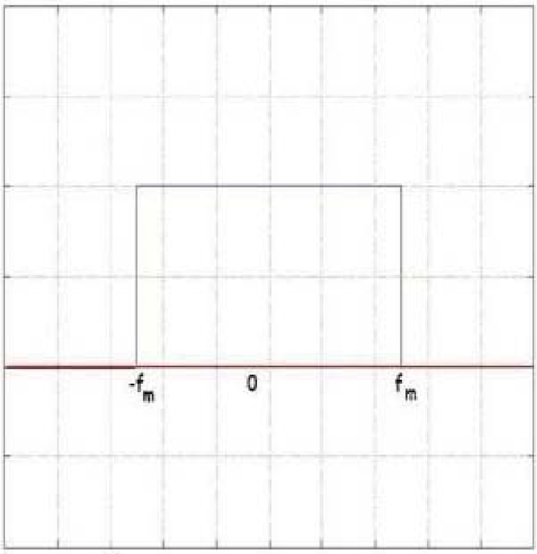
then $f(x,y)$ can be reconstructed by its samples taken

every w units in the x direction and every h units in the y direction, and we can write

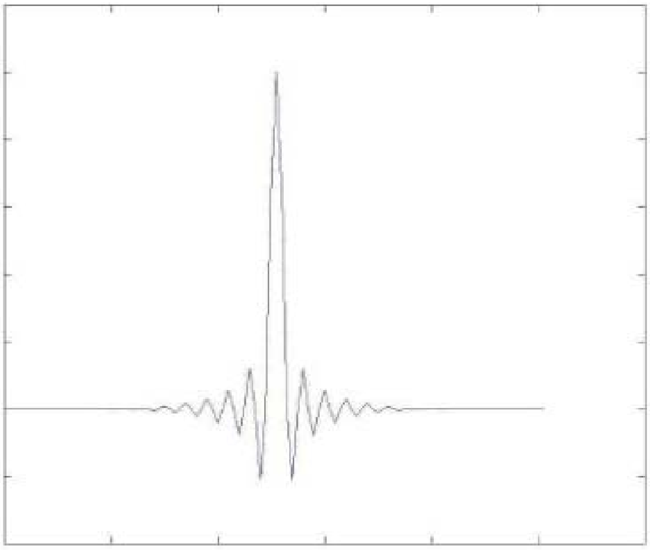
$$f(x,y) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} f(k\omega, hl) \frac{\sin(\pi(x/\omega - k))}{\pi(x/\omega - k)} \frac{\sin(\pi(y/h - l))}{\pi(y/h - l)}$$

However in no practical case the image is band limited. Therefore, there will always be some overlapping between original and its aliased copies. Hence, we need to band limit the signal in order to avoid aliasing. And this is done by **LOW PASS FILTER**.

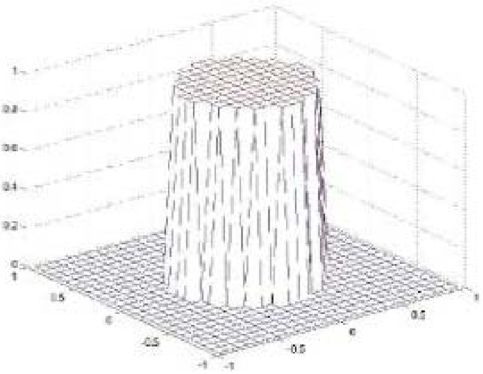
1-dimensional plot of ideal low pass filter.



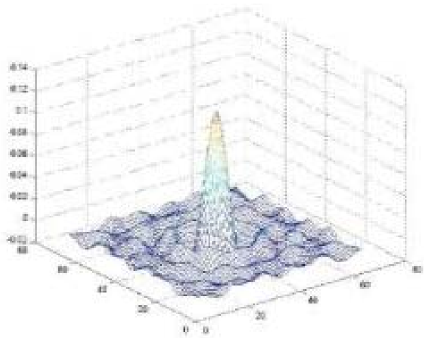
Impulse response of an ideal low-pass filter.



<u>A similar extension of above idea leads to 2 dimensional low pass filter .</u>



2-dimensional ideal low pass filter



impulse response of the filter

Interpolation

Definition: *Interpolation is the process used to estimate the image value at a location in between image pixels.*

Almost every image in practice is not band limited, hence it is certain that we will loose some of the information . Therefore if we try to reconstruct the signal, we will have to use approximations.
3 simplest kind of methods for approximate reconstruction are :

Nearest-neighbor interpolation: The output pixel is assigned the value of the pixel that the point falls within. No other pixels are considered.
Bilinear interpolation: the output pixel value is a weighted average of pixels in the nearest 2-by-2 neighborhood.
Bicubic interpolation, the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighborhood.

This is over sampled version of original image

Interpolated image(Pixels not visible)

Image Enhancement

1. Spatial domain:
2. Frequency domain:

Here we will mainly concentrate on image enhancement techniques in frequency domain. We know the discrete Fourier transform (DFT) of a function (image) f(x, y) of size $M \times N$ is given by the equation

$$F(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \exp(-j2\pi(ux/M + vy/N)$$

and inverse Fourier transform is given by equation

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) \exp(j2\pi(ux/M + vy/N)$$

Basics of filtering in the frequency domain:

(1). Multiply the input image by $(-1)^{x+y}$ to center the transform.
(2). Compute F (u, v), the DFT of the image from 1.
(3). Multiply F (u, v) by a filter function H (u, v).
(4)Compute the inverse DFT of the result in (3).
(5) Obtain the real part of the result in (4).
(6) Multiply the result in (5) by $(-1)^{x+y}$.

H (u, v) is called a **filter** as it suppresses certain frequencies in the transform while leaving others unchanged.

Let f (x, y) be the input image in step 1 and F (u, v) its Fourier transform.
Then Fourier transform of the output image is given by

G (u, v) = H (u, v) F (u, v)

Filtered image is obtained by taking the inverse Fourier transform of G (u, v)
Filtered image = $F^{-1}[G(u,v)]$.

Any filter can be used in the following processes
(1) smoothing
(2) sharpening
(3) histogram equalization of intensity images

5.1. Smoothing or blurring: It is achieved in the frequency domain by attenuating a specified range of high frequency components in the transform of the given image. For this we generally use the following filters

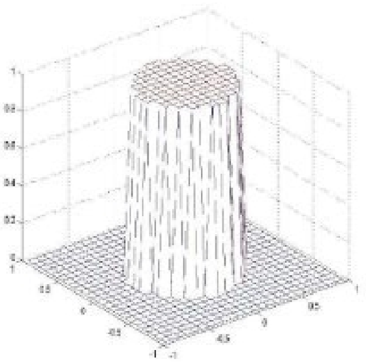
1) Ideal low pass filter
2) Gaussian filter

IDEAL LOWPASS FILTER:

This cut's off all high frequency components of the Fourier transform that are at a distance greater than a specified distance Do from the origin of the transform.

It has a transfer function

$$H(u,v) = \begin{cases} 0 \text{ if } D(u,v) \leq D_0 \\ 1 \text{ if } D(u,v) > D_0 \end{cases}$$



we have already seen the result of passing an image from above filter.

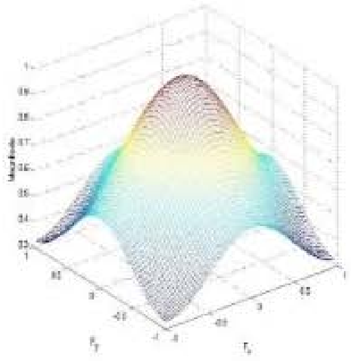
GAUSSIAN FILTER:

The transform function for these filters is given by

$$H(u,v) = \exp(-D^2(u,v)/2D_0^{\,2})$$

D (u, v) is the distance of the Fourier transform from the origin.

D_0 is the cut off frequency.



following is the result of passing the same image (that we considered previously) through Gaussian filter

5.2. SHARPENING:

This can be achieved in the frequency domain by a high pass filtering process, which attenuates the low frequency components without disturbing high-frequency information in the Fourier transform.

For this also we generally use:

1) Ideal high pass filter
2) Gaussian high pass filter

1) IDEAL HIGHPASS FILTER:

This filter is opposite to ideal low pass filter in the sense that it it sets to zero all the frequencies in the circle of radius D_0, while passing, without attenuation all the frequencies outside the circle.

$$H(u,v)=\begin{cases}0 \text{ if } D(u,v)\le D_0\\1 \text{ if } D(u,v) > D_0\end{cases}$$

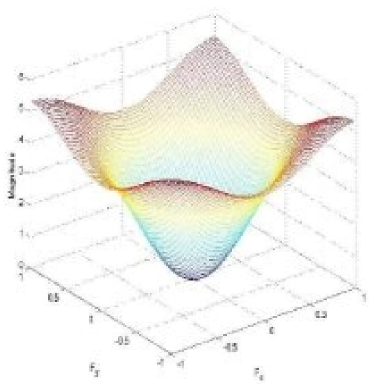
2) **GAUSSIAN HIGHPASS FILTERS:**

The transform function for these filters is given by

$$H(u,v)=1-\exp(-D^2(u,v)/2D_0^2)$$

D_0 is the cut-off frequency location from the origin.

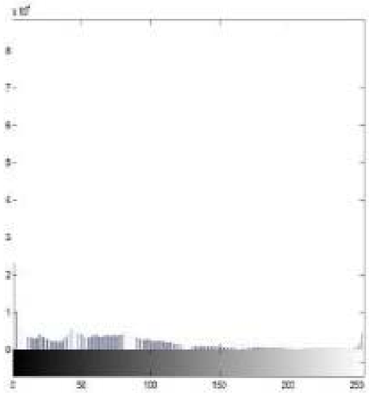
Gaussian high pass filter looks like



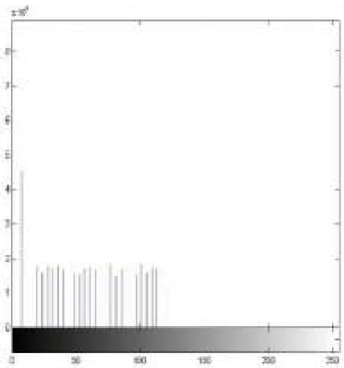
5.3 Histogram equalization of intensity images

The histogram of intensity images are not generally well spaced throughout the gray scale. Histogram equalization technique is done by spreading the histogram over the gray scale. For example, consider the following image

Original image Histogram



Now the following image is obtained by equalizing the histogram of above



.

Above we considered some of the basic terminology, mathematical tools and there effect on digital images. However there are many more filters and techniques that can be understood with the help of mathematical tools that we discussed . For example , Unsharp masking ,Edge detection etc. These simple techniques have very vide application in many areas of science like **Biomedical, space exploration.**

references

1) Digital image processing : by Rafael C . Gonzalez, Richard E. Woods.
2) Signals and Systems : by Oppenheim.
3) Matlab documentation(www.mathwork.com)

CODE DIVISION MULTIPLE ACCESS

by

Sainyum P Gupta (02D07034)

Satish Meena (02D07038)

Aditya Agrawal (02D07020)

Abstract:

This paper studies CDMA data transmission technology. CDMA spread spectrum signaling techniques have gained increasing importance over the past few years. Long used in military systems such as GPS, they are finding increased usage in the cellular communications market. CDMA uses revolutionary techniques such as Rake receivers, Soft Hand-offs and Variable Rate Vocoders to provide a superior cellular service to its subscribers making it the choice data transmission technique to dominate the data transmission market in the near future.

Introduction:

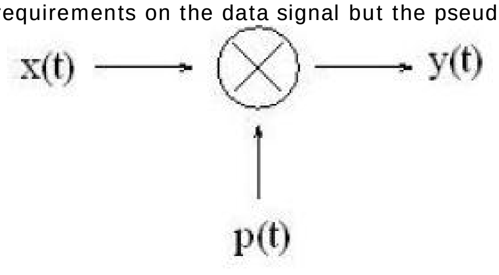
Today's world practically survives on information. From finance to sports, communication has become a crucial factor in running day-to-day activities. In today's competitive markets, people need cheaper and more efficient ways of communicating over long distances. CDMA technology is the solution to this demand. Using ingenious, revolutionary signal transmission techniques, it provides cheaper, clear voice transmission between cellular phones. Besides this CDMA ensures data privacy during the transmission, something alien to outdated transmission systems like TDMA and FDMA. A more factual proof of the advantages of CDMA technology can be inferred from the formation of CDMA based cellular service providing companies like Reliance and Tata, in India, and their wide success. In this presentation, we study an overview of CDMA encoding and transmission systems and techniques.

MOTIVATION:

After careful consideration and debating, we decided to choose a topic for this Presentation that was significant in today's world and would be a useful subject of study that would assist us in the future. CDMA is a rapidly evolving field with countless applications and possibilities of expansion and has taken the cellular phone industry by storm. It stands out against competition from various other data transmission techniques like GSM by head and shoulders. CDMA is the base of the cellular phone industry of the future as proved by its sudden uprising in today's world. It is a more effective technology providing clearer sound quality, lower cost requirements, and smooth inter-cellular functioning capabilities. It ensures data-privacy for the users unlike outdated technologies which can be easily deciphered. In today's world more people are switching to cellular phone networks using CDMA techniques for data-transmission providing cheaper and better services. CDMA makes an intriguing topic of study for us and we are only glad to get a chance to study it in detail.

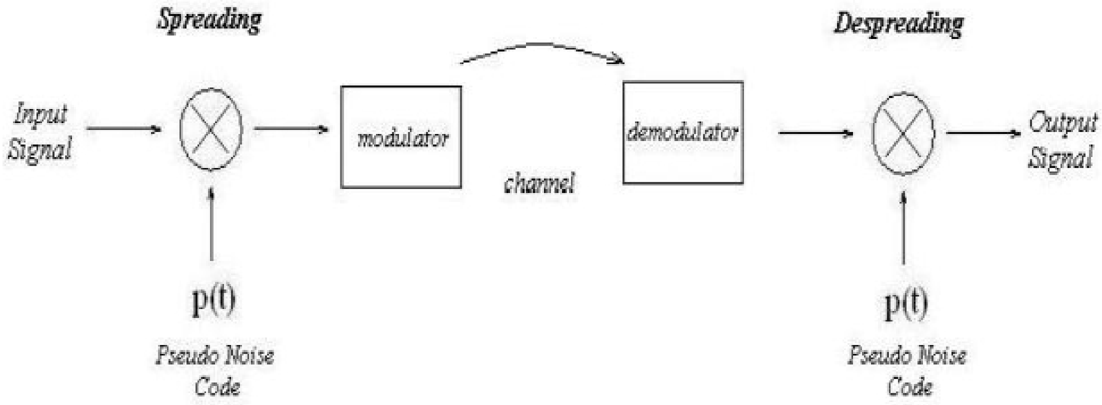
System Specification:

As I said earlier, we are just going to focus on the way signals are transmitted so as to avoid interference and data loss or in their presence minimize it. The system essentially consists of the input or data signal- [x(t)], pseudo noise sequence generator which generates a signal [p(t)] that is unique. Then there is a modulator which modulates the product of the above two signals [y(t)]. This modulated signal is transmitted and at the receiver it is demodulated and the original data signal recovered. There are no particular requirements on the data signal but the pseudo noise needs to have certain properties that help in better transmission.



Working:

As we have stated earlier it works on the principle of Spread Spectrum Technology. In this transmission technique, at the transmitter end, a pseudo – noise code, which is independent of the information data, is employed as a modulation waveform to "spread" the signal energy over a bandwidth much greater than the signal information bandwidth. At the receiver end, the spread signal is received and demodulated using the same pseudo – noise code. This operation is called as Despreading.



Theory Behind Working

CDMA (Code Division Multiple Access) is based on the principle of DSSS (Direct Sequence Spread Spectrum), which is one of the prevailing SS (Spread Spectrum) Technologies in today's world. Spread Spectrum:

In SS technique, the same bandwidth is shared by multiple users, without significantly interfering with each other. The spreading waveform is controlled by a Pseudo – Noise (PN) sequence, which is binary random sequence. This PN is then multiplied with the original baseband signal, which has lower frequency, which yields a spread waveform that has a noise like properties. In the receiver, the opposite happens - the passband signal is first demodulated, and then despread using the same PN waveform. An important factor here is the synchronization between the two great sequences.

Pseudo Noise (PN):

It is the key factor in DSSS systems.
A Pseudo – Noise or Pseudo – Random sequence is a binary sequence with an autocorrelation that resembles, over a period, the autocorrelation of a random binary sequence. It is generated using a shift Register, and a combinational logic. They have the following important properties:

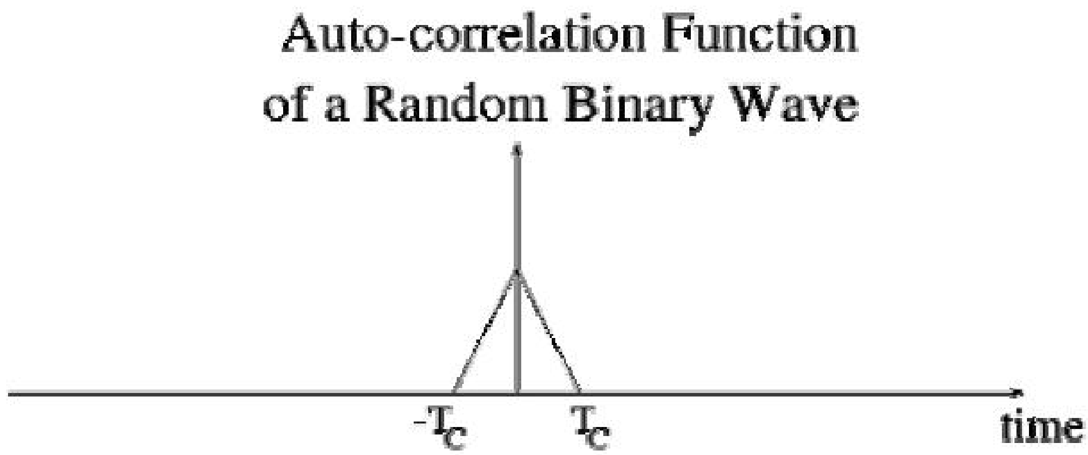
· **Balanced:** The codes should be ``balanced'': The difference between ones and zeros in the code may only be 1. This last requirement stands for good spectral density properties (equally spreading the energy over the whole frequency-band)

· **Single Peak auto-correlation function:** The codes must have a sharp (1-chip wide) autocorrelation peak to enable codesynchronization. · **Deterministic:** The subscriber station must be able to

independently generate the code that matches the base station code. It must appear random to a listener without prior knowledge of the code

Process Gain:

The auto-correlation function of a random binary sequence is a triangular waveform as in the following figure, where TC is the period of one chip.

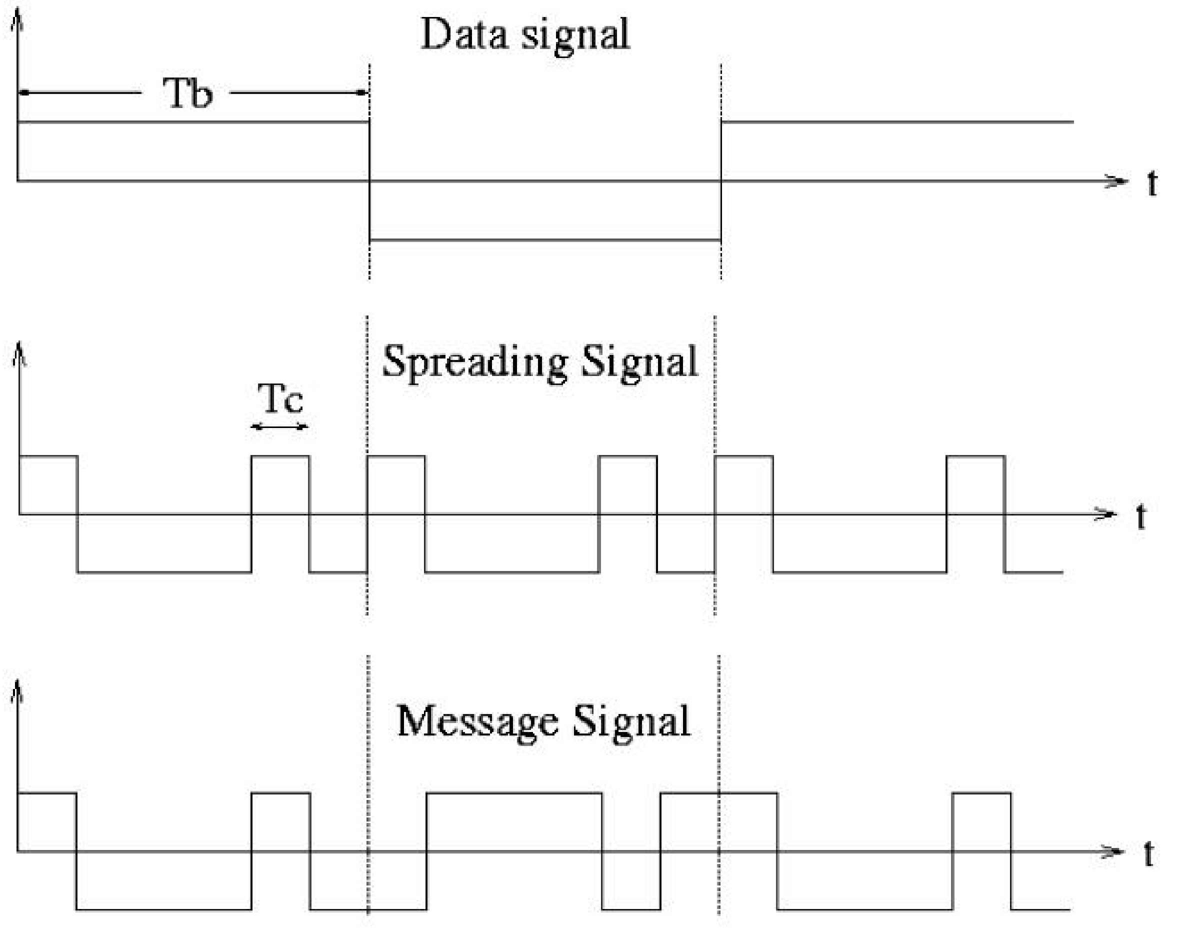


The spectral density of such a waveform is a Sinc function squared, with first zeros at ± 1/TC

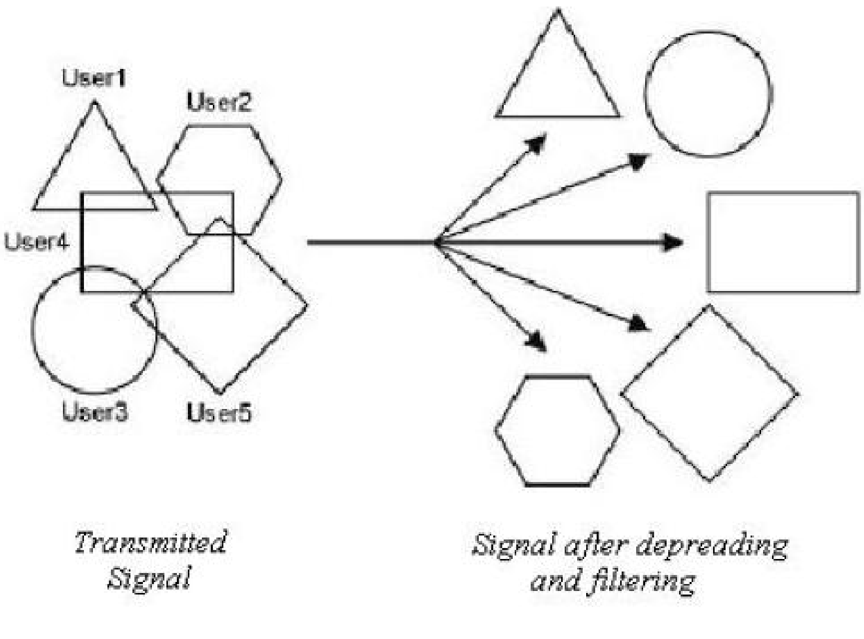
Process Gain: Since multiplication in the time domain corresponds to convolution in the frequency domain, a narrow band signal multiplied by a wide band signal ends up being wide band. One way of doing this is to use a binary waveform as a spreading function, at a higher rate than the data signal. Here the three signals corresponds to x(t), p(t) and y(t) discussed below. The first two signals are multiplied together to give the third waveform. Bits of the spreading signal are called chips. In the figure shown below, Tb represents the period of one input data bit and Tc represents the period of one chip. The chip rate, 1/Tc, is often used to characterize a spread spectrum transmission system. The Processing Gain or sometimes called the Spreading Factor is defined as the ratio of the information bit duration over the chip duration:

$$PG = SF = T_b / T_c = \frac{\text{Bandwidth of transmitted signal}}{\text{Bandwidth of input signal}}$$

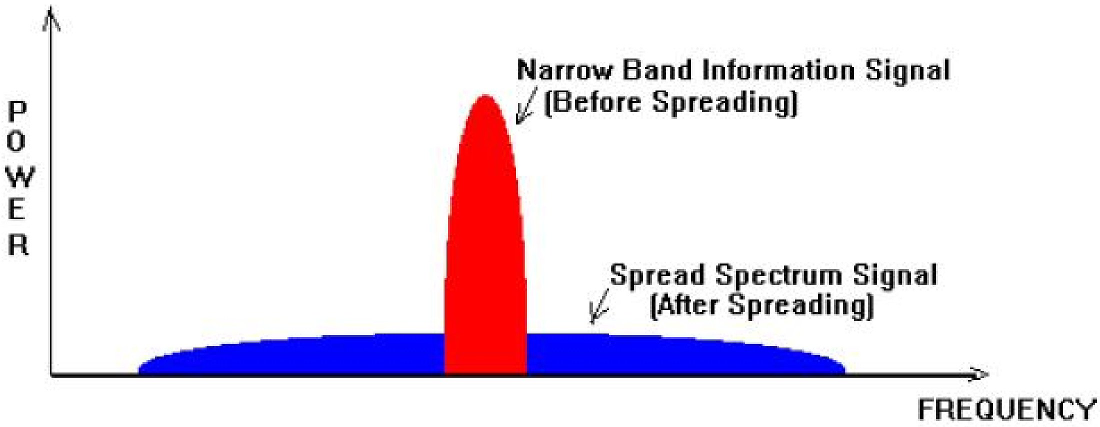
Hence, it represents the number of chips contained in one data bit. Higher Processing Gain (PG) means more spreading. High PG also means that more codes can be allocated on the same frequency channel.



Real Life Application



The most common application of spread systems is in mobile phones (in general multiple access communication systems). They use a particular type of direct sequence spread spectrum transmission called Code Division Multiple Access or CDMA in short. In this all users are given a unique pseudo noise code. When they transmit signals the signals are multiplied by this code and hence only a receiver having the same code sequence can recover data from the signal. The signals sent by other users will be equivalent to noise. The main advantage of this technique is that the users can use the whole of the bandwidth assigned and there is no frequency or time restriction as observed in FDMA or TDMA techniques.(refer Glossary)



CDMA is not exclusively restricted to DSSS. Frequency Hopping technique can also be incorporated in the signal transmission so as to increase the security and resistance to interference. Spread Spectrum techniques are also used in Wireless Networks and the recently developed Bluetooth technology.

Advantages of CDMA over others:

· **Capacity:** Because of its unique SS technology, mobile phone service providers can handle more customers on a CDMA network than previous networks.

· **Quality:** Improved call quality, with better and more consistent sound. Dropped calls are minimized.

· **Enhanced Privacy:** Data bits used to convey information are mixed with PN code that is known only to the base station and the individual mobile phone.

· Improved Coverage, allowing for the possibility of fewer cell sites.

· Increased talk time for mobile phones.

· **Bandwidth on Demand:** When one mobile phone is using less bandwidth, more is available for others.

Conclusion:

Thus we have seen how spread spectrum works and its advantages in day to day life. There are however other techniques in SS that are also equally advantageous. One of this is Frequency Hopping Spread Spectrum technique. Bluetooth – a technology being extensively developed for wireless networking involves the use of frequency hopping. There are however some disadvantages of SS the main being making the system complex and not easy to understand. Another is the implementation costs though this is likely to reduce once the chips used are more easily available. However due to the lack of better techniques of signal propagation SS is here to stay.

Glossary:

FDMA: In Frequency Division Multiple Access, each user is given a particular frequency which he can use at all times. Hence this technique requires very good filtering abilities and interference will be high if users are high.

TDMA: In Time Division Multiple Access, the entire bandwidth is available for use but only at particular times for one user. Therefore this demands high degree of synchronization between the users.

Auto-correlation: The auto-correlation for a periodic signal of period T is defined as follow:

$$R_i(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} W_i(t)W_i(t-\tau)dt$$

It defines how much a function correlates with a time shifted version of itself, with respect to that time shift.

Cross-correlation The cross-correlation for periodic signals of period T is defined as:

$$R_i(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} W_i(t)W_i(t-\tau)dt$$

It measures how much two different signals, Wi and Wj, one shifted in time with respect to the other, correlate as a function of that time shift.

Acknowledgements

We thank Prof V.M.Gadre, our course instructor, for giving us this opportunity to do this presentation, for encouraging us to explore the practical applications of this course. It was a learning experience and fun. We also thank the course associates for reviewing the material and extend our gratitude to all those who have helped us in making this presentation possible.

References

· Wireless Digital Communications (Modulation and Spread Spectrum Applications)---Dr.Kamilo Feher.
· Communication Techniques for Digital and Analog Signals --- Morton kanefsky.
· Wireless Networking Notes --- Techfest 2003
· http://www.sss-mag.com/ss.html#tutorial
· http://www.sss-mag.com/primer.html#vpintro
· http://www.cwt.vt.edu/faq/dsss.htm

SPEAKER VERIFICATION

by

Vinodkumar Pralia(02D07039)

Dharamveer Meena(02D07037)

Abstract:

This paper discusses the text dependent Speaker Verification system. It shows the basic principles behind this technology. Some important applications of this speaker verification system are discussed in brief.

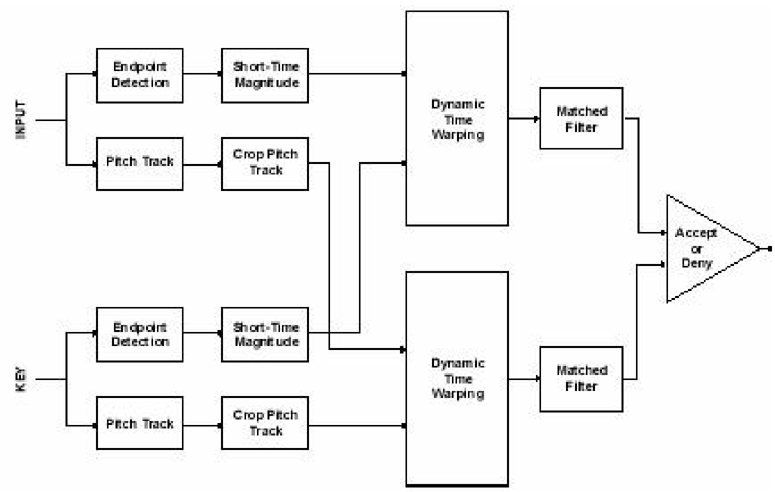
INTRODUCTION:

Speech contains information about the identity of the speaker. Speaker Verification is confirming the identity of an individual from his speech. There are several possible applications for speaker verification in security and identification systems. Some examples include automatic teller machines (ATM) and home or vehicle access. There are several possible implementations for systems that perform speaker verification. The implementation which most commonly comes to mind is that of the matched filter. There are various characteristics of signals which can be matched for instance magnitude, energy ,frequency etc.

MOTIVATION

Security being the major cause of concern in modern world , speaker verification is a good option because of its reliability and user friendly nature. Comparing to other biometric methods like fingerprint or face recognition, speaker verification systems do not require expensive specialized equipments and are effective especially for remote identity verification.

System Specifications



Working:

4.1 Endpoint Detection

The implementation for the speaker verification system first addresses the issue of finding the endpoints of speech in a waveform. Endpoints are the points on the time axis at which the actual speech signal starts and terminates. The code's algorithm finds the start and end of speech in a given waveform, allowing the speech to be removed and analyzed. It is important to note that this algorithm gives the entire region where speech exists in an input signal. This speech could include voiced regions as well as unvoiced regions. Voiced speech includes hard consonants such as "ggg" and "ddd", while unvoiced speech includes fricatives such as "fff" and "sss". For the short-time magnitude of a speech signal, it is necessary to include all speech which would be located by this algorithm. However, for short-time pitch, one is only concerned with voiced regions of speech. As a result, this algorithm is not used, and instead, we use the energy in the signal to find the voiced and unvoiced regions of the pitch track.

4.2 Dynamic Time Warping (DTW)

"Dynamic Time Warping" (DTW) is a procedure which computes a non-linear mapping of one signal onto another by minimizing the distances between the two signals i.e the input signal and the key signal. Suppose a speaker verification system has a password of "Project". If the user says "Prrroooject" instead of "Project", here the rate of speech has varied during the input. Obviously, a simple linear squeezing of this longer password will not match the key signal because the user slowed down the first syllable while he kept a normal speed for the "ject" syllable. We need a way to non-linearly time-scale the input signal to the key signal so that we can line up appropriate sections of the signals (i.e. so we can compare "Prrrooo" to "Pro" and "ject" to "ject"). The solution to this problem is "Dynamic Time Warping" (DTW).

4.3 Feature Extraction

The final implementation for our speaker verification system includes matched filtering the pitch and magnitude of a signal. By comparing the "short-time" magnitude and "short-time" pitch of two speech signals, we are separating the waveforms into their slowly and quickly varying components, respectively, thus eliminating the problem of aligning the phase of the two input waveforms. "short-time" processing methods => short segments of the speech signal are "isolated" and "processed" as if they were short segments from a "sustained" sound with fixed (non-time-varying) properties.

4.3.1 Short-Time Magnitude

One portion of the final implementation is the comparison of the short-time magnitude of two speech signals. Using an endpoint detection process, the speech is selected from the two input waveforms. Then, their short-time magnitudes are determined.

4.3.2 Short-Time Frequency

A simple model of the human vocal tract is a cylinder with a flap at one end. When air is forced through the tract, the vibration of the flap is periodic. The inverse of this period is known as the fundamental frequency, or pitch. This frequency, combined with the shape of the vocal tract, produces the tone that is your voice. Variations in people's vocal tracts result in different fundamentals even when the same word is said. Therefore, pitch is another characteristic of speech that can be matched. For short-time pitch, one is only concerned with voiced regions of speech. As a result, the end point detection process is not used, and instead, we use the energy in the signal to find the voiced and unvoiced regions of the pitch track.

4.4 Training

When a speaker attempts to verify himself with this system, his incoming signal is compared to that of a "key". This key should be a signal that produces a high correlation for both magnitude and pitch data when the authorized user utters the password, but not in cases where:

- the user says the wrong word (the password is forgotten)
- an intruder says either the password or a wrong word

To develop such a key, the system is trained for recognition of the speaker. In this instance, the speaker first chooses a password, and it is acquired five separate times. The pitch and magnitude information are recorded for each. The signal that matches the other four signals best in both cases is chosen as the key.

4.5 Verification

Once a key has been established, an authorization attempt breaks down into the following steps:

1. The person speaks a password into the microphone.
2. The short-time magnitude of the signal is found and recorded, as is the pitch track.
3. Each is cropped and dynamically time warped so that (possible) corresponding points on the signals are aligned.
4. Now that the signals rest on top of each other, matched filter (for both magnitude and pitch) to determine a numerical value for their correlation.
5. These numbers are compared to the thresholds set when the key was first created. If both the magnitude and pitch correlations are above this threshold, the speaker has been verified.
6. Allow user to enter top-secret hangout.

Theory behind the Application

5.1 Endpoint Detection

The endpoint detection algorithm functions as follows:

1) The process removes any DC offset in the signal. This is a very important step because the zero-crossing rate of the signal is calculated and plays a role in determining where unvoiced sections of speech exist. If the DC offset is not removed, we will be unable to find the zero-crossing rate of noise in order to eliminate it from our signal.

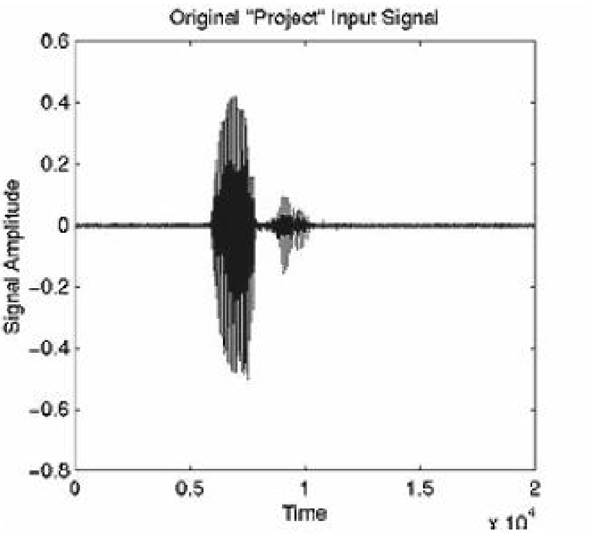
2) Compute the average magnitude and zero-crossing rate of the signal as well as the average magnitude and zero-crossing rate of background noise. The average magnitude and zero-crossing rate of the noise is taken from the first hundred milliseconds of the signal. The means and standard deviations of both the average magnitude and zero-crossing rate of noise are calculated, enabling us to determine thresholds for each to separate the actual speech signal from the background noise.

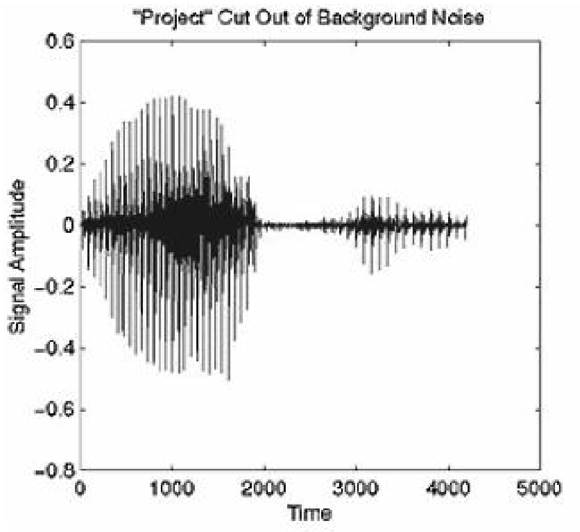
3) At the beginning of the signal, we search for the first point where the signal magnitude exceeds the previously set threshold for the average magnitude. This location marks the beginning of the voiced section of the speech.

4) From this point, search backwards until the magnitude drops below a lower magnitude threshold.

5) From here, we search the previous twenty-five frames of the signal to locate if and when a point exists where the zero-crossing rate drops below the previously set threshold. This point, if it is found, demonstrates that the speech begins with an unvoiced sound and allows the algorithm to return a starting point for the speech, which includes any unvoiced section at the start of the phrase.

6) The above process will be repeated for the end of the speech signal to locate an endpoint for the speech.





5.2 Dynamic Time Warping

The end result of the "short time magnitude" and "short time frequency" is a new, time-varying sequence that serves as a new representation of the speech signal. Suppose in the fig.1 the two signals are the new representation of the input signal and the key signal. As we can see from the fig.1 the appropriate sections on the two signals are not aligned as the second lobe of the input signal is bit delayed compared to the second lobe of the key signal. In such a situation comparison of the two signals will yield poor results. Now if the above two signals are processed through Dynamic Time Warping (DTW), it non-linearly maps one of the two signals onto the other by minimizing the distance between the two signals. Hence the two signals have all their appropriate sections aligned as shown in the fig 2. On comparison of these newly mapped signals will yield far better results than before.

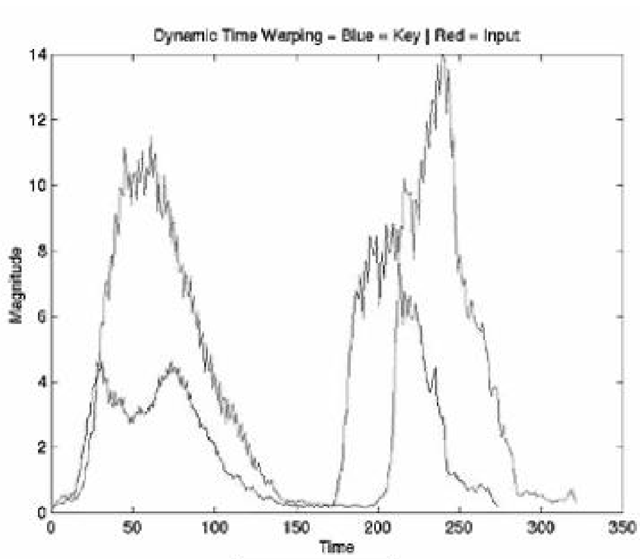


Fig. 1

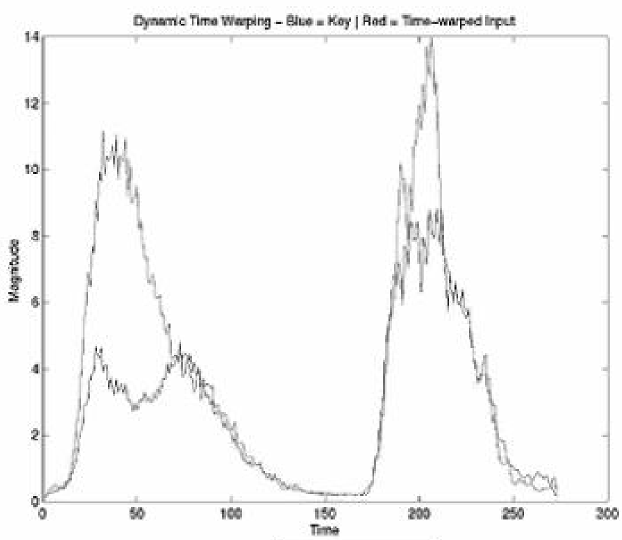
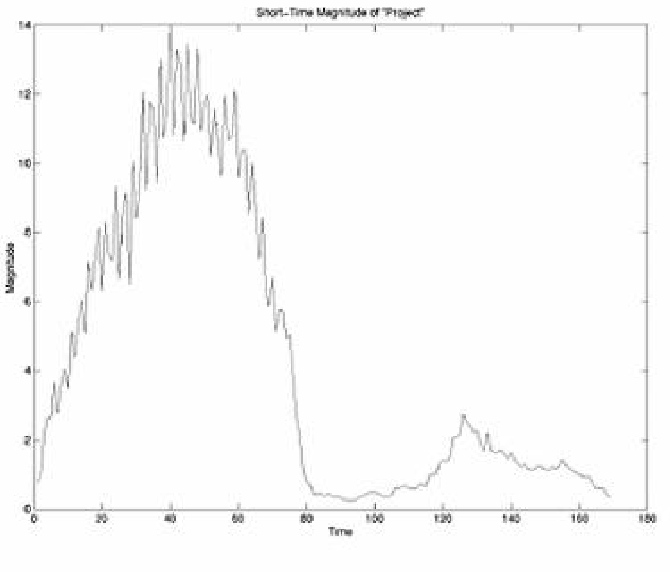


Fig. 2

5.3 Feature Extraction

5.3.1 Short-Time Magnitude

The short-time magnitude characterizes the envelope of a speech signal by lowpass filtering it with a rectangular window. The magnitude function follows these steps: The bounds of the signal are determined and each end is zero-padded. The signal is convolved with a rectangular window. As the window is swept across the signal, the magnitude of the signal contained within it is summed and plotted at the midpoint of the window's location. One magnitude plot is discrete time warped onto the other. The dot product of the two waveforms is computed, and this number is divided by the product of the signals' norms as mandated by the Cauchy- Schwarz Inequality. This calculation results in a percentage of how similar one signal is to the other.

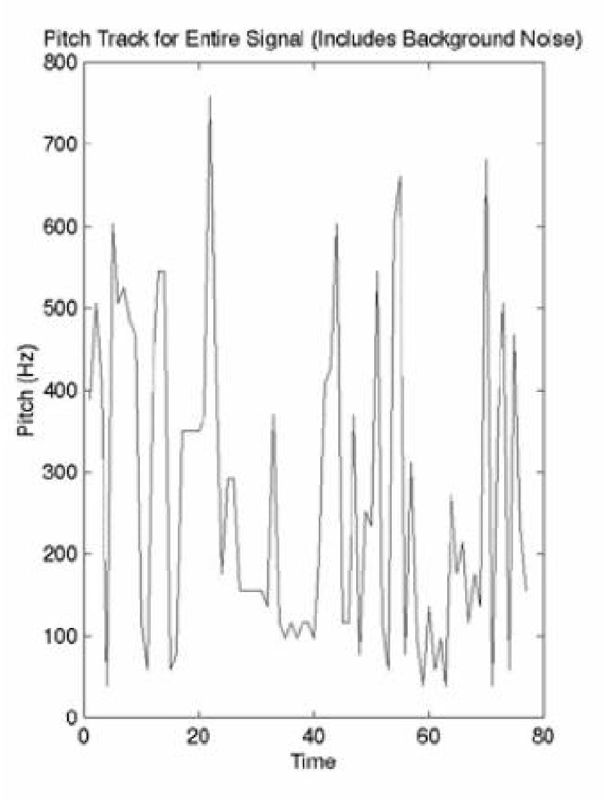


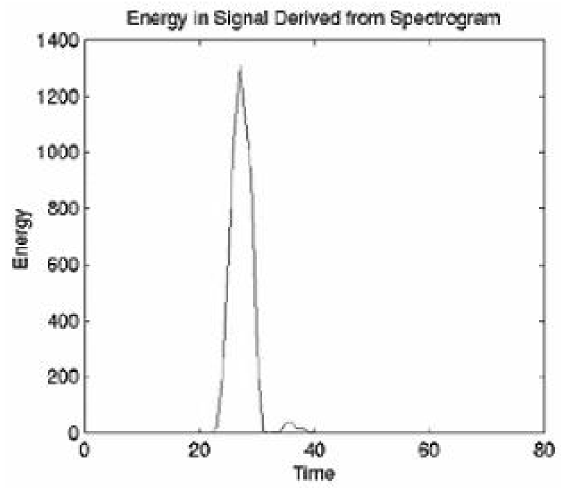
5.3.2 Short-Time Frequency

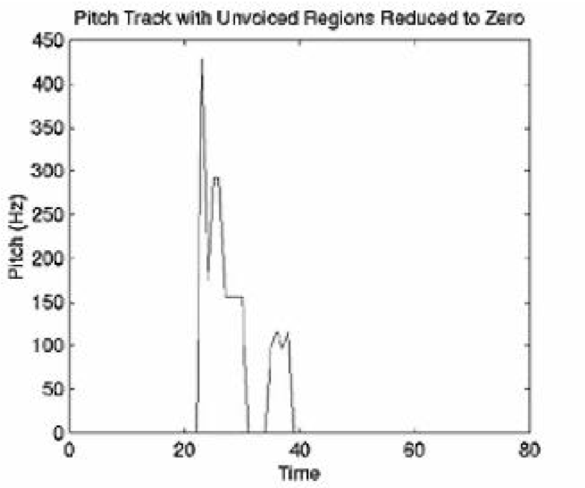
To extract pitch from our signals, we make use of a harmonic-peak-based method. Since harmonic peaks occur at integer multiples of the pitch frequency, we can compare peak frequencies at each time t to locate the fundamental. Our implementation finds the three highest-magnitude peaks for each time. Then we compute the differences between them. Since the peaks should be found at multiples of the fundamental, we know that their differences should represent multiples as well. Thus, the differences should be integer multiples of one another. Using the differences, we can derive our estimate for the fundamental frequency. Derivation:

Let f1 = lowest-frequency high-magnitude peak
Let f2 = middle-frequency high-magnitude peak
Let f3 = highest-frequency high-magnitude peak
Then d1 = f2-f1 and d2 = f3-f2.
Assuming that d2 > d1, let n = d2/d1. Then our estimate for the fundamental frequency is F0 = (d1+d2) / (n+1).

We find the pitch frequency at each time and this gives us our pitch track. A major advantage to this method is that it is very noise-resistive. Even as noise increases, the peak frequencies should still be detectable above the noise. In order to do this,
We first find the signal's spectrogram. We assume that the fundamental frequency (pitch) of any person's voice will be at or below 1200 Hz, so when finding the three largest peaks, we only want to consider sub- 1200 Hz frequencies. Thus, we want to cut out the rest of the spectrogram. Before we do this, however, we must use the whole spectrogram to find the signal's energy. The signal's energy at each time shows the voiced and unvoiced areas of the signal, with voiced areas having the higher energies. Since we are using our pitch track to compare pitch between signals, we want to be certain that we are only comparing the voiced portions, as they are the areas where pitch will be distinct between two different people. A plot of energy versus time can actually be used to window our pitch track, so that we may get only the voiced portions. To find the energy versus time window, we take the absolute magnitude of the spectrogram and then square it. By Parseval's Theorem, we know that adding up the squares of the frequencies at each time gives us the energy of the signal there. Plotting this versus time gives us our window. Once this is done, we cut our spectrogram and move on to finding the three largest peaks at each time. This is done by observing the peaks on the new edited spectrogram. When the three largest peaks at each time are known , the fundamental frequency at each time is calculated using the above derived mathematical expressions. The graph of the pitch track of only the voiced region is obtained from these fundamental frequencies.







System Characteristics

1. Memory:- The system has memory as the output depends on the current input as well on the input give at the time of training.

2. Invertibility:- The system is non-invertible as there are only two outputs accept or deny for all the possible inputs.

3. Causality:- The system is causal as the output depends only on the input signal and the key signal stored during the training process.

4. Stability:- The system is stable as for any given input the output is bounded i.e. either its accept or deny.

5. Time invariance:- The system is time invariant as on keeping the input signal constant, the output is also unchanged irrespective of the time.

6. Linearity:- The system is non-linear as on addition of two input signals does not give an output which is the sum of their respective outputs. Homogeneity also doesn't hold as the output doesn't scale for a scaled input.

USE OF APPLICATION IN REAL LIFE:

a) Speaker verification is used for secure access to services via the telephone, including home shopping, home banking and telecom services.

b) In a forensic context, speaker verification can be deployed in the processing of telephone taps, either to identify a talking suspect, or track the time intervals where a suspect is talking in a lengthy telephone tap . c) Another emerging market is secure access to information which can be obtained through the internet.

Multimodality is a hot issue, here as well as elsewhere: there is a need for a better understanding of the ways in which natural speech can be combined with other media for both input and output.

d) When speech of a visitor is available and security is an issue, speaker verification can be deployed to verify the visitor's identity .

e) One other important application is probation monitoring, where speaker verification is applied to monitor the whereabouts of persons who are not allowed to travel freely, e.g. rehabilitating prisoners.

ACKNOWLEDGMENT:

We would like to thank Prof. V. M. Gadre and the course associates of EE210 for their valuable time and guidance.

REFERENCES

1. ROBERT D. RODMAN : Computer Speech Technology
2. DOUGLAS O' SHAUGHNESSY : Speech Communication- Human and Machines
3. SADAVKI FURUI , MOHAN SONDHI : Advances in Speech Signal Processing
4. 1999 IEEE International Conference on Acoustic, Speech and Signal Processing. Volume I.

PATTERN RECOGNITION IN FOREX RATES USING NEURAL NETWORKS

by

Arijit R. Sarkar

U. Jayakrishnan

Abstract:

This presentation deals with the use of neural networks as a tool for pattern recognition, especially when applied to something as volatile as foreign exchange rates. We look at the nature of Artificial Neural Networks (ANN), and how they can be trained to recognize patterns, and interpolate/extrapolate data streams, more efficiently than conventional regression techniques. Specifically, we look at the kind of algorithms used while training a network, with focus on the Back-Propagation Algorithm. As an application of the pattern recognition capabilities of Artificial Neural Networks, we examine a neural network developed to forecast Foreign Exchange Rates. We look at the various factors affecting the Foreign Exchange Rates, and how these can be fed into a trained neural network to obtain predictions of future Foreign Exchange Rates. Finally, we examine presently available commercial software which use this principle, their drawbacks, and scope for further improvement in this field using wavelets.

Introduction:

Artificial Neural Networks (ANN) are inspired by the high level of information processing capabilities observed in a human brain. Despite technological breakthroughs which have immensely increased the speed of conventional digital processing, as in computers, many tasks can be performed with a much greater speed and efficiency by a human brain. Taking motivation from this, neural networks are attempts to recreate these amazing properties of the brain, albeit on a smaller scale. Tasks such as image processing, pattern recognition and interpolation of data points are increasingly being entrusted to neural networks because of their robust, dynamic and efficient nature.

Neural networks involve a paradigm shift in information processing ideas. They are basically models of the structural nature of a brain. Hence, their units, also called neuronal nodes, model actual neurons.

Neural networks always undergo a training period as such, when input data is fed to it, and adjustments made to improve the correlation between desired and obtained outputs. These adjustments are made to the synaptic weights, which are numbers associated with the interconnections between two neurons. Thus, the entire knowledge, or experience of a network, acquired during training, is stored in these interconnections (synapses). Neural networks, once trained, are assumed to give a minimum average error for the input data which was used to train it. However, trained neural networks have also been observed to be excellent at generalization, i.e. they yield fairly accurate results for interpolated and extrapolated data points. Hence, their outputs are reasonable even when faced with data inputs not previously encountered. This is one of the chief strengths of neural networks, i.e. their ability to handle cases for which they weren't programmed specifically.

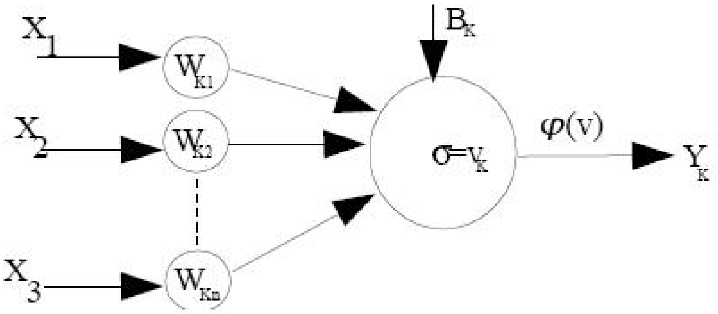
A definition of Artificial Neural Networks which encompasses these details is as follows (ref. S. Haykin):
.An Artificial Neural Network (ANN) is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:
1. Knowledge is acquired by the network from its environment through a learning process.
2. Inter neuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

Some advantages of neural networks are their non-linearity, adaptivity, implementability in both software and hardware, and especially their ease of implementation in VLSI. Coming to Foreign Exchange Rates (forex), we find that it is an area where application of neural networks is particularly apt. To be specific, we will be looking at the exchange rate between Indian Rupees and the U.S. Dollars. The major problem associated with forecasting the exchange rates is to decide exactly which factors affect it. In the present Indian context, this is an extremely tough decision. We can try and quantify some of these factors, such as interest rates, inflation rates, GDP, forex reserves, oil prices, etc. However, some of these cannot be quantified, such as fiscal and monetary policies, political scenarios, etc.

Making several assumptions about the constancy of most of these factors, we can arrive at a certain bouquet of critical variables. However, determining the output (tomorrow's forex rate) as a function of the current value of these variables is also a mammoth task. This is where neural networks enter the picture. With regards to accuracy of prediction, neural networks generally function much better than conventional regression techniques and numerical methods. This is chiefly due to the massively parallel nature of a neural network, with (typically) several layers of interconnections being used. As the number of synapses increases, the number of free parameters to be adjusted (synaptic weights) also increases. This yields better predictions and reduces errors. Now, let us look at the mathematical modeling of neural networks, and the kind of algorithms used to train these networks.

A Neuron

The fundamental structural component of any neural network is a neuron. It is basically a model of the natural neuron present in the human nervous system. It can be mathematically described by a relation between an output and a set of inputs. The diagram depicting a neuron is shown below.



This means that theoretically, we can look at each neuron as composed of three logical units:
1. The synapses, or interconnections, each with a number wkj associated with it. This number multiplies the input coming to the neuron along this synapse. It is called the synaptic strength.
Note: The nomenclature is as follows: k refers to the k'thneuron, and j to the j'thinput from the previous layer. i.e. input along the synapse between the j'th input and the neuron is wkj xj
2. The linear combiner, or adder, which adds all the weighted inputs, along with a bias bk.

$$u_k = \sum_{j=1}^{m} w_{kj} x_j$$

$$v_k = u_k + b_k$$

Sometimes bk can be looked upon as wk0, associated with another synapse at which a constant input x_0=1 is applied. Thus, the previous equation becomes:

$$v_k = \sum_{j=0}^{m} w_{kj} x_j$$

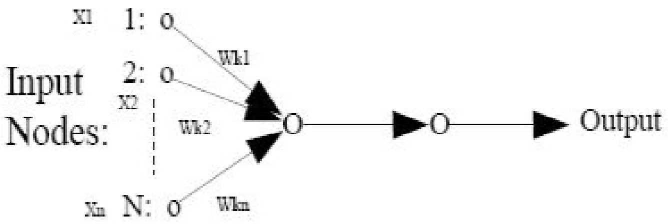
3. The activator / squasher (associated with an activating/squashing function φ) is the final stage of a neuron.

$$y_k = \varphi(v_k)$$

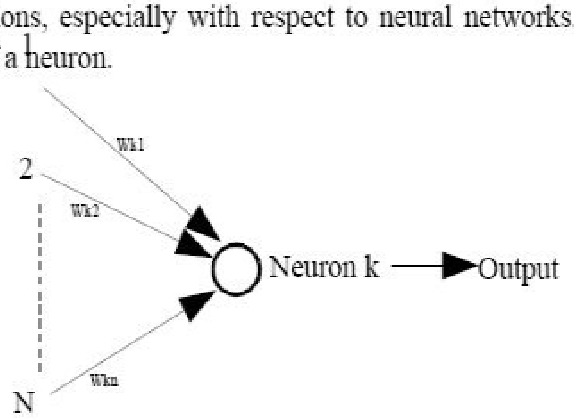
Here φ can have various forms, and various such squashing functions can be used, depending on the ease of computation and accuracy desired.
eg. For the McCulloch – Pitts model, φ is the Heaviside step function,

$$\varphi(v)=0 \quad \{v < 0\}$$
$$=1 \quad \{v > 0\}.$$

For brevity of representation, the following diagram can be said to represent a neuron completely (Also called the signal-flow graph of a neuron).



In fact, we can further ignore the intricacies inside a neuron, if we' re only interested in inter-neuron connections, especially with respect to neural networks. This gives us the architectural graph of a neuron.

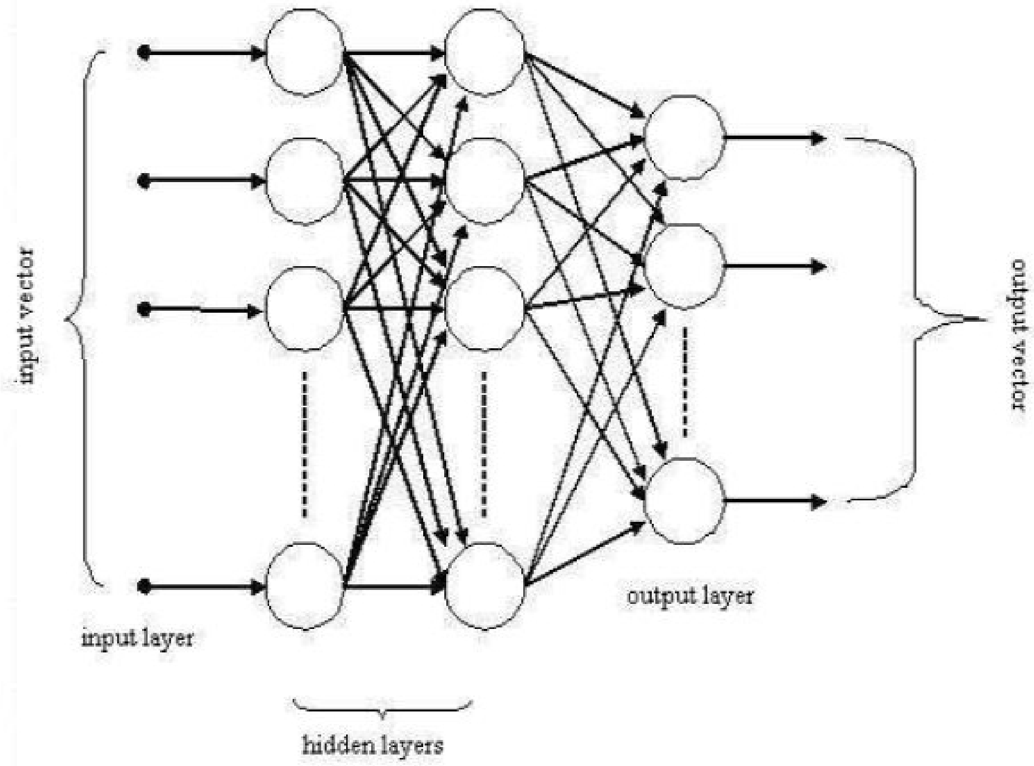


Neural networks are basically collections of neurons. We' ll chiefly be looking at networks where the neurons are arranged in layers, with the outputs of neurons in one layer being used as input for the next. Thus, all layers would be hidden to external measurements except for the last layer.

Now, we will look at an example of such a neural network, specifically the one we use in our application.

The Neural Network Used

For our pattern recognition task, we shall consider a multilayer, feedforward (i.e. with no feedback loops) neural network, as depicted in the figure below. It has an input layer (where the external input vector is applied to the network), an output layer (of neurons, whose outputs constitute the output of the network) and intermediate hidden layers of neurons. We will consider a fully connected network, in which the output of each neuron in a hidden layer is applied to every neuron in the next layer. Such a network is also called a Multilayer Perceptron.



Multilayer perceptron with 2 hidden layers

It is, at this stage beyond us to justify why such a structure of a neural network is employed for our pattern recognition task. All we can say is, such networks are known to do a good job, with suitable choice of synaptic weights and biasing parameters, and that algorithms for evaluating these parameters optimally are known. We shall discuss in detail one such training algorithm a little later, however, let us first make clear what these algorithms do for us. We start off with a set of input vectors, and the corresponding responses we expect from the system. For instance, with forex forecasting, if we have determined that factors a, b, c and d determine tomorrow'sforex rate, we start off with a set (typically large) of previously observed values of a, b, c and d and the correspondingly observed forex rate of the next day. This constitutes what is called the training data. The learning algorithms adjust the free parameters of the network so that the responses of the network to the input samples in the training data are close to the corresponding desired responses.

Now, once the training is done, what happens when the network is presented with an input vector not encountered before during training? It is observed that under certain conditions, the response of the neural network is actually close to the response we would expect for that input. This property of the trained neural network is called generalization. It is this generalization we rely on when we input, say today'svalues of the factors a, b, c, and d to get a forecast of tomorrow'sforex rate. Neural networks are able to pick up very complex dependencies of the dependent variable(s) on the independent variables (that constitute the components of the input vector), and hence do a better job at function extrapolation / interpolation than conventional regression techniques.

A natural thought arises: surely this kind of generalization must demand certain conditions on the function involved and the training data used. The answer is yes, and some of the intuitive conditions are:
1. The components of the input vector must be chosen, so that a dependence does in fact exist, between these and the dependent variable. You obviously will not expect the neural network to generalize well with incomplete input dependencies. For example, to say the forex rate depends only on inflation rate and price index in the two countries is of course grossly crude. Hence, if you train a network with only these as input fields, you can't expect the network top redict forex rates accurately.

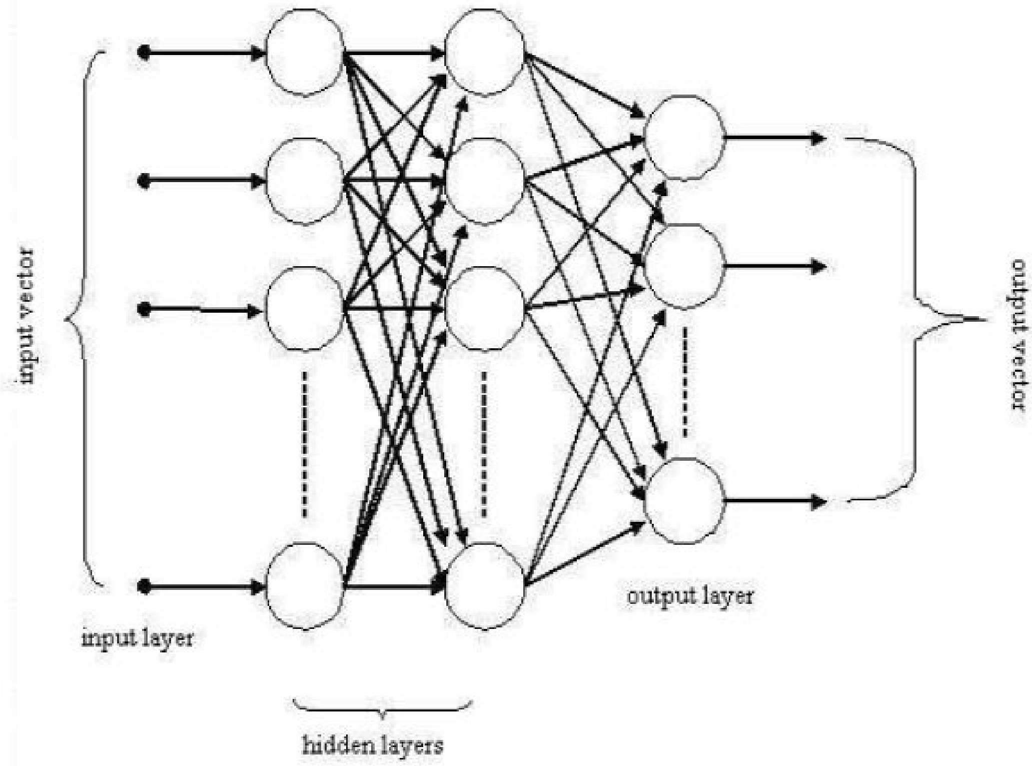
2. The data samples must be large in number and well spread over the input space.

3. The function involved must be smooth, i.e. a small change in one of the factors should cause a small change in the dependent variable. An arbitrary function obviously can't be estimated from its samples!

Now, we move on to a study of the Back Propagation Algorithm, one of the algorithms used to train a multilayer feedforward neural network.

Back Propagation Algorithm

To start off, the parameters of the neural network are arbitrary. We have with us a set of sample input vector-desired output pairs. At first, these examples are fed to the network one after the other and in each case the network adjusts its parameters as explained below. After all pairs are presented once, they are presented again in a shuffled order, then yet again on another order.......till the performance is found satisfactory(this will be elaborated on later) . Below, we explain what is done for each sample presented to the network.



Multilayer perceptron with 2 hidden layers

Say some input sample is fed to the network. Let the output of the k'thneuron in the output layer be yk and let the desired response of this k'thneuron be dk . We define the error ek to be:
ek = dk - yk

We define the error energy for neuron k as ½ e_k^2 , and the total energy E as:

$$E = \frac{1}{2} \sum_k e_k^2$$

where k runs over all neurons in the output layer. Clearly E is a function of all the free parameters of the network. Our objective in this iteration is to decrease E by changing appropriately the free parameters. We do this by shifting the parameters in the direction opposite to the gradient of E (with respect to these parameters). Note the gradient points in the direction of steepest descent of the function and giving the independent variable a small shift opposite to the gradient will decrease the value of the function. Thus our objective is to find $\partial E/\partial w_{kj}$, where w_{kj} is a synaptic weight of neuron k for that synapse emerging from neuron j in the previous layer. We have now to evaluate $\partial E/\partial w_{kj}$ for all synaptic weights of the network.

1. If k is in the output layer, this is easy:
   By chain rule,

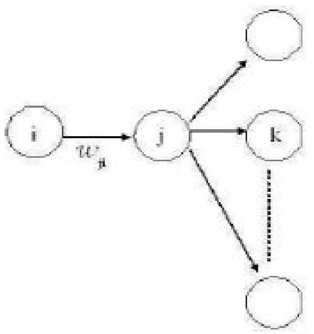
   $\partial E/\partial w_{kj} = (\partial E/\partial e_k)\ (\partial e_k/\partial y_k)\ (\partial y_k/\partial v_k)\ (\partial v_k/\partial w_{kj})$
   note v_k is the induced local field of neuron k.

   where:

$\partial E/\partial e_k = e_k$;
$\partial e_k/\partial y_k = -1$; because $e_k = d_k - y_k$
$\partial y_k/\partial v_k = \Phi_k'(v_k)$; $\Phi_k(.)$ is the activation function of neuron k
$\partial v_k/\partial w_{kj} = y_j$

Thus,
$\partial E/\partial w_{kj} = -e_k\ \Phi'(v_k)\ y_j$
$\partial E/\partial w_{kj} = -\partial_k\ y_j$ where $\partial_k = \partial E/\partial v_k$

2. We can find $\partial E/\partial w_{ji}$ where j belongs to a hidden layer as follows :



By chain rule,

$$\partial E/\partial w_{ji} = \sum_k (\partial E/\partial v_k)(\partial v_k/\partial y_j)(\partial y_j/\partial w_{ji})$$

where the summation is over all neurons of the next layer !
Therefore,

$$\partial E/\partial w_{ji} = \sum_k \partial_k w_{kj} (\partial y_j/\partial w_{ji})$$

i.e. $$\partial E/\partial w_{ji} = (\partial y_j/\partial w_{ji}) \sum_k \partial_k w_{kj}$$

where $\partial y_j/\partial w_{ji} = \Phi_j'(v_j)\, y_i$

What this means is: we can first find partial derivatives for all synapses that are inputs to the output layer. This has been illustrated in step 1. In the process we also find ∂_k where k is any neuron in the output layer. Using these ∂_k ' ,swe can find the partial derivatives $\partial E/\partial w_{ji}$ where j is any neuron in the previous layer. Using these ∂_j ' ,swe can find the partial derivatives for the layer previous to that......
That is why the algorithm is called the Back-Propagation algorithm. The partial derivatives of E with respect to all synaptic weights are evaluated layer by layer, backwards, starting from those incident on the output layer.

Corrections are made to the synaptic weights as follows:
$\Delta w_{kj} = -\,\eta\, \partial E/\partial w_{kj}$

where η is called the learning parameter. It is typically a small positive real number. Too small a value makes learning too slow and too large a value could introduce make the network unstable.

This procedure is repeated over all samples, then all over again and again by shuffling the order of the samples. The algorithm may be terminated when the average value of E over a cycle is found to be small enough. The Back-Propagation algorithm cannot be theoretically shown to converge and its utility comes predominantly from the fact that it works. Correspondingly there are many termination criteria as well, each with its own practical merits.

Forex Forecasting

The Backward Propagation Algorithm is presently used by commercial software in the training of neural networks for prediction of forex rates. An example of such a software is NeuNetPro. Though various factors influence forex rates, a statistical report (Waghela – Mallya) obtained a reasonable amount of accuracy in forecasting using only the following factors:
1. Call money rates
2. 91 Day – Treasury bill yield
3. WPI
4. GM LIBOR
5. Reserves
6. US Inflation

Conclusion:

Due to the ability of neural networks to generalize very well, they can be employed wherever something varies as a very complicated function of a number of variables, and where it is difficult to define the function from the data we have. For instance neural networks may be used by direct mail advertisers to decide out of a database which customers to target. More complex uses include handwriting recognition, fingerprint matching, facial recognition and so on.

Bibliography

1. Neural Networks, A comprehensive foundation – Simon Haykin
2. Forex Forecasting – Prashant Waghela and Shailesh Mallya
3. http://www.cormactech.com/neunet/
4. comp.ai.neural-nets FAQ
5. A case study on using neural networks to perform technical forecasting of forex, Neurocomputing, Volume 34, Issues 1-4, September 2000, Pages 79-98 Jingtao Yao and Chew Lim Tan

Use of Signals in Seti

by

Jagabathula Srikanth(02007005)

Salim Dewani (02007004)

Abstract:

The Search for Extra-terrestrial Intelligence (SETI) is an ongoing project directed towards determining the existence of intelligent life beyond our planet. This is done through the detection and analysis of signals from outer space. This document describes the application of signal processing in SETI. The following topics are covered:
1. Origin and brief history of SETI.
2. Nature of signals sought by SETI.
3. Detection and Analysis of these signals.
4. Optical SETI project and its comparative advantages.
5. Current status of research and exciting future prospects of SETI.

Motivation:

The ideas of galactic civilizations and alien landscapes have been elegantly built by science fiction writers over the years. The SETI project seeks to investigate these ideas scientifically. The use of simple concepts in signal analysis such as the discrete Fourier transform and the Doppler shift in the search for extra-terrestrial intelligence is a fascinating realworld application of the concepts of the EE210 course.

So what is SS technology?

The SETI project can be traced back to 1959 when Cornell physicists Giuseppi Cocconi and Philip Morrison published their findings regarding the potential for using microwave radio to communicate between the stars. Another astronomer, Frank Drake, in the spring of 1960 began a series of experiments which turned out to be the first microwave radio search for signals from other solar systems. He used a single-channel receiver tuned to the magic frequency of the 21 cm (1,420 MHz) line of neutral hydrogen. This came to be known as Project Ozma and it subsequently captured the interest of astronomers worldwide in spite of failing to detect any indicative signals. Thus began an era of Soviet dominated SETI that used antennas to observe large areas of the sky, assuming that some very advanced civilizations might be capable of radiating enormous amounts of transmitter power.

In the early 1970's, NASA conducted a comprehensive study for known as Project Cyclops. The Cyclops report provided an analysis of SETI science and technology issues that is the foundation upon which much subsequent work is based. Radio astronomers now conducted searches employing steadily improving technology. Amongst these are the Planetary Society's Project META, the University of California's SERENDIP project, and a longstanding observing program at Ohio State University. In 1989 NASA withdrew its involvement and SETI continued its large-scale programs with private funding.

Why do we believe in Extra terrestrial Life:

Drake's Equation:
Frank Drake came up with an equation to estimate the number of communicating civilizations in our Milky Way. The equation is given below: $N = R * f(p) * n(e) * f(l) * f(i) * f(c) * L$

Here:
1. N = the number of communicating civilizations in our Milky Way galaxy.
2. R = the rate of "suitable" star formation in the galaxy.
3. f(p) = the fraction of stars that have planets.

4. n(e) = the number of these planets around any star within the suitable ecosphere of the star.
5. f(l) = the fraction of suitable planets on which life actually appears.
6. f(i) = the fraction of life bearing planets on which intelligent life emerges.
7. f(c) = the fraction of those planets where intelligent life develops a technology and attempts communication.
8. L = the length of time that an intelligent, communicating civilization lasts.

(An "ecosphere" is a shell surrounding a star within which the conditions suitable for life are observed to exist.) Let us estimate the value of N and see if it is worthwhile to search for extra-terrestrial intelligence. The rate of star formation is about 20 stars per year (R=20). We can safely say that one-half of the stars form planetary systems and the other half form binary star systems (f(p) = 0.5). Since small stars are at lower temperatures, planets capable of life have to revolve very close to the star. Such planets may be locked in position such that one side always faces the star leaving the other side very cold. Also huge stars have a short life and so they have no chance to promote life on their revolving planets. An average sized star like the sun has 2 planets that support life. Lets take n(e)=1. Also we can conservatively estimate f(l)=0.2. Assuming that the development of intelligent life is a part of evolution we can say that f(i) = 1. We see human beings making attempts to communicate with extra-terrestrial intelligence while other moderately intelligent species on earth don't do so. We could once again estimate f(c) to be equal to 0.5.

Substituting the above estimates in the Drake equation yields: N=L
It is indeed interesting to note that the number of intelligent communicating civilizations in the galaxy equals the number of years that such a civilization lasts. Using the fact that humans are successfully using technology over the last 50 years, we may say that L might at least be 50. We thus see that scientific reasoning tells us that there is indeed a strong possibility of the existence of extra-terrestrial life trying to communicate with us.

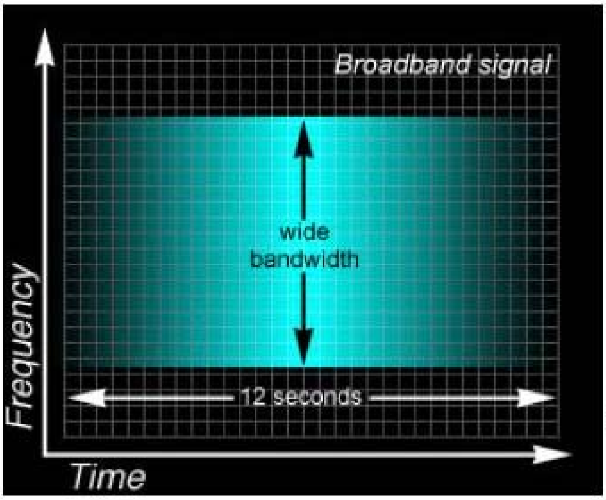
WHAT KIND OF SIGNALS DO WE LOOK FOR?

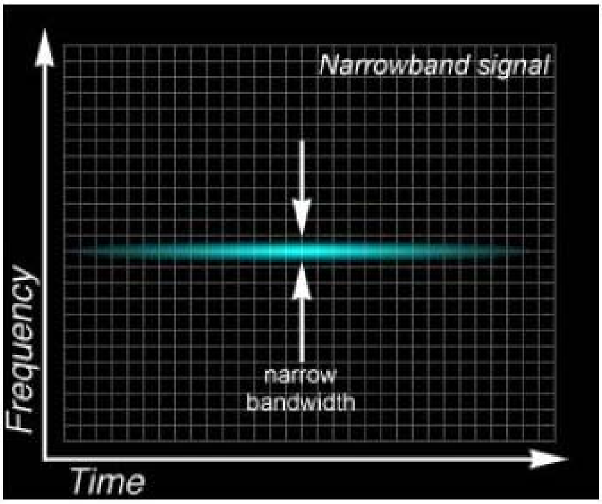
The best communication method known to us is through radio waves. Radio waves have the information carrying capacity and they can be transmitted using equipment that is cheap to build. The information also travels at the speed of light. Moreover we can receive signals from various directions and communicate simultaneously with many different civilizations. The electromagnetic spectrum is very large and the signals can be sent at any frequency. So we should find a reasonably small region of it to begin our search.

The galaxy produces noise at low frequencies and the atmosphere at higher frequencies. Between the two noisy regions we have a relatively quiet region -- from about 1 GHz to about 10 GHz. Neutral hydrogen gas emits radio signals at 1.42 GHz and the hydroxyl emits signals at around 1.64 GHz. We know that the water molecule is composed of these two species. Accepting the fundamental need of water for life to exist, we find ourselves with a frequency range between these two emissions which is a quiet region of the spectrum and is termed "the water hole." This gives us a nice limited frequency range to start our search.

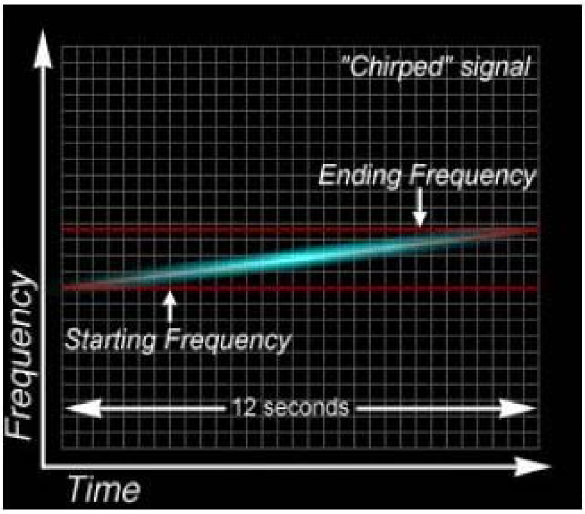
Nature of the spectrum of the signals we look for:

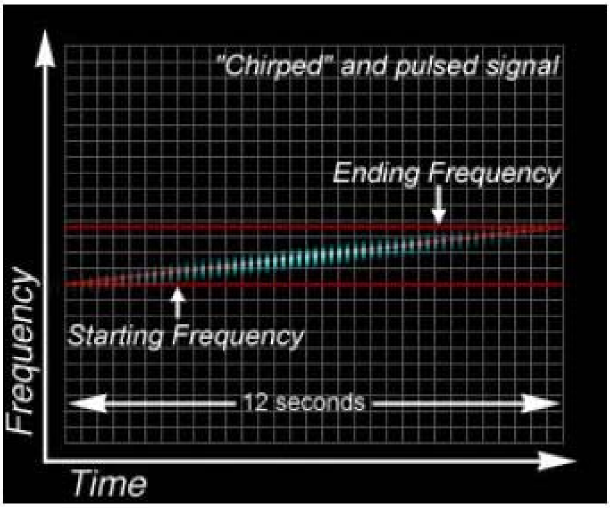
A message transmitted on many frequencies is not efficient as it consumes lots of power. However, if one selects a very narrow frequency bandwidth to concentrate the power of the signal, then the signal can be transmitted easier through the background noise. Considering the distances over which these signals must travel before they reach our planet, we expect intelligent extra-terrestrials to send a very specific frequency message.





To distinguish these signals from local earth-based signals, we recognize the fact that local signals maintain more or less constant intensities with time. However in case of a SETI telescope, it is the sky that is seen to drift past the focus of the telescope. In 12 seconds a target typically crosses the focus (target beam) of the telescope. Thus an extraterrestrial signal is expected to grow in intensity and then diminish over the 12 second time period. This shape is described by the Gaussian curve.





Due to the relative motion of planets, we are likely to observe a Doppler shifting or frequency shifting of the signal due to relative motion of the telescope during the 12 second period. This would cause the signal to rise or fall in frequency slightly over the time length of 12 seconds. These signals are called chirped signals. Also if alien civilizations are sending information our way then the signals are surely going to be pulsed in nature. Thus we may refine our search by checking for chirped signals containing pulses.

DETECTION AND ANALYSIS OF SIGNALS

DETECTION:

The signals that we seek may be of two kinds. Either they are intentionally sent out to us by civilizations in an attempt to contact us, or they simply be leaking out of their planets unintentionally.

Project Phoenix:

Project Phoenix is the world's most sensitive and comprehensive search for extraterrestrial intelligence. It scans for both kind of signals, ie those deliberately beamed our way, and those that are inadvertently transmitted from another planet. Phoenix began observations in February, 1995 using the Parkes 210 foot radio telescope in New South Wales, Australia. This is the largest radio telescope in the Southern Hemisphere. Phoenix scans the skies around stars that resemble the sun in their size and composition. We expect the probability of finding life around such stars to be comparatively higher. Thus stars surrounded by planets are closely observed in preference to binary star systems. There are about one thousand stars targeted for observation by Project Phoenix. All are within 200 light-years distance.

Arecibo Radio Telescope:

It is the largest radio telescope in the world and the is used in the Berkeley SETI search. The large dish reflects and concentrates the weak celestial signals on the receiving antennas hung above. The Observatory operates on a continuous basis, 24 hours a day every day, providing observing time, electronic, computer, travel and logistic support to visiting scientists. The Observatory is located on the Caribbean island of Puerto Rico.

Arecibo Observatory
(courtesy of the NAIC - Arecibo Observatory, a facility of the NSF, David Parker/ Science Photo Library)

ANALYSIS OF SIGNALS:

The output received from the telescopes is digitized and converted into a baseband signal using various filtering and mixing devices. This digitized signal is subjected to further analysis as described below.

De-Chirping:

The relative motion of planets due rotations and revolution introduces Doppler shifts into the signals received. These chirped signals are received by the telescope. Since we are unaware of the rest frame of the transmitting source in the laboratory's topocentric reference frame, softwares are used to examine thousands of different Doppler acceleration frames of rest (dubbed chirp rates), ranging from -10 Hz/sec to +10 Hz/sec.

De-chirping the data is accomplished by multiplying the time domain data by the complex vector V:
V = e(-j*c*t2)
where:
t = time
c = chirp rate (ranges from -10 Hz/sec to +10 Hz/sec)

The resulting de-chirped signal is then further analyzed using the discrete flourie r transform(DFT).

Discrete Fourier Transform:

The output signal from the telescope is digitized and we have the sampled signal. We need to analyze the signal in the frequency domain in order to distinguish the signal from noise. In the frequency domain, the analysis is carried out by sampling the signal at various frequency resolutions (bandwidths). The sampling of the signal in the frequency domain is done by taking the Discrete Fourier Transform(DFT).

The Discrete Fourier Transform of a discrete periodic signal, x[n] ,with a period N is defined by:

$$x[n] = \sum_{k=0}^{N-1} (f[k] \times e^{\frac{-j2\Pi nk}{N}})$$

Let x(t) be a signal which is non-zero for t between 0 and T. Let X(f) be its Fourier Transform. Let x[n] be the signal obtained by taking N samples of the signal x(t). We now argue that what we obtain on sampling X(f) with a sampling frequency of 1/T is approximately the DFT of the periodic extension of x[n] with a period N.

$$X(\frac{n}{T}) = \int_{-\infty}^{+\infty} x(t)e^{\frac{-j2\Pi nt}{T}} dt$$

But x(t) is non-zero only between 0 and T, hence

$$X(\frac{n}{T}) = \int_{0}^{T} x(t)e^{\frac{-j2\Pi nt}{T}} dt$$

The sample points are t_k=kT/N.

Now approximating the signal with the left end point Riemann summation we get:

$$X(\frac{n}{T}) = \frac{T}{N}\sum_{0}^{N-1} x(t_k)e^{\frac{-j2\Pi kn}{T}}$$

Thus we see that X(n/T)=T/N *X[n]
Hence we use DFT to analyze the signal in the frequency domain at different frequency resolutions.

In practice,however we use software to determine the DFT of a signal. It requires a lot of calculations to determine the DFT of a signal. In order to increase the efficiency we determine the Fast Fourier Transform(FFT) instead of DFT. It serves the same purpose as that of the DFT and can also determined with fewer calculations. The FFT is carried out on the sample at different frequency resolutions. The finest frequency resolution is 0.07Hz. This means that T=1/0.075 =13.375 seconds. The output from the telescope is a 107 second data. Hence it is divided in to 8 chunks of 13.375 seconds each. Now FFT is carried out on each of the 8 chunks of data. Then the FFT is carried out at 0.15 Hz. For this resolution T=1/0.15=6.7 seconds. Thus the data is divided into 16 chunks each of 6.7 seconds and 16 FFTs are determined. This process is carried out for 15 frequency resolutions fn=0.075 * 2n Hz, n = 0, ... , 14 and FFTs are determined in each of these cases. Finally the signals that show a strong power which increases and then decreases over the 12 second period can be tentatively considered extraterrestrial in nature.

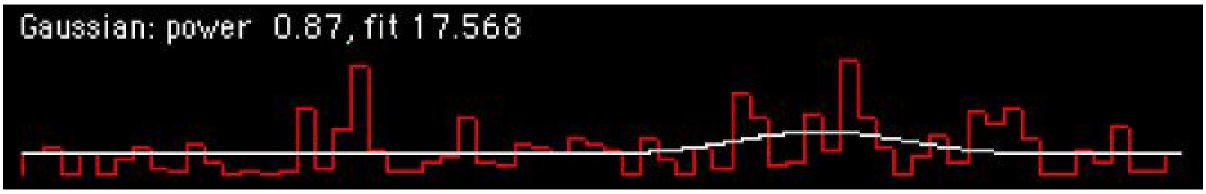
For each of the above frequency resolutions and chirp rates we analyze the signal for the following four types of signals:

1. Spikes: Spikes are radio waves occurring at single frequencies that are strong enough to be distinguished from general noise. The data is examined for signals whose power exceeds 22 times the mean noise power.

2. Gaussian: Gaussians are signals whose power goes up an down as the telescope moves across the sky. As a radio source drifts through the field of view of the telescope, the measured power will vary depending on the telescope's beam profile, which is approximately Gaussian. A beam fitting algorithm is used to fit a Gaussian curve at each time and frequency of the form:

$$\mathbf{P = B + A\, e^{-((t-t_0)/b)2)}}$$

where
P = predicted power
B = baseline power
A = peak power
t = time
t0 = time of Gaussian peak
b = half power beamwidth
The parameters B, A, and t0 are unknown, but the beamwidth is known. It can be calculated from the slew rate of the telescope beam.



3. Triplet: We assumed that the aliens would send pulsed signals in an attempt to contact us. So we analyze the data for the presence of pulsed signals. A triplet is a type of pulsed signal. In order to determine the presence of triplet signals we look for pulses above a certain threshold value for every frequency slice in the spectrum. A reasonable value of the threshold frequency is chosen so that a reasonable number of pulses are revealed, yet we are not overwhelmed by the noise which leads to a number of useless calculations. For every pair of pulses above the threshold, the we look for a pulse exactly in between the two. The algorithm used for doing this avoids any sort of repetition in trying all pairs. This process is repeated for every frequency slice.

4. Pulse: Pulses are signals which seem to repeat themselves at constant intervals. An algorithm called fast folding algorithm is used to determine the pulses. This analysis is carried out in every frequency slice of the 10Hz data. We try to detect many small repeating pulses in the data. Quite often these small pulses so weak that get lost in the noise and are undetectable. We select a frequency slice from the data and examine the power vs time data for the pulses. The data is divided into uniform sized time chunks and these chunks are added together. If the size of the time chunk is the same as (or a multiple of) the period of the pulses, all the pulses will add one on top of the other and we will see the pulses grow out of the noise. As we have no idea about the frequency of the pulses, we have to try all the time periods. Again, the algorithm does this in such a way that it will not repeat work already done.

Optical Seti

If we use a high-intensity pulsed laser coupled to a moderate sized transmitting telescope, we create an efficient signaling device. Such a laser transmitter with its slender beam would appear a thousand times brighter than our sun in broadband visible light to a distant observer. Even at distances up to a 1000 light years away, a single nanosecond laser pulse would transmit a thousand photons to a 10-meter receiving telescope. Using research grade equipment, detecting signals from distances up to 1000 light-years are feasible and this is the motivation behind the Optical SETI project which scans the skies for pulses of laser light in or near the visible portion of the light spectrum.

Merits of Optical SETI over Radio SETI:

1. The higher frequency of light waves allows lasers to produce carrier signals at much higher frequencies allowing extremely fast data transfer.

2. Visible light is typically not affected by interference as compared to microwaves.

3. Nanosecond pulses of light of natural origin are not known to exist.

4. Dispersion, which spectrally broadens radio pulses, is completely negligible at optical frequencies.

5. Intensive computer analysis required in sensitive microwave searches today is not needed for optical SETI.

The analysis of microwaves suffers from false triggers created by spark plugs, etc. However brief spikes or pulses of visible light due to natural origin are rare and the detection of a laser pulse signifies a higher probability of the signal originating from intelligent life and not from a natural astronomical event. In fact, nanosecond spikes of light are thought to be nonexistent in the universe.

Procedure of detection:

We are looking for pulses of visible light, several nanoseconds in duration. The amplitude of the pulse must significantly exceed the strength of illumination from the actual star. Each star is scanned for a certain duration of time which should exceed the time period of the expected pulses. Also we do not seek any specific frequency of light since even the lasers made out of current day technology, if sampled at a nanosecond pulse without filtering, still outshine the parent star by 30 times.

False Triggers: Any strong flash of light in nature could be a potential source for a false trigger. However every such known source can be compensated for as elaborated below:

Cerenkov Flashes: A Cerenkov light flash is emitted when objects exceed the speed of light in media other than vacuum, such as air or water. These particles subsequently emit Cerenkov radiation, which is visible as a flash of blue light. However, these flashes release only a single photon in a nanosecond which is insignificant for detection purposes and thus do not affect OSETI readings. Lightning too does not affect readings as it cannot generate pulses in the nanosecond range.

Electrical arcs across high voltage terminals can trigger detectors. To compensate for this, we use two detectors, and reject readings that are not common to both simultaneously. This gets rid of the problems created by electrical arcing in the instruments. Cosmic rays entering the atmosphere can cause discharge of high energy photons that could trigger the OSETI detection setup. To prevent this, data from various adjacent detectors is compared.

CURRENT STATUS OF RESEARCH AND FUTURE PROSPECTS

Clearly the SETI project has no obvious hunting ground or no clear route to discovery. In 40 years of data collection, actual experimental evidence has been missing. We still have not detected any undebatable signals from the extra terrestrial intelligence that we presume is somewhere out there. With 400 billion stars in the Milky Way itself, the search is analogous to seeking a needle in a haystack. Also the astronomical distances that signals need to cover before they reach earth causes them to diminish in intensity.

At present Project Phoenix uses the Arecibo Telescope to examine nearby, Sun-like stars for narrow band microwave signals. Other SETI searches also sweep wide areas of the sky in the hunt for such signals. Over the years of observation we have found evidence that planets are abundant, but we still do not know about the existence of life-supporting conditions on them. Also over the decades we have developed powerful lasers and supercomputers.

Today research involves studying both radio and optical signals for extraterrestrial signals. Recent developments include an improved detector for targeted optical SETI that reduces the number of false alarms from typically once a night to once a year! Radio SETI searches too benefit from the technology leaps. The ideal SETI radio telescope would monitor every point on the sky, in every radio channel from one end of the microwave window of the earth to the other all the time. This projects called the Omnidirectional Search System, or OSS.

The Moore's Law in computing predicts a thousandfold improvement in computational power in 15 years. If this happens then OSS can be a reality at the end of that time period. Presently the most sensitive project is Phoenix. But the Phoenix is not sensitive enough to even pick up leakage satellite and television signals from the earth at a distance equal to that of the Alpha Centaur (the nearest star to earth.) For sheer sensitivity, SETI's dream instrument is the proposed Square Kilometer Array, or SKA. This grand radio telescopes would be a mammoth device with a total collecting area of one square kilometer. No one can predict the ultimate outcome the search. However the powerful next generation instruments will surely raise the probability of a finding if such extra-terrestrial intelligence actually exists.

ACKNOWLEDGEMENTS

We would like to thank Prof. V. M. Gadre for the wonderful opportunity he provided us to explore this fascinating subject through the assignment.

REFERENCES

The SETI Institute.
SETI@Home Project
Princeton OSETI Center
SERENDIP Project
Sky and Telescope Magazine
Discrete Fourier Transform – David Gilliam (Texas Tech Univ.)

Signals and Systems in Radio Astronomy

by

B. Srikar (02007001)

Chaitanya Ghone (02007028)

Varun Bhalerao (02007006)

Abstract:

We look at radio telescopes in the context of signals and systems and try to analyse their functioning. Instead of mathematical rigor, we resort to an intuitive understanding of the concepts (without detailed proof). We describe in this spirit the properties of a single antenna, as well as a basic two antenna interferometer in terms of standard system descriptions.

Index terms:Radio astronomy, Signals and Systems, Application Poster presentation Example

Introduction:

Let us first take a brief look at radio astronomy. Radio astronomy began in 1894 – Oliver Lodge made the first [unsuccessful] attempts to detect radio emissions from sun. An engineer at bell labs – Karl Jansky – made the first detection of radio waves as “a steady hiss interference in radio transmissions”. This hiss was associated with stars because of the periodicity, and thus began radio astronomy. Now it is a well developed field with different types of antennas and receivers used all over the world. The world’s largest meter-wave radio telescope is in India. Today, these telescopes and detectors are at the cutting edge of technology, far better than those used for commercial purposes.

A radio telescope differs from an optical telescope in the sense that we can’t look “into” a radio telescope and see the image. A radio telescope is basically an antenna, with a receiver that stores the received data. The data is processed to generate some sort of an image. In contrast to an optical telescope which looks in a specific direction, the radio antenna “lobes” are rather spread out, giving lesser directionality in observations. Another factor that comes in (which we shall see in more detail later) is the resolution – it is not feasible to create a single radio telescope that has as much resolution as an optical telescope. So, interferometers are used – which are radio telescopes working together to form an image. All this as we shall see, is closely related to the concepts of a “system” processing input “signals” to give an output “signal”.

Motivation:

We were motivated in our choice of this application due to our basic interests in astronomy. Yet, radio astronomy was a field which we had not explored, and wished to take a glimpse at. Looking into it with this point of view, we realized that it was, in great detail, an application of precisely what was covered in the course material. This drove us further into choosing radio astronomy as a system for analysis..

THE RADIO TELESCOPE AS A SYSTEM

The first system we take for analysis is a simple radio antenna. We examine the antenna in terms of the system properties – looking at them from a radio astronomy perspective. Radio telescopes are different from conventional idea of “Telescope” – it is actually a standard (and sometimes a non-conventional) radio antenna. In this sense, we can also say that the radio telescope dose not directly form an images, as does an optical telescope. When an optical telescope takes an image, it is collecting rays from specific directions in the sky and each pixel stores information only about the light it received – we cannot get any details about the pattern within the pixel.

In case of radio telescopes, however, the beam is not as directional as is the case for optical telescopes. Also, we can say that there is a single “pixel” or detector in most radio telescope antennas. (There are some notable exceptions where multiple detectors are put around the primary focus of parabolic antennas). Before we go into more details, let us formalize the radio telescope as a system under consideration. The input to the system is a function of time and some space coordinates. Typically the space coordinates are the spherical polar coordinates theta and phi. The output of the system is a plot of the variation of the intensity as a function of some linear coordinate, say the x axis. If we are imaging a part of the sky rather than a point source, the output may be plotted as a contour map on the x-y plane, at some given instant of time. Here it is worth noting that since practically we cannot measure the intensity of all the points at once, such maps are plotted only for sources where the intensity is not a function of time, or is constant over the time scales involved in imaging the object.

This system is then seen to have the following properties:

1. Linearity: The system is linear [upto the practical limit of saturation of detectors, this limit is never attained in practice.]

2. Shift invariance: The system is shift invariant in time – i.e. whether we image a steady source now or later, the output recorded is the same. We note that the system as it stands is NOT shift invariant in space [as has been discussed above – the magnitude of response recoded depends on where the object is. More on this is discussed below]. However, if we tracking the source, the coordinate frame is with respect to the principal axis of the telescope. In that case, we see that the system is shift invariant in space also – but this interpretation is open to discussion, or to be more precise, depends on the convention adopted.

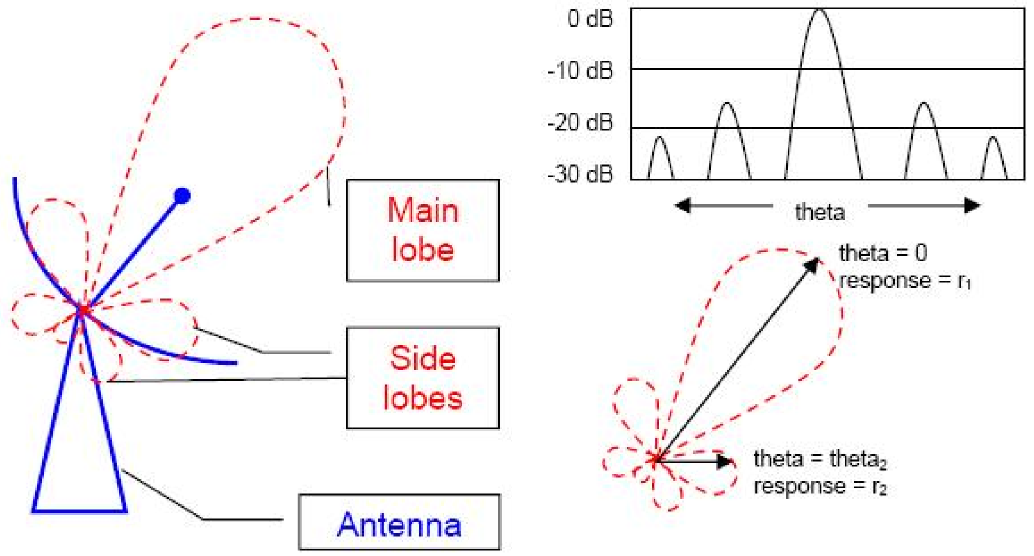
3. Stability: The system is stable. This is rather intuitive, as no practical system can ever give an unbounded output – limiting factors will always take it to a saturation level. We can talk of this in slightly more formal terms by noting that given an input $x(t)$ to the system, the output $y(t)$ can be obtained by $kx(t)$ where k is an appropriate constant with dimensions. Then we can see that if $x(t)$ is bounded, $y(t)$ is also bounded.

4. Causality: The system is causal. Although we will later see that there is a convolution involved, it is not “looking into the future” – it is a convolution in the space coordinates. The output of the system at any given time depends only on the input at that instant. Here we are neglecting the tiny processing delays that may creep in. In that case, the system just adds a constant delay, but still remains causal.

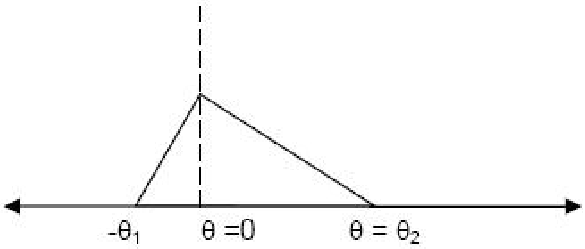
5. Memory: The antenna and detector by themselves are memory-less. They take the input of radio waves, and give the corresponding output. All operations are point-by-point. However, if we record the output on any device which we consider to be a part of the system, it can be said to have memory. Again, this is a limited version of memory, as the output variable is not dependent on time, but is dependent on another quantity (say the x variable) proportional to time.

6. Invertibility: The system is not invertible. Although it is linear and one-one in mapping the input to the output, we cannot expect the telescope to emit radio waves given the graph plotted at the output. However, if we feed in electrical signals to the antennas, they are capable of radiating radio waves. We do not make a general statement here about whether this system with radio waves incident on the antenna and some electrical signal as output, is invertible or not. There are several factors which come into play here before we get the output signal, and the system still need not be invertible.

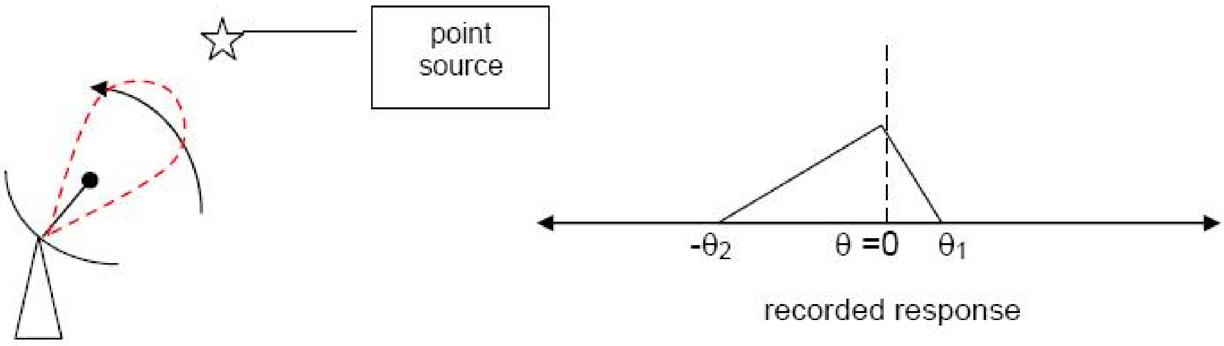
Now, having taken a look at the basic system properties, let us consider an antenna observing something in the sky. Had we been talking of an optical telescope, we would have said that the telescope being pointed in a certain direction, is taking an image of the sky in that direction. The case of a radio telescope, however, is different: the antenna does not have such a narrow beam. The antenna responds to radiation coming from different directions. We can arbitrarily fix the maximum response of the antenna as unity and plot the response in different directions on an r-theta plot. A typical such plot is given below:



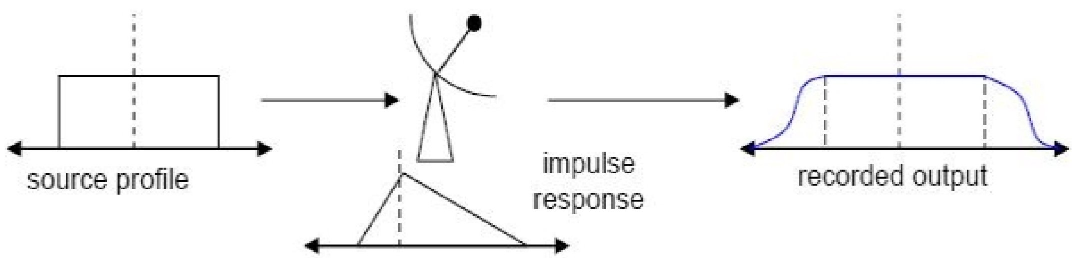
We see that the response plotted in r-theta shows “lobes” – the “r” is the fraction of incident radiation along that direction that the antenna would detect. The direction along which the response is unity is called the principle axis. This is usually normal to the antenna. The coordinate theta (or phi, as the case may be) is measured with respect to this axis. On the right we have plotted the antenna response or the so called “lobe pattern” on an xy graph: the angle theta is plotted on the x axis, while on the y axis we plot the response of the antenna to radiation form that direction, in decibels. The response of the antenna is the total amount of radiation coming from the sky. Here, we note that the antenna can give the same response to several possible sources: as an example, we see that the antenna would give the same output for a source of intensity k/r1 along the principle axis, as for a source of intensity k/r2 along the direction theta2 shown in the figure. Now we take a look at how a radio telescope images the sky. Instead of considering the realistic lobe pattern depicted above, let us take up a simplified lobe pattern for analysis. This pattern is shown below with theta (measured from the principle axis) plotted on the horizontal axis and intensity plotted on the vertical axis.



Now, let us consider the antenna sweeping across a point source – what we see on the graph is as follows:



Notice that we have the recorded impulse as some sort of an impulse response of the system. Since we are talking of an LSI system, this impulse response characterizes the system completely. So, if we have an extended source in the sky, and the antenna sweeps over it, what we record is the convolution of the source with the antenna lobe pattern. This is illustrated in the figure below:

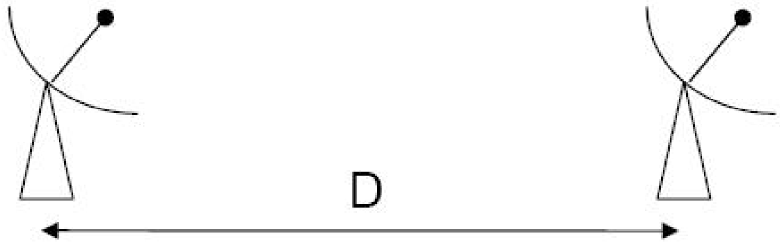


Using this output, we can reconstruct the source profile if we know the lobe pattern (impulse response) of our antenna. Here, we shift our attention to a more advanced technique of studying radio sources – interferometry.

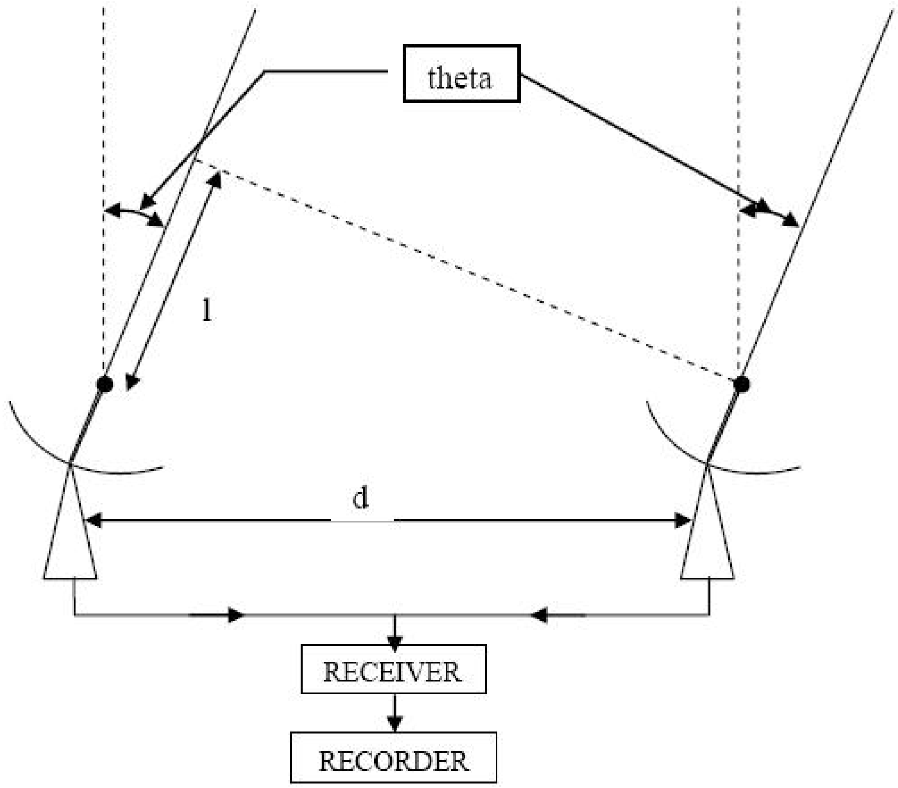
THE SIMPLE ADDING INTERFEROMETER

An interferometer is a combination of two or more radio antennas working together to image the same source. The use of an interferometer is primarily motivated by the need for better resolution. In a radio telescope, the resolution (minimum angle between two point sources that can still be detected to be separate) is equal to the wavelength upon the diameter of the telescope. So, even for a 45 meter size dish of GMRT (Giant Meterwave Radio Telescope, India), the resolution is about 1/45 rad = 1.27 deg, while a small 6” optical telescope has a resolution of 0.00025 deg (about 6,000 times better!) Clearly, we need better resolution in radio telescopes in order to coordinate with other observational astronomers and gain good images of celestial sources. Now, if we were to use a single dish to have resolution of 10-3 degrees at meter wavelengths, the size of the antenna would have to be a few tens of kilometers. This problem is overcome by the use of interferometers, which to some extent simulate a large telescope antenna and thus give better resolution. In terms of the lobe pattern of the antenna, this interferometer as a whole has a narrower main lobe, so is closer to the ideal “impulse response” that we would want from a radio telescope.

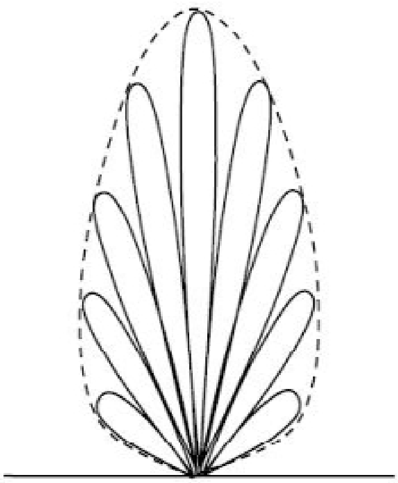
The simple adding interferometer consists of two antennas separated by a distance D between them:



This is in some sense like a double slit for observing an interference pattern. Indeed, it can be proved that the pattern we get by observing uniform radiation in the sky using such a device is identical to a double slit diffraction pattern. Let us do some analysis of the interferometer for getting an insight into its working.



The figure shows two rays coming to the interferometer from a distant, narrow source. The rays are essentially parallel to each other. By simple geometry, we see that the path difference in the rays received by the two antennas is $l = d \sin \theta$, and the phase difference is thus $\varphi = 2\pi d/\lambda \times \sin \theta$. Since we have directly connected the cables of the two antennas together, we will get the maximum output at the receiver only if the two signals are in phase ($\varphi = 0, 2\pi \ldots$) with each other, while the output is zero when the signals are out of phase ($\varphi = \pi, 3\pi \ldots$). Now the net result obtained is show in the polar plot below:



Beam pattern (polar) of a simple two antenna adding interferometer is made up of several beams of width $\Delta\theta$ within the general envelope given by the lobe patterns of the individual antennas.

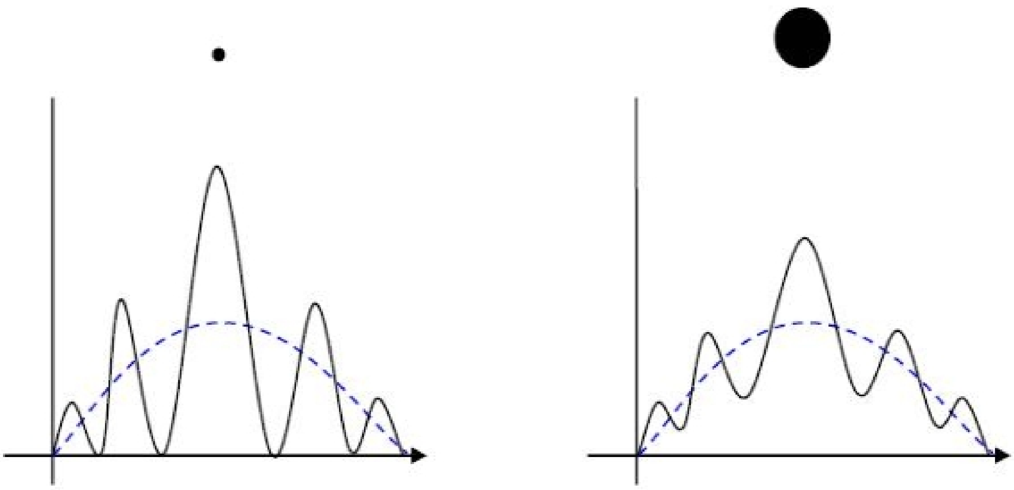
It can be shown that the apparent power at the receiver is given by,

$$P' = p[1 + \cos(\frac{2\pi d\theta}{\lambda})]$$

where P is the power available from one of the antennas. Here, θ is assumed to be a small angle. From this equation, we see that the width of a beam ($\Delta\theta$) is given by λ/d where d is the separation between the two antennas. The resolution of the combined antennas, i.e. the interferometer is $\Delta\theta$, but we cannot ensure which lobe the source is in. For example, as this beam sweeps across a point source, we get the following response at the recorder as shown in the first figure on next page.

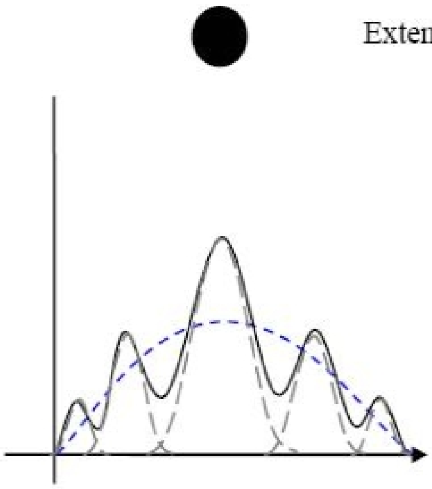
For the small source, which we can approximate as a point source, we get the output as a *convolution* of the signal and the interferometer "impulse response". We can also think of this as *sampling* the input signal at several angles (separated by $\Delta\theta$). That provides an insight into the output of the interferometer swept across an extended source with angular size greater than $\Delta\theta$. Here we can think of the output as the sum of the output recorded by the different lobes (shown in gray dotted lines below). This is a form of *aliasing* and in this case since the "bandwidth" (angular size) of the input is greater than the sampling period ($\Delta\theta$), we get an overlap in the

Output recorded by the two antenna interferometer for two sources
(See explanation below)



Small source | Extended source

Extended source



The output given by the interferometer when swept across an extended source, interpreted as the sum of the outputs given by each individual lobe as it sweeps across the source

The given this output, we can get back the signal if we know the impulse response of our system, i.e. the lobe pattern of the antenna. However, here we note that we have still considered only simplified cases of a one-dimensional scan across a source, while in real astronomy, astronomers are interested in forming 2-D maps of the sky. That adds another degree of complicated mathematics to the process of reconstructing the signal from the output, however, the basic underlying theme remains the same. The actual reconstruction is done from the combined output of several antennas, each with a different baseline (separation), and working together as more advanced types of interferometers. This process is very computation-intensive and professionals use highly specialized hardware to do this job.

References

[1] John D. Kraus, Radio Astronomy, McGraw-Hill, 1966.
[2] Hanbury Brown, A. C. B. Lovell, The Exploration of Space by Radio, Chapman and Hall Ltd., London, 1957.

Signals and Systems in Economic Analysis

by

Vineet Rathi

Sumit Kendurkar

Jiteshu Godara

Abstract:

There ain'tn o such thing as a free lunch economic knowledge in one sentence

An economy is the institutional structure through which individuals in a society coordinate their diverse wants and desires. An ECONOMIC SYSTEM is the system by which the economy is organized. For example, if an economy is organized through markets, it is a market economic system. The study of economics involves the following three central problems:

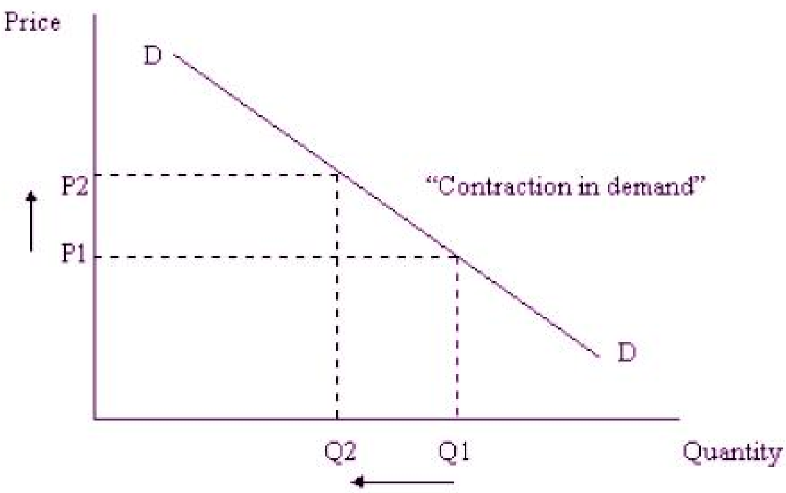
1.What, and how much, to produce.
2.How to produce it.
3. For whom to produce it

There are various factors that become the input signals to the economic system and decide various matters in the economy regarding aggregates (total amount of goods & services produced by society), absolute levels of prices, level of growth of national output (GNP & GDP), interest rates, stock exchange rates, unemployment and inflation. As an example we'll be considering the stock market and the ways in which it functions as a system which takes in diverse signals (inputs) with the prices as outcomes.

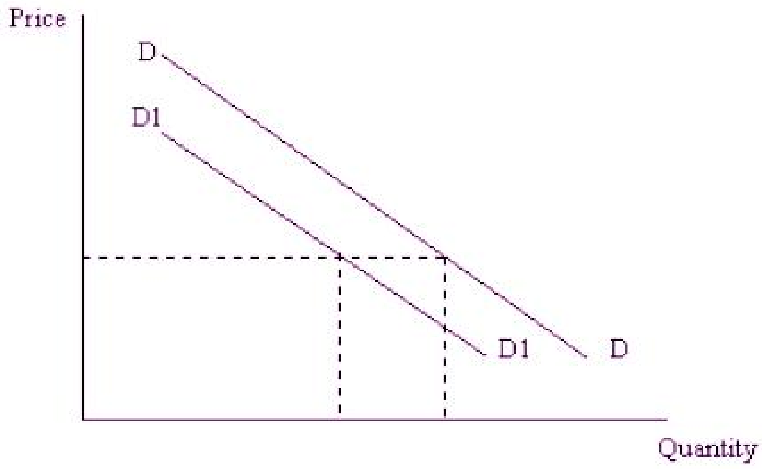
Introduction

THE ECONOMY: Desire refers to people's willingness to own a good. Demand is the amount of a good that consumers are willing and able to buy at a given price.The demand curve labeled DD in the figure below shows the amount of a good one or more consumers are willing and able to buy at different prices

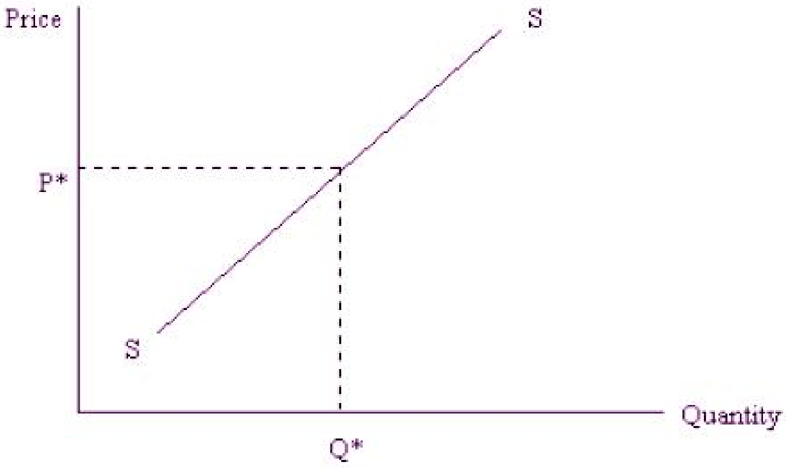
Movements Along and Shifts in Demand Curves: A change in price never shifts the demand curve for that good. In the figure below an increase in price results in a movement up the demand curve. The fall in the quantity demanded from Q1 to Q2 is sometimes called a contraction in demand.



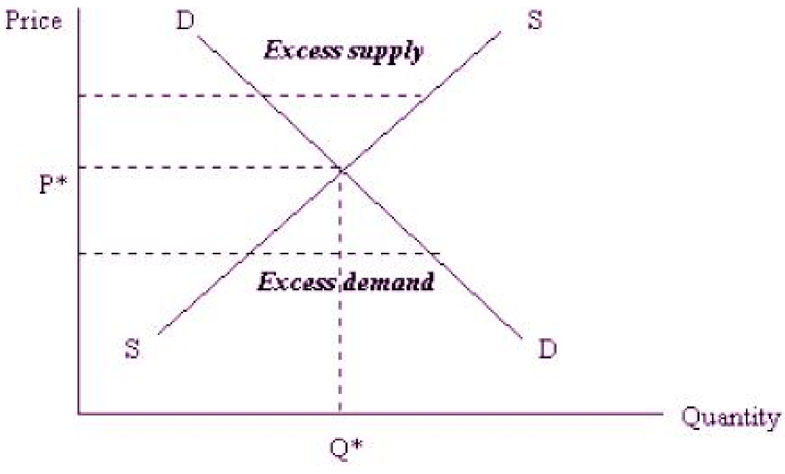
A demand curve shifts only if there is a change in income, in taste or in the demand for substitutes or complements. In the diagram below a decrease in demand has shifted the demand curve to the left. The new demand curve is D1 D1.



Supply: The supply curve labeled SS in the figure below shows the amount of a good one or more producers are prepared to sell at different prices.



Market Price: At prices above the equilibrium (P*) there is excess supply while at prices below the equilibrium (P*) there is excess demand. The effect of excess



supply is to force the price down, while excess demand creates shortages and forces the price up. The price where the amount consumers want to buy equals the amount producers are prepared to sell is the equilibrium market price. All these situations are shown in the diagram below: Now, we are ready to look at the stock market and how these concepts apply there.

Stock Market: Stock is a share in the ownership of a company. We are interested in looking at how the shares trade. In fact, this is the system we will be looking at.

Most stocks are traded on exchanges, which are places where buyers and sellers meet and decide on a price. Before we go on, we should distinguish between the "primary" market and the "secondary" market. The primary market is where securities are created while, in the secondary market, investors trade previouslyissued securities without the involvement of the issuingcompanies. The secondary market is what people are referring to when they talk about "the stock market. Stock prices change everyday by market forces. By this we mean that share prices change because of supply and demand. If more people want to buy a stock (demand) than sell it (supply), then the price moves up. Conversely, if more people wanted to sell a stock than buy it, there would be greater supply than demand, and the price would fall. The principal theory is that the price movement of a stock indicates what investors feel a company is worth. Price times the number of shares outstanding (market capitalization) is the value of a company. But the price of a stock doesn't only reflect a company'sc urrent valueit also reflects the growth that investors expect in the future. We will explore these ideas further in the next section.

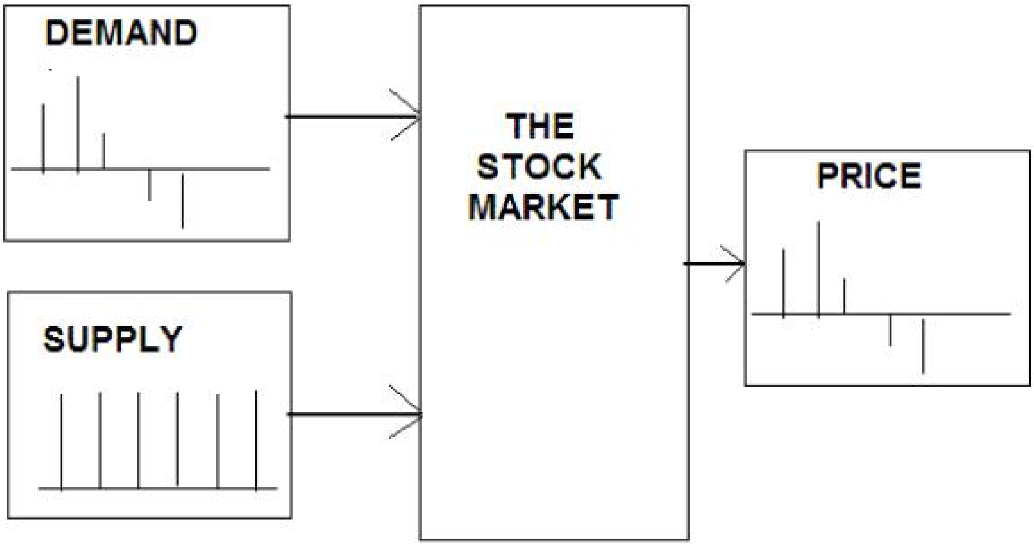
System Specifications & Working

System Specifications:

THE SHARE MARKET AS A SYSTEM:

We look at the share market as a system which determines the price of the share of a company as a function of time. The share market, being a market, has as its determinants, the market forces of supply and demand. Thus, the input to the system are the “signals” of supply and demand as a function of time, giving the price of the share as its output.We only consider the price and the demand at the end of the day, not taking into account the 'intraday't rading which is usually based on speculation.

Working:



Since, we are considering the secondary market, as it is more dynamic and hence interesting, we treat “supply” as constant. If we thought that changes in the supply of stocks are the main cause of stock market rises and declines then, according to this rule, when a company issues new stock we would expect the price of stock to decline. If stock prices are largely determined by the supply of stocks and the market declines prior to an economic decline, we should see a flood of new stock issues before a recession. This does not happen in practice, as new stock issues tend to occur as the economy enters a growth period. This is because the money made from a stock issue is used to increase the output of the company, which causes economic growth to rise. Hence we focus our attention on the forces influencing demand.

Theory behind the working

FORCES INFLUENCING DEMAND: Demand can again be thought of to be determined by a system :

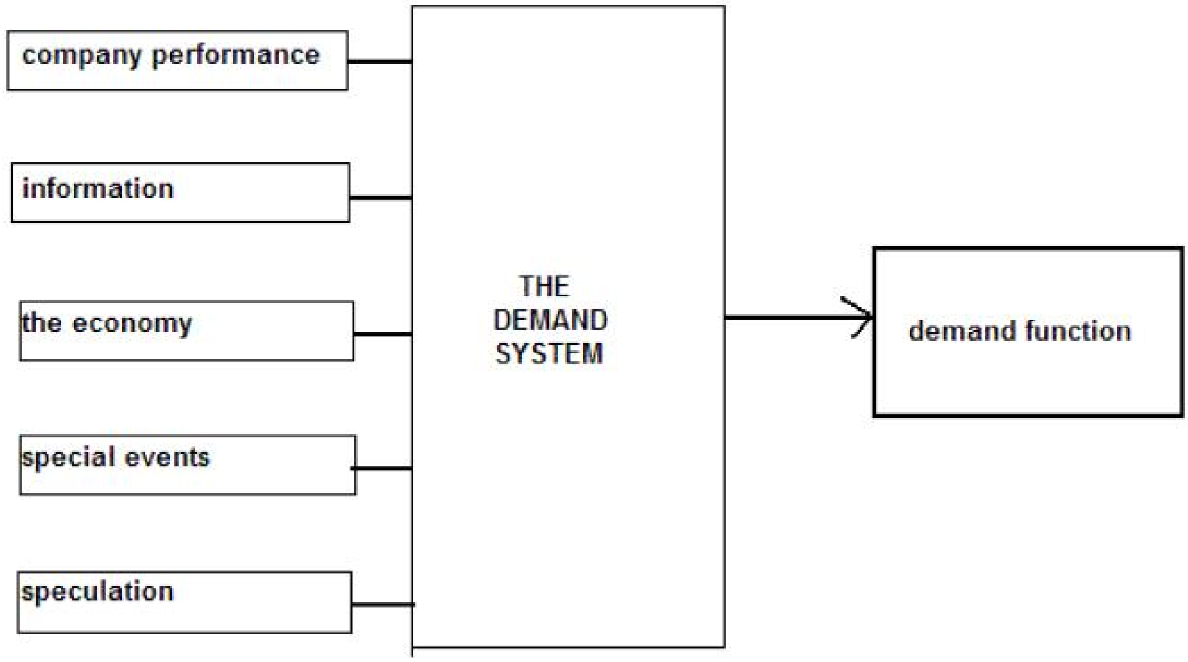
1. Company performance : The quarterly results of a company have a great bearing on the price of a stock, also the dividend paid by a company decides the valuation of its stock. As the revenues earned by the compay increase, the dividend paid increases; buyers increase.

2. Information about status of the company: Information that helps a company's reputation or earnings increases demand. This may range from information like technical advances, new research projects or collabrations taken up by the company, or information about new government policies, or even things such as good weather.

3. The economy : If investors believe a recession is coming, they will sell the stock, driving the price of the stock down. If investors believe a boom is coming, they will increase their estimates of the inherent value because future earnings should be higher than they previously expected. So investors buy the stock. This leads the price of the stock to rise.

4. Special events: Wars / strikes /great depressions/ political instability tend to drive buyers away from the market.

5. Speculation: Past experiences, intelligence, statistical analysis all leads to people trying to anticipate the market's moves, which affect their actions and hence the demand.

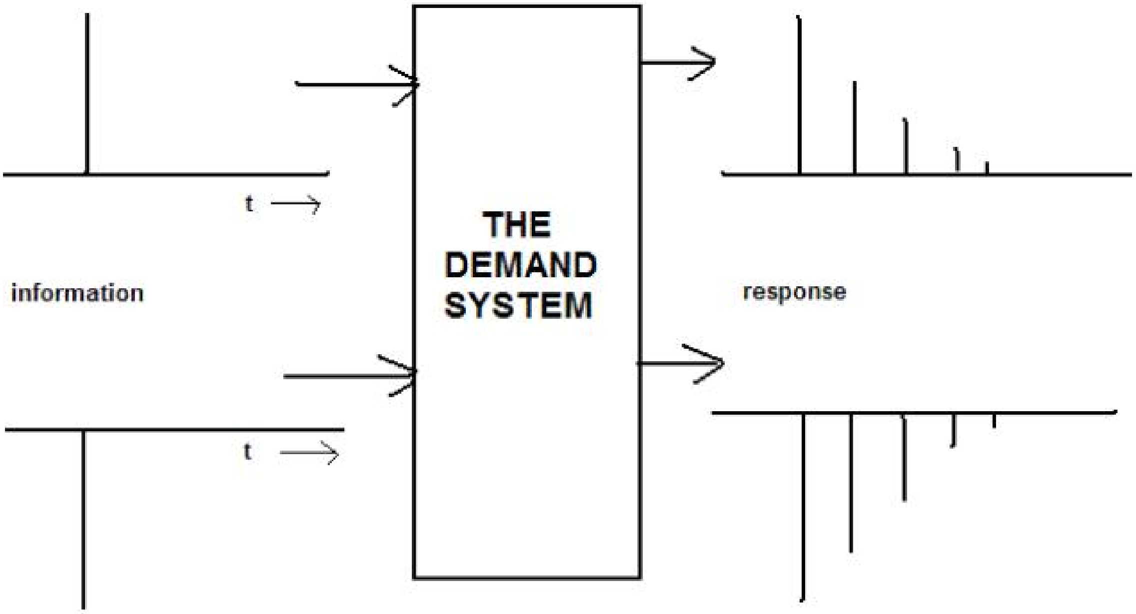


Before we conceptualize the system, we make a few important assumptions:

1. No animal behaviour: everyone behaves according to rational thoughts and ideas based on careful observation of the present and the past state of the market. There are no wild guesses.
2. Honesty: everyone adheres to the norms of the market. No one tries to manipulate the market by using political influences, scams, etc. 3. Mood of the market is not taken into account.
4. The information is assumed to be freely available and everyone is informed. There are no secrets.

Properties of the system

Keeping the above factors in mind we form the following system:The input to our system is the information we get everyday related to the company, and the output is the change in demand . We assume the system to be a discrete time system wherein we take into account the demand change at the end of the day.We can think of this system in terms of a very simple model which takes it's input as a positive spike or impulse, for a piece of information which is beneficial to the company performance, and the output is increase in demand for the stock which persists for several days though reducing in magnitude. Similarly a bad news can be represented as a negative impulse , and the output will be decrease in demand whose magnitude will go on eroding with time.



Though it cannot be quantified , we take an abstract assumption that the magnitude of the impulse represents the extent to which the information is beneficial or detrimental to the company'sp rospects. Now, we analyse our system on the basis of the properties that it may possess:

1) Memory : The system has memory. This is evident from the impulse response of the system which demonstrates that an information has a persisting effect on the demand for a stock. For example If a company comes out with good quarterly results , the stock is likely to remain in high demand for some time.

2) Causality : The system we are considering is noncausal. Investors are constantly watching macroeconomic variables to try and determine when the next downturn in the economy will happen. Investors are often right when they predict the future growth rate of the economy. As a result, they often sell off their shares before the economy goes into a decline making it look like the stock market is causing a recession. In reality the causality runs the other way because the two things that causes price to change are changes in supply or changes in demand. An interesting observation can be made here that a system based just upon the company's performance as inputs would be causal , it is human analysis and the presence of information that make it noncausal.

3) Stability : The system is stable, there exists a balance in the market due to the presence of bulls and bears. An upward march in prices is checked by the presence of bears who earn money on falling stocks, while the bulls ensure that stocks don't plummet to abysmal values.

4) Shiftinvariance: We cannot really comment whether the system is shiftinvariant, this is because the system itself is variable with time, for example the number of the buyers and sellers might change , the initial price of the stock might be different and so on. On the premise that the system does not vary with time, if we consider one input signal at a time keeping others fixed, we find that the system still remains shiftvariant.

5) Linearity: Considering the magnitude of the two informations received to be the same, the change in price would be double the individual change that would have otherwise occurred. This is if we assume the buyers to have an infinite purchasing power, and hence the prices can go on increasing indefinitely without any constraint on the buyer. Also the influence of the bulls and the bears is neglected. But since these assumptions are not justified in the real scenario, the system is not linear. But we can see that the market exhibits liearity in a certain range of prices.

6) Invertibility: Since the system maps more than one input signal to an output signal, we cannot comment upon its invertibility. Now, let us return to the basic system that we were considering. This system has as its inputs the time varying signals of supply and demand and the output is the price at the end of the day. Also, the system works with discretetime signals.

We observe that since the supply is constant the %change in price curve follows the %change in demand curve. Hence the system properties:

1) **Memory :**The system has no memory, as the price is solely determined by the demand at that particular instance of time.

2) Causality: The system is causal, since it possesses no memory.

3) Shiftinvariance: The system is shiftvariant, as the change in price is dependent upon the original price of the stock. For example if a stock is already overpriced the same demand would cause less change in price than it would cause to an undervalued stock.

4) Stability: The system is stable, since bounded change in demand will always result in bounded change in price of the stock.

5) Linearity: The system is linear within a price range but practically due to limited purchasing power and the effect of bulls and bears it is non linear.

6) Invertibility: The system is not invertible since one value of change in demand can produce different changes in price depending upon the original price.

Thus, we conclude that although the practical model for the stock market is fairly complicated, we can get a good idea of how it functions by making a few simplifying assumptions and looking at it from the “signals and systems” point of view.

Use Of Application In Real Life

Since we are dealing with a real life application i.e the stock market, the theory we've proposed can be used to study the stock market in a simpler manner.

Acknowledgement

We would like to thank Prof. V.M .Gadre for his invaluable guidance.

Signals and Systems in Meteorology

by

Sidharth Kumar

Saurabh Shrivastava

Priyesh Neema

Abstract:

In meteorology, short-term forecasts are the main application of radar data. In other fields, the radar data is generally used as a source of data that will help to perform other types of forecasts (like river flow). Short term forecasting of weather events that go beyond simple extrapolation follows several avenues, depending on the method used for forecasting. For example, work on convection initiation is being persuaded to establish radar-based algorithms to predict the outbreak of storms. Another approach is to try to assimilate radar data into numerical models, a process by which the initial conditions of a model are being modified until they can reproduce the storm evolution observed by radar, after which a forecast can be made.

Other applications are many, and are primarily limited by the lack of knowledge of the capabilities of radar data. Hydrologists, both in rural (rivers) and urban (sewers) areas, are becoming natural users of radar data. Agriculture (from plant growth to pest control) and aviation uses (wind and severe weather detection) are also rapidly growing. Some of the important uses of radar in meteorology:

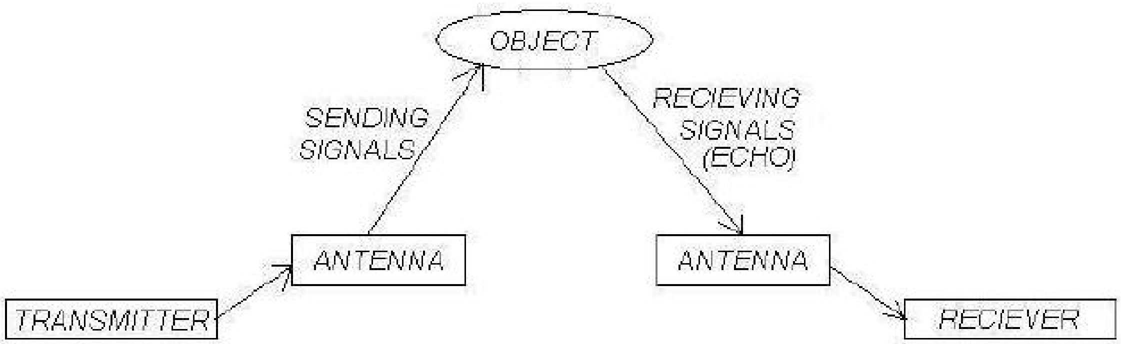
1.Short-term public weather forecasts.
2.Short-term forecasts for aviation.
3.River and sewer flow forecasting.

Motivation:

Radar is one of the rare instruments that can be used to observe the time evolution of weather events in three dimensions, and a unique tool to observe mesoscale events. The radar technology mainly evolved and improved during World War II. Radar was originally developed to satisfy the needs of the military. So, how this weapon can be used to determine weather. Further radar is not capable of recognizing the colours and other specifications of the object. So, how do such a simple thing, which can't even differentiate between the basic colours, can be used to approximate the near future weather. The storm and flood warnings are issued in advance. So, how does meteorologists come to such big conclusion, using weather radars. This invokes curiosity to know about the matter and motivate us to take on this topic for this presentation.

System Specifications

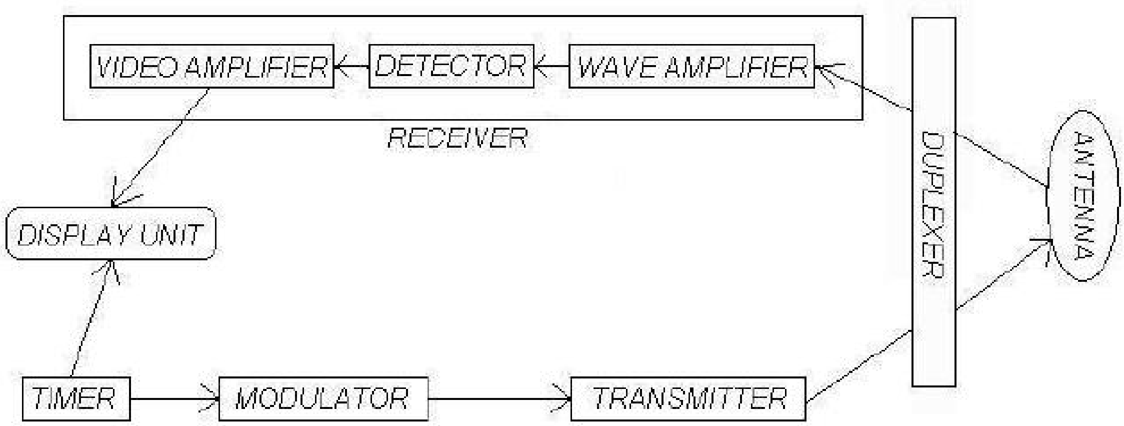
The objective of radar is to transmit radio frequency energy toward a target and to receive and process the reflected energy. Radar equipment can be divided into three basic component groups: the transmitter, the receiver, and the antenna.



Block diagram of Radar

Theory & Working:

Radar operates on the principle that objects in the path of radio waves reflect or reradiate some portion of the wave, enabling those objects to be detected and tracked at long-range distances. Radar is an instrument that basically includes transmitter, antenna, and receiver. Besides them there are timer, modulator and display unit. The complexity of the transmitter and receiver depends upon the modulation format and the signal processing methods to be used and the complexity of antenna depends upon the radar frequency and operational expectations. Usually, radars are monostatic and duplexer is used to switch it between transmission and reception. The basic parts of radar are shown below.



1) Timer: It is also known as the synchronizer. The timer generates the trigger pulses at a specific frequency. Each pulse turns on the modulator. The timer has two outputs, one to the modulator and other to the display unit. The main function of the timer is to produce trigger pulses that start the transmitter and display unit.

2) Modulator: The modulator is turned on by the pulses produced by the timer. It in turn produces another pulse of the required shape and duration with sufficient power and passes it to the transmitter. The modulator also maintains the intervals between pulses. The peak power of the transmitted radio signal pulse depends on the amplitude of the modulator pulse.

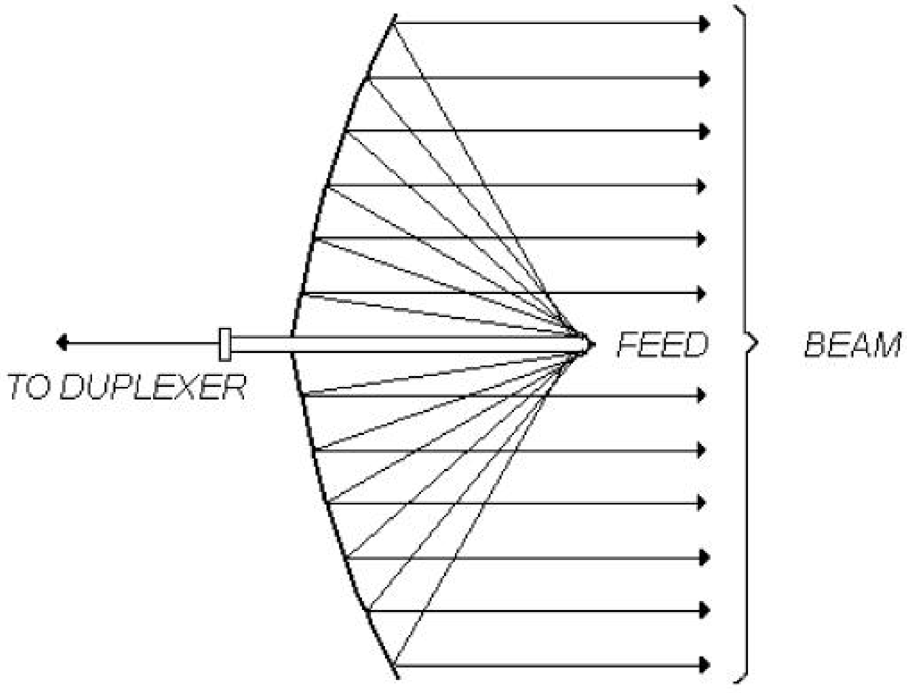
3) Transmitter: The transmitter is turned on by the pulses produced by the modulator. It is basically a simple oscillator that produces a radio wave output for the duration of the pulse provided by the modulator and sends it to the duplexer.

4) Duplexer: The antenna is shared for the transmission of energy and the receiving of the returned energy. It is a fast-acting electronic switch that permits the single antenna to be shared for both the transmission and reception. It is responsible for channeling the transmitted power to the antenna and the returned energy to the receiver.

5) Antenna: The antenna is connected to the duplexer. The transmitter power is radiated into space by a highly directional, high-gain antenna that concentrates the energy into a narrow beam. Most antennas used in radar applications have parabolic reflector. The antenna radiates the radio waves into the reflector. The reflector reflects the waves in the form of a narrow beam that is parallel to the feed. See figure. The narrow, directive beam not only concentrates the energy on target but also measures the direction to the target. On reception, the antenna collects the energy contained in the returned signal and delivers it to the duplexer. The shape and size of the antenna depends upon its use and the radio signals it radiates and receives. The design of the antenna determines its directivity and gain.

6) Receiver: The function of the receiver is to detect desired signals in the presence of noise and interference. The receiver processes the returned energy and sends the signal to a display unit. Target detection is done within the receiver. The received signal passes through a wave amplifier where it is amplified and then passed on to the detector. The detector converts it to DC and filters the signal resulting in a signal in video form. The signal pulse then passes to the video amplifier where its strength is increased and the signal is cleaned up. The signal pulse is then displayed.

Meteorologists recognized the applicability of radar in meteorological studies and based their initial work on radar pattern recognition. The most conceptually simple method is that of contouring. A description of contours of a digitally defined echo area may be derived from a Fourier series expansion. Fourier expansion approximates a digital contour and it can be seen that as few as five harmonics may be used to provide quite a detailed description of an echo if only one intensity level is used.



Antenna

Radar data is usually used for short range forecasting. The meteorological predictions are made by tracking echo areas from one radar picture to the next. This is done by echo matching or echo correlation. The simplest matching procedure is to cross correlate the centroids of each echo (radar picture) with those of the subsequent ones. However cross-correlation techniques may also be used to match portions of one radar picture with portions of subsequent pictures for more accurate results. If there is no significant change from one echo to another the entire picture can be matched. Once the echo movement is defined, then short-period forecasts can be made.

Use of application in Real life:

Weather radars play a vital role in short-term weather forecasting and for meteorological research. They are being used routinely in meteorology to monitor storms and follow their evolution, as well as observe winds and detect regions where severe weather might develop. Some of the important uses are:

1. Short-term public weather forecasts: One of the primary applications of radar data is short-term (0-4 hr) weather forecasting, also known as “now-casting”. It proves extremely useful to issue warnings when severe weather develops rapidly in the vicinity.

2. Short-term forecasts for aviation: Another application of radar is to improve short-term weather forecasts around airports, also known as "terminal forecasts". Aircrafts are very susceptible to severe weather. Icing, hail, very heavy rains, rapid wind shifts and downbursts creates difficult situations for aircrafts. All these meteorological conditions can be detected at least to some extent by weather radars and therefore radar has an important role to help aircrafts steer out of dangerous areas.

3. Radar can be used to find rain - no rain areas using cloud imagery. The rainfall amount can also be approximated.

4. River and sewer flow forecasting: Rivers take their water from rain. Radar measures rainfall and snowfall rates over large areas, and can also be used to make short-term predictions of precipitation amounts. Therefore, radar can be used, in conjunction with appropriate models of river flow, to predict the flow and the level of the water in a river. Sewers can be viewed as a special kind of river. They cover small areas, but in many ways they can be viewed as rivers. If sewers are predicted to overflow, special measures (like diverting water from one part of the network to the next) can be taken to prevent damage to property.

5. Wind velocity: This is done by using Doppler radar which measures the radial velocity, which is the component of the wind going in the direction of the radar (either towards or away). There are many other practical applications of radar data in meteorology like it can be used to determine the refractive index, measurements of moisture, modeling of mold on crops, etc.

Acknowledgement:

We would like to thank professor V. M. Gadre for his excellent guidance throughout this course and also the teaching associates who have helped us a lot in making this report.

Applications of Signals and Systems to fourier optics

by

Adnan Raja

Shriram Shivaraman

Aashwit mahajan

Abstract

In the following report, we discuss application of Fourier analysis in understanding the phenomenon of Diffraction in Optics. We show that under certain approximations (Fraunhofer and Fresnel), diffraction can shown to be a linear system with an impulse response description. A discussion of Fourier Transform in Two Dimensions is included in the initial portion of the report. Then an elementary discussion on Holography follows, with a lucid description of wavefront recording and reconstruction. In the concluding section, we briefly look at some of the applications of Holography.

Keywords: Fourier Transform in Two Dimensions, Scalar Diffraction, Fraunhofer, Fresnel, Holography, Wavefront recording and reconstruction.

Optics and Communication theory

The field of optics is in a way related to communication theory and information sciences. This seems quite reasonable to expect, for both communication systems and imaging systems are designed to collect or convey information. In the former case, the information is generally of a temporal nature (i.e a modulated voltage or current waveform), while in the latter case, it is of a spatial nature (i.e, a light amplitude or intensity distribution over space). Light, as an electromagnetic wave, can be considered as a medium which carries information.

The light signal falls on an object, depending on the transmittance and reflectance of the object plane, the amplitude gets modulated. The signal is received by the eye and converted to nerve signals, which are transmitted to the brain. The brain performs the function of interpreting the information contained in the signals. Here, the eye acts as a signal processing system.

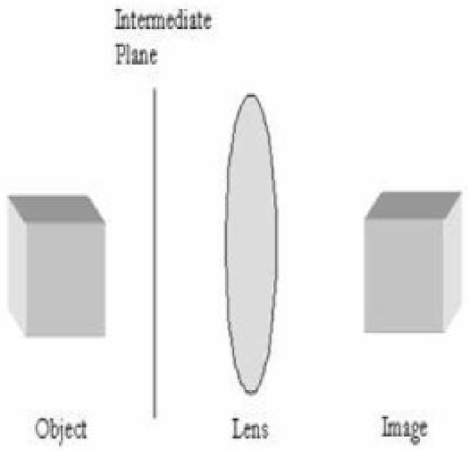


Fig.1 Information concerning the object exists in every plane between the object and the image

On the other hand, if the light signals pass through slits or lenses, physical phenomenon like diffraction, refraction etc. take place and as a result, patterns and images are formed. Information concerning the object is transmitted by light to the image plane. This information is also definitely present at all intermediate planes as the light propagates. If this propagating information is recorded in some way, and at a later time and place allowed to resume its propagation, the original image can be reconstructed. These, roughly, were the thoughts that led Dennis Gabor to conceptualize holography.

Before we go further, let us have a look at an important mathematical tool required for this

Fourier analysis in two dimensions

The Fourier Transform (alternatively, the Fourier spectrum) of a complex function **g** of two independent variables, x and y, will be represented here as F(**g**) and is defined by

$$\mathcal{F}(g) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\exp[-j2\pi(f_X x + f_Y y)]dxdy$$

Clearly, the transform is itself a complex valued function of two independent variables f_X and f_Y, which are referred to as frequencies.

Similarly, the inverse Fourier Transform of a function $G(f_X, f_Y)$ is represented by $F^{-1}(G)$ and is defined as

$$\mathcal{F}^{-1}(\mathcal{G}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathcal{G}(f_X, f_Y)\exp[j2\pi(f_X x + f_Y y)]df_X df_Y$$

As in the case of single dimension case, the set of sufficient conditions for the existence of the Fourier Transform in two dimensions are:

1. g must be absolutely integrable over the infinite x-y plane.
2. g must have only a finite number of discontinuities and a finite number of maxima and minima in any finite rectangle.
3. g must have no infinite discontinuities.

The properties of the Fourier Transform in two dimensions, viz. linearity, scaling, convolution theorem etc., are analogous to those in single dimensional case. For the case of separable functions, the Fourier Transform in two dimensions takes a very simple form, i.e. it is equal to product of the one-dimensional Fourier Transforms of the two single variable functions they are composed of.

$$\begin{aligned} &\textit{If } f(x,y) = g(x)h(y) \\ &\textit{then } \mathcal{F}(f(x,y)) = \mathcal{F}_X(g)\mathcal{F}_Y(h) \end{aligned}$$

Now, we define certain frequently encountered functions in Fourier Optics:

Function	Transform
$\exp[-\pi(x^2+y^2)]$	$\exp[-\pi(f_X^2+f_Y^2)]$
$rect(x)rect(y)$	$\mathrm{sinc}(f_X)\mathrm{sinc}(f_Y)$
$\Lambda(x)\Lambda(y)$	$\mathrm{sinc}^2(f_X)\mathrm{sinc}^2(f_Y)$
$\delta(x,y)$	1
$\exp[j\pi(x+y)]$	$\delta(f_X-1/2, f_Y-1/2)$
$\mathrm{sgn}(x)\,\mathrm{sgn}(y)$	$1/[(j\pi f_X)(j\pi f_Y)]$
$comb(x)comb(y)$	$comb(f_X)comb(f_Y)$

1. Rectangle Function:

$$rect(x) = \begin{cases} 1 & |x| \le \frac{1}{2} \\ 0 & otherwise \end{cases}$$

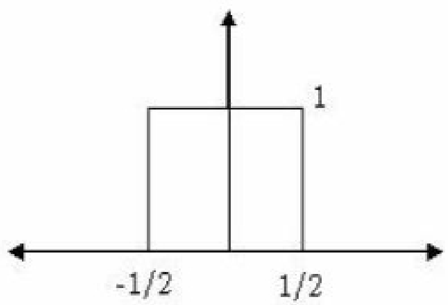
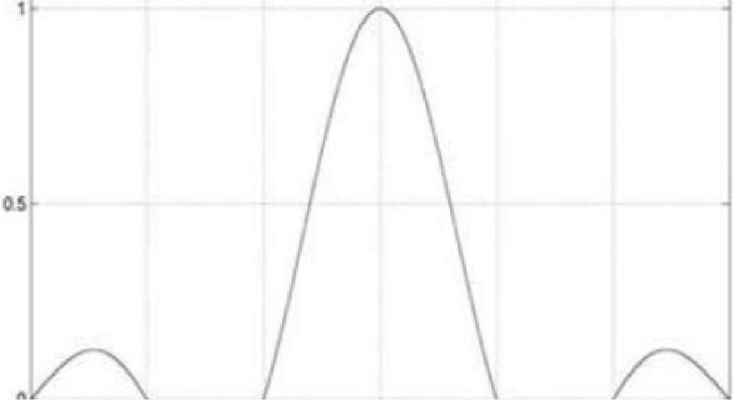


Fig.2 : The rectangle function

2. Sinc Function:

$$\mathrm{sinc(x)} = \sin(\pi x)/(\pi x)$$



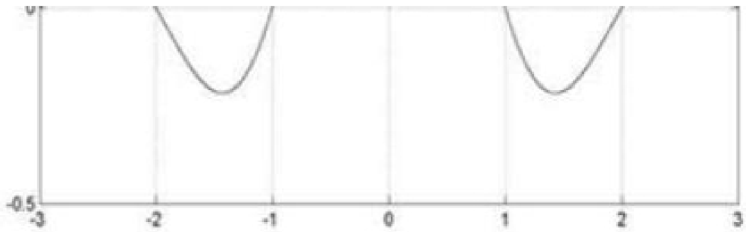


Fig.3 : The sinc function

1. Sign Function:

$$\mathrm{sgn}(x)=\begin{cases}1 & x>0\\ 0 & x=0\\ -1 & x<0\end{cases}$$

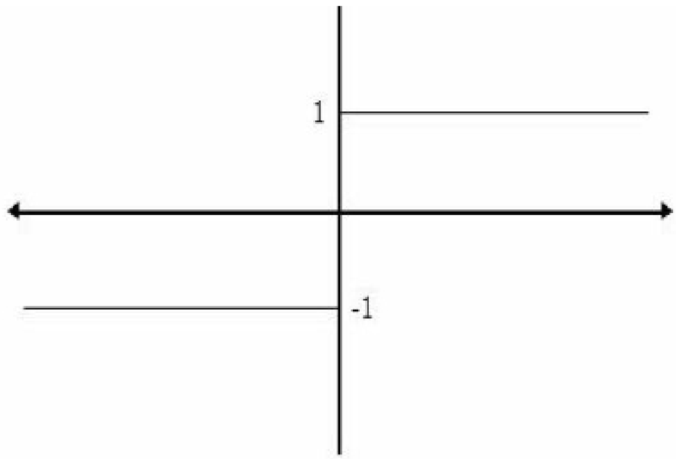


Fig.4 : The sign function

1. Triangle Function:

$$\Lambda(x)=\begin{cases}1-|x| & |x|\leq 1\\ 0 & otherwise\end{cases}$$

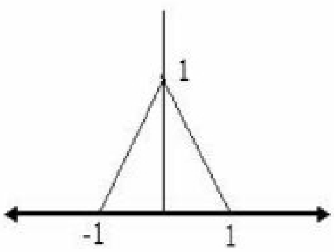


Fig 5: The triangle function

1. Comb Function:

$$comb(x)=\sum_{n=-\infty}^{+\infty}\delta(x-n)$$

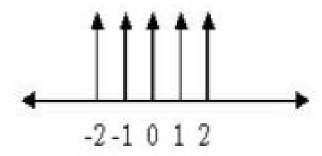


Fig. 6 : The Comb function

1. Circle Function:

$$circ(\sqrt{x^2+y^2}) = \begin{cases} 1 & \sqrt{x^2+y^2} \le 1 \\ 0 & otherwise \end{cases}$$

Now, we look at the Fourier Transforms of some separable functions of two variables, which are composed of the above functions and are frequently encountered in Fourier Optics:

Scalar Differentiation

According to this approximation, light is treated as a scalar phenomenon, i.e only the scalar amplitude of one transverse component of either the electric or magnetic fields is considered. Assumption is made that the other components can be treated similarly. Such an approach neglects the fact that the various components of the electric and magnetic field vectors are coupled through Maxwell Equations, so they cannot be treated independently. However, the scalar theory leads to accurate results in the microwave region if two conditions are met:

1. Diffracting Aperture must be large compared to the wavelength.
2. Diffracted fields must not be observed too close to the aperture.

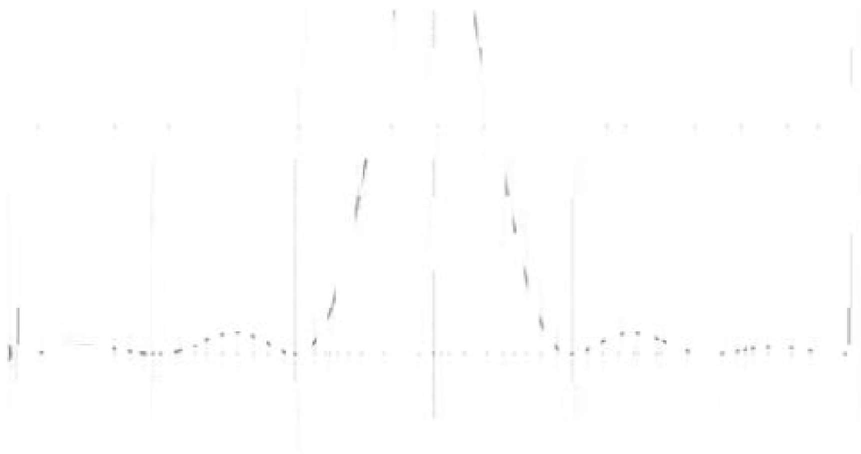
However, these conditions are generally satisfied in most of the practical situations encountered.

Angular Spectrum of Plane Waves

We here consider monochromatic light having wavelength λ. Also assume that the wave is incident on x-y plane and traveling in positive direction.

Let the complex field across the plane z=0 be represented by $U_1(x_1, y_1)$ and across the plane $z=z_0$ be represented by $U_2(x_1, y_1)$. Now consider the Fourier Transform A_1 of U_1,

$$A_1(u_1, v_1) = \iint U_1(x_1, y_1) \exp[-j2\pi(u_1x_1 + v_1y_1)]dx_1dy_1$$



Cross-section of the Fraunhofer Diffraction pattern of a rectangular aperture

The Fourier transformation can be regarded as a decomposition of a complicated function into plane waves propagating with direction cosines <α, β, γ> such that

$$\alpha^2 + \beta^2 + \gamma^2 = 1$$

and, $u = \alpha / \lambda$, $v = \beta / \lambda$

For, $\alpha^2 + \beta^2 > 1$, the component waves decay exponentially along the z-axes. These are called the evanescent waves and are neglected.

Therefore, in the angular spectrum $(-1/\lambda) \leq v, u \leq (1/\lambda)$, we allow each of these component waves to propagate to the plane x_2-y_2 plane to give the corresponding angular spectrum $A_2(u, v)$. Since, every complex disturbance satisfies the Helmholtz equation $(\nabla^2 + k^2)U = 0$ at all source-free points, this gives $A_2(u, v)$ (on applying the Helmholtz equation to U_1 and solving the subsequent differential equation) as:

$$A_2(\boldsymbol{u},\boldsymbol{v}) = A_1(\boldsymbol{u},\boldsymbol{v})\exp[\,j2\pi z\sqrt{1-\boldsymbol{\alpha}^2-\boldsymbol{\beta}^2}\,]$$

Fresnel Diffraction

In this case, the distance of the screen from the source is very large as compared to the size of the aperture. Hence, here α, β are very small, so using the Binomial Expansion and approximating $\sqrt{1-\alpha^2-\beta^2}$ by $1-(\alpha^2+\beta^2)/2$, we get

$$A_2(\boldsymbol{u},\boldsymbol{v}) = A_1(\boldsymbol{u},\boldsymbol{v})\exp[\,jkz\,]\exp[-j\pi\lambda(\boldsymbol{u}^2+\boldsymbol{v}^2)]$$

So, taking the inverse Fourier transform to obtain the disturbance $U_2(x, y)$ gives

So, taking the inverse Fourier transform to obtain the disturbance $U_2(x, y)$, gives

$$U_2(x, y) = \mathcal{F}^{-1}\{A_1(u,v)\exp[jkz]\exp[-j\pi\lambda z(u^2+v^2)]\}$$

$$\Rightarrow \quad U_1(x, y) = \frac{\exp[jkz]}{j\lambda z}.U_1(x, y) * \exp[jk(\frac{x^2+y^2}{2z})]$$

(Approximation to Huygen-Fresnel Principle)

$$\therefore \quad U_2(x, y) = U_1(x, y) * H(x, y)$$

where, $H(x,y) = \frac{\exp[jkz]}{j\lambda z}\exp[jk(\frac{x^2+y^2}{2z})]$

Thus, as shown above, Fresnel Diffraction is an LSI system with impulse response H(x, y).

Properties:

1. Linear: As has been shown above.

$$|H(x,y)| = \frac{1}{\lambda z}$$

Clearly, the impulse response is bounded and hence, the system has BIBO stability.

Fraunhofer Diffraction:

In this case, the distance of the screen is considered to be even larger so that $z >> \max\{\frac{1}{2}k(x^2+y^2)\}$, then the argument of the exponential term $\exp[jk\frac{x^2+y^2}{2z}]$ tends to zero and this term can be taken as 1. So, the impulse response can be written as

$$H(x,y) = \frac{\exp[jkz]}{j\lambda z}$$

and, the Fraunhofer distribution takes the form

$$U_1(x, y) = \frac{\exp[jkz]\exp[jk\frac{x^2+y^2}{2z}]}{j\lambda z}\iint U_1(\alpha,\beta)\exp[-j\frac{2\pi}{\lambda z}(x\alpha + y\beta)]d\alpha d\beta$$

Example of Fraunhofer Diffraction Pattern

Consider, a rectangular aperture with an amplitude transmittance given by

$$t(x,y) = rect(x/a)rect(y/b)$$

The constants a and b represent the dimensions of the aperture in the x and y directions respectively. If the aperture is illuminated normally by a unit-amplitude monochromatic plane wave, the field distribution across the aperture is equal to the transmittance function. The Fraunhofer Diffraction pattern of the aperture can then be written as :

$$U_2(x,y) = \frac{\exp[jkz]\exp[jk\frac{x^2+y^2}{2z}]}{j\lambda z}\mathcal{F}\{U_1(x,y)\}\Big|_{\substack{f_X = x/\lambda z \\ f_Y = y/\lambda z}}$$

Noting that $\mathcal{F}\{U_1(x,y)\} = ab\,\mathrm{sinc}(af_X)\mathrm{sinc}(bf_Y)$

We find

$$U_1(x,y) = \frac{\exp[jkz]\exp[jk\frac{x^2+y^2}{2z}]}{j\lambda z}ab\,\mathrm{sinc}(ax/\lambda z)\mathrm{sinc}(by/\lambda z)$$

Therefore, Intensity $I(x,y) = [ab\,\mathrm{sinc}(ax/\lambda z)\mathrm{sinc}(by/\lambda z)/\lambda z]^2$

Holography-An Elementary discussion

In 1948, Dennis Gabor proposed a novel two-step, lensless imaging process which he called wavefront reconstruction. The information required for reconstruction is present in both the phase and amplitude of the light waves diffracted by the object. The wavefront reconstruction process consists of two distinct operations : a recording or information storage operation, and a final reconstruction operation. These two aspects are discussed below.

Recording Amplitude and Phase:

As discussed above, it is necessary to store information about both the amplitude and phase of the

Image Formation by Wavefront Reconstruction

In the previous section, we considered the problem of reconstructing a wavefront from the recording surface. Now, we will consider the wavefront reconstruction process as a means of image formation. To adopt this point of view, we note that the wave component U_3 in equation (1), which is just a duplication of the original object wave A(x, y), must appear to an observer to be diverging from the original object, in spite of the object being not present. Thus, during reconstruction, the transmitted wave component U_3 may be regarded as generating a virtual image of the original object. In a similar fashion, when the conjugate reference wave $\overline{B(x,y)}$ is used as the illumination, the wave component U_4 of equation (2) also generates an image, but this time the image is real.

therefore, necessary to somehow convert the phase information into intensity variations for recording purposes. A standard technique for accomplishing this task is interferometry, i.e a second wavefront of known amplitude and phase is added to the unknown wavefront, as shown in fig. The intensity of the sum then depends on both the amplitude and the phase of the original wavefront. Thus, if

$$A(x,y) = a(x,y)\exp[-j\phi(x,y)]$$

represents the wavefront to be recorded, and

$$B(x,y) = b(x,y)\exp[-j\psi(x,y)]$$

represents the "reference" wavefront with which A(x, y) interferes, the intensity of the sum is given by

$$\mathcal{I}(x,y) = |B(x,y)|^2 + |A(x,y)|^2 + 2b(x,y)a(x,y)\cos[\psi(x,y) - \phi(x,y)]$$

While the first two terms of the above expression depend only on the intensities of the two waves, the third depends on their relative phases. Thus, information about both the phase and amplitude of A(x, y) has been recorded. Such a recording of the interference pattern between two wavefronts may be regarded as a Hologram.

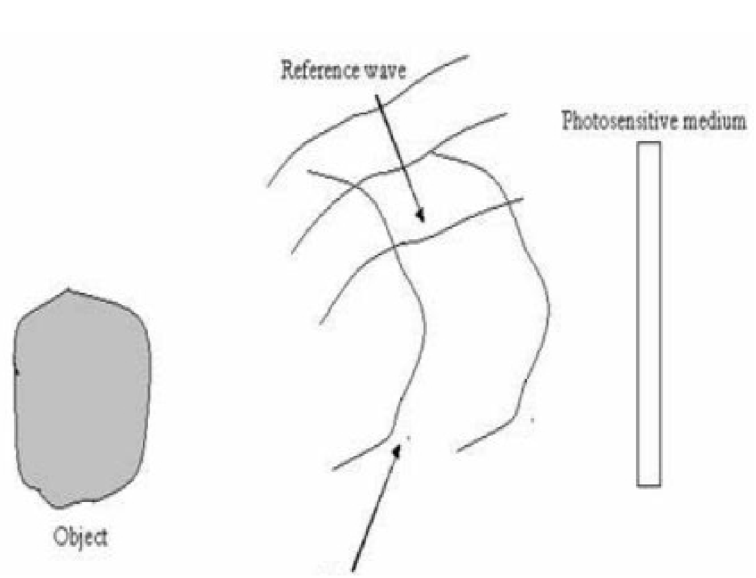


Fig. 8: Wavefront Recording

Wavefront Reconstruction

Here we assume that the amplitude transmittance of a developed film is of the functional form

Now, suppose that the developed transparency is illuminated by a reconstruction wave C(x, y). Then the light transmitted by the transparency is

$$C(x,y)t_f(x,y) = t_b C + \beta A\bar{A}C + \beta \bar{B}CA + \beta BC\bar{A}$$
$$= U_1 + U_2 + U_3 + U_4$$

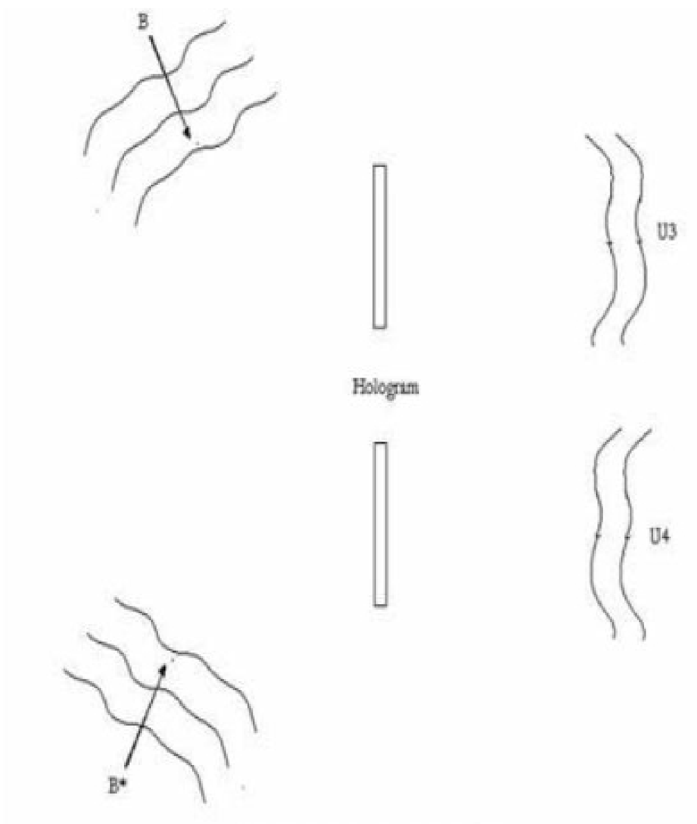


Fig.9 Wavefront Reconstruction

Consider the special case, when C(x, y) is simply an exact duplication of the original reference

Consider the special case, when C(x, y) is simply an exact duplication of the original reference wavefront B(x, y). Then

$$U_3(x,y) = \beta \,|\, B(x,y) \,|^2\, A(x,y) \qquad (1)$$

From the above expression it is clear that U_3 is an exact duplication of the original object wavefront A(x, y) upto a simple multiplicative constant.

In a similar fashion, if $C(x,y) = \overline{B(x,y)}$ then

$$U_4(x,y) = \beta \,|\, B(x,y) \,|^2\, \overline{A(x,y)} \qquad (2)$$

which is proportional to the conjugate of the original object waveform. The terms other than U_3, U_4 in the equation can be regarded as extraneous interference components, which may be separated out using a suitable filter.

Holography-An Elementary discussion

1. **Microscopy:**

An important application of holography is in the construction of X-ray microscopes. The basic concept here is the use of X-ray illumination for the recording of a hologram and the use of optical illumination during reconstruction. Enormous magnifications can be achieved by this change of wavelength but, more importantly, the use of X-ray for illumination can yield resolutions of few Angstroms.

1. **Interferometry:**

Holography offers the capability of performing several kinds of interferometry. Any use of holography to achieve superposition of two images will result in a potential method of interferometry. One unique feature of holographic interferometry is that the wave that serves as a standard of comparison can be stored and used at a later time.
One of the dramatic demonstrations of the potential of holographic interferometry is outlined in the following experiment performed by Brookes et al. to outline the shock waves generated by a bullet in air as it passes over a diffuse background.
The first pulse records the hologram of only the diffuse background, while the second pulse records the hologram of the bullet in flight in front of the same diffuse background. The shock waves generated by the bullet produce fluctuations in the refractive index of air. As a consequence, the two images of the diffuse background will mutually interfere, producing interference fringes which outline the shock wave generated by the bullet.

ACKNOWLEDGMENTS

We are thankful to Prof. V.M Gadre and all course associates of EE210, Signals and Systems, for providing us the motivation and encouragement for exploring this domain of signals and systems.

APPLICATION OF POWER CONTROL AND CDMA IN ADHOC NETWORKS

by

Shagun Dusad

Anshul Khandelwal

Shirish Mehta

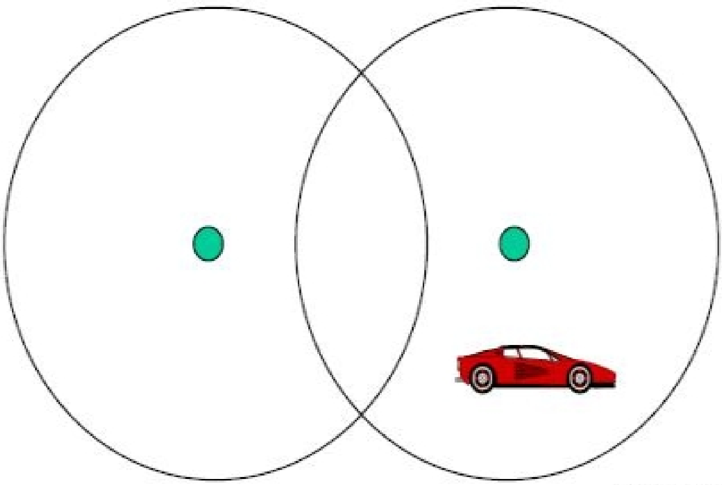
Abstract

Here we present some basic concepts related to CDMA (code division multiple access) and mobile adhoc networks (MANETs). We discuss how CDMA based power controlled MAC (medium access control) protocols can be used to increase the network capacity and throughput. CDMA is relatively new technology in the field of MAC and has its own advantages. In the case of transmission with the fixed power for any destination, the node, transmitting signal, enforces the neighborhood nodes to remain silent for the transmission duration. Thus prevents the reuse of channel. This disadvantage can be alleviated with the use of power controlled MAC protocols. Use of CDMA in MAC protocols allows the multiple simultaneous transmission in the network on the basis of interference level considerations. Some of the issues and design aspects in using CDMA in MAC layer are discussed. Finally, we conclude with some real life applications of adhoc networks.

Index Terms: CDMA, Power Control, Adhoc Networks, CSMA/CA, Spread Spectrum, MAC.

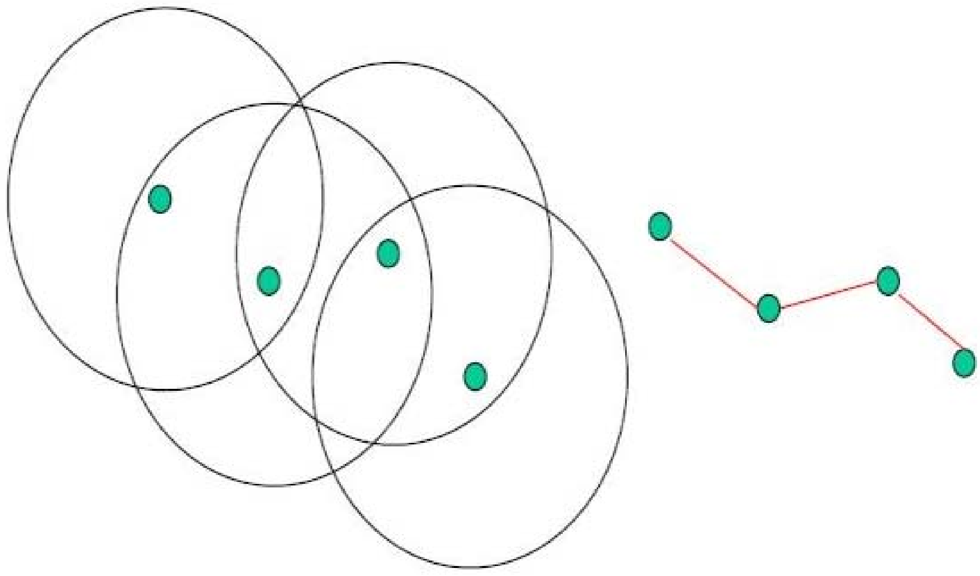
Introduction

Mobile Adhoc Networks: Mobile adhoc networks (MANETs) have been a subject of extensive research for the past few years. These adhoc networks consist of autonomous nodes that collaborate in data transfer, acting both as end-systems and as routers simultaneously. In case of cellular networks there is single hop wireless connectivity to the wired world i.e., the space is divided into cells and the base station is responsible to communicate with hosts in its cell. Hand-off occurs when a mobile host start communicating via a new base station.



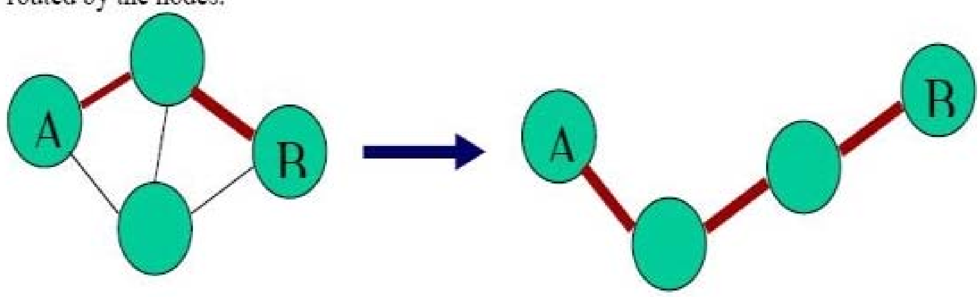
source:shilling

On the contrary, in Adhoc domain we need we need to traverse multiple links to reach destination.



source:schilling

Here the host movement is frequent causing topology change. So the data must be routed by the nodes.



CDMA

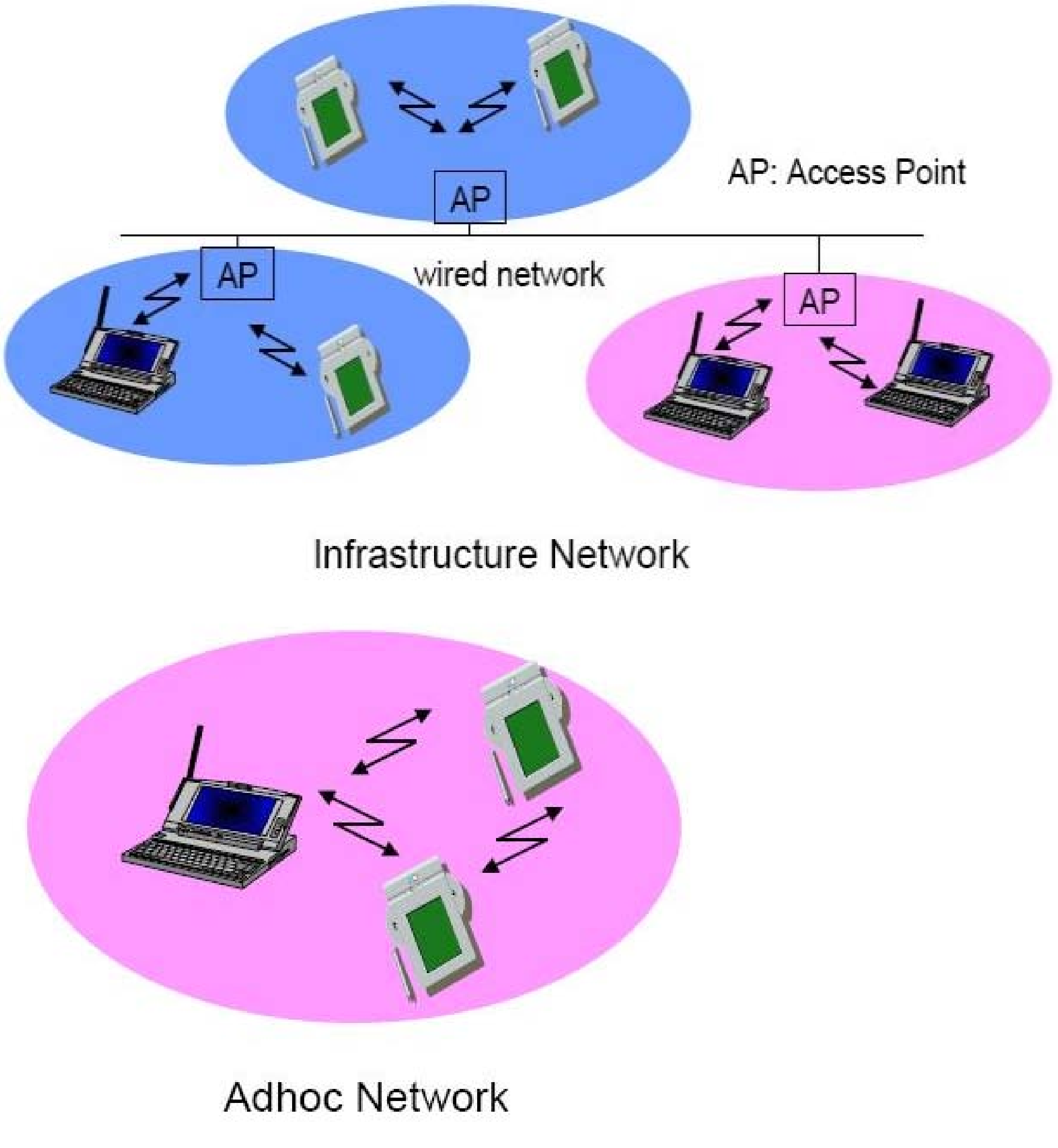
In wireless networks, one of the fundamental challenges is the developing of efficient MAC (medium access control) protocols. MAC comprises all mechanisms that regulate user access to a medium so that many users may use the same medium with minimum interference. The base technologies used in MAC are FDMA (frequency division multiple access), TDMA(time division multiple access) and CDMA (code division multiple access). Extensive research and deployment has fortified that, in the above three technologies, CDMA is the best technology for the power controlled multiple access systems providing high network capacity. So in this section we present some basic concepts related to CDMA. CDMA is based on spread spectrum techniques in which each user's packet transmissions acquire the entire spectrum, and there is no allocation of bandwidth for any user. A digital signal of bandwidth R bits/s occupies a bandwidth of R Hz if baseband pulse transmission is used for actually transmitting packets. In CDMA, this signal is multiplied (or the XOR operation is performed which is discussed in the "system specification" section) using a pseudo-random sequence or PN code of bandwidth W bits/s (and hence occupies a spectrum of W Hz). These PN codes may be unique to a particular node or may be common between all the nodes. The resultant signal itself occupies a bandwidth of W Hz and this process is called spreading. Ideally this PN sequence must be a perfectly random noise, but this will be impossible to generate at the transmitter, and more-so impossible to regenerate at the receiver for reception. So, the pseudo-random sequences are the next-feasible option which approximate the binary random sequence sufficiently. The pseudo random codes (noise), with which the transmitted signal is multiplied, should have a good autocorrelation and should be orthogonal to other codes. Orthogonal in code space has the same meaning as in the standard space(i.e. the three dimensional space) that the inner product with the other codes should be zero. Having good autocorrelation imply that the inner product with itself produces large value but when it is shifted by one unit, inner product stays at the low value until the code matches itself again perfectly.

Spreading the transmissions provides multi-faceted benefits and the transmitted signal is more "robust" than ordinary baseband pulse transmissions. Benefits of CDMA include increased received quality due to protection against multipath fading, protection against jamming, frequency diversity, interference cancellation and better performance with regard to mobility of the nodes.

CDMA based power controlled MAC protocol for Adhoc Networks: The channel access in adhoc networks is contention based due to its distributed nature which leads to a low throughout under high traffic loads on the system. Currently, the de facto mechanism for MAC protocol follows the "carrier sense multiple access with collision avoidance (CSMA/CA)" paradigm. It has been observed that throughput of a network is significantly low in CSMA/CA based MAC implementations. The low throughput is attributed to the contention-based nature of CSMA/CA. In CSMA/CA with fixed transmission power based schemes, the protocol executes a handshake mechanism to reserve the channel in order to avoid collision from nodes located within the transmission range of transmitter and receiver. This results in a "silence zone" around the source and the destination, where no node can transmit or receive packets until the data transmission is complete. This class of protocols do not take advantage of the relative distance between the transmitter and the receiver, which results in transmissions done with excess power than that required for correct reception at the destination node. Therefore these protocols are not efficient with respect to power consumption. But power controlled MAC protocols for adhoc networks makes use of this aspect by transmitting only at a power level required for correct reception. This concept of power control is also known as Interference Limited Access as by controlling power, Multiple Access Interference(MAI) is bounded. This power control scheme has been effectively used in cellular wireless systems to increase capacity and performance, especially in CDMA networks. In a cellular environment, the nature of any transmission is point-to-multipoint with Base Station acting as the coordinator to measure the MAI in the system. In contrast, the transmission in an adhoc network is multipoint to multipoint with lack of any coordination. Hence, the multiple access capabilities of CDMA are extremely useful in adhoc networks.

Motivation:

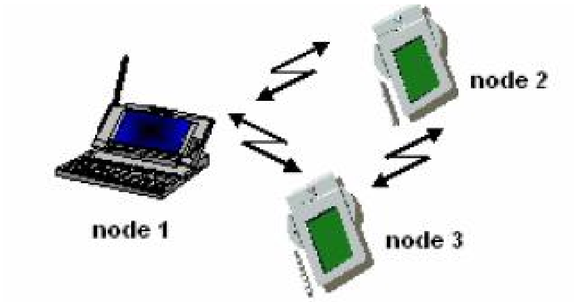
In wireless network, setting of fixed access points and backbone infrastructure is not always viable. For instance, infrastructure may not be present in the disaster area or war zone or in remote terrain. In such arenas, Mobile Adhoc Networks (MANETs) are extremely useful. Apart from such scenarios, the widespread availability of mobile computing devices and handheld devices like palm-tops, portables and personal digital assistants have increased the user base of Wireless Local Area Networks (WLANs) which has typical features of adhoc networks. They are easy to deploy and do not use backbone infrastructure support. But design of efficient media access protocols for such self-organizing networks is fraught with many challenges. There is randomness inherently associated with adhoc networks. It is really fascinating that these random attributes make them more useful and flexible as compared to infrastructured wireless design in their realm. The art of optimization of available resources displays how this chaotic nature of adhoc network wins over its counterparts. The flexibility they offer is admirable. These thoughts were a source inspiration and motivation to write about how their efficiency can be increased by the use of CDMA and Power control.



Source: schilling

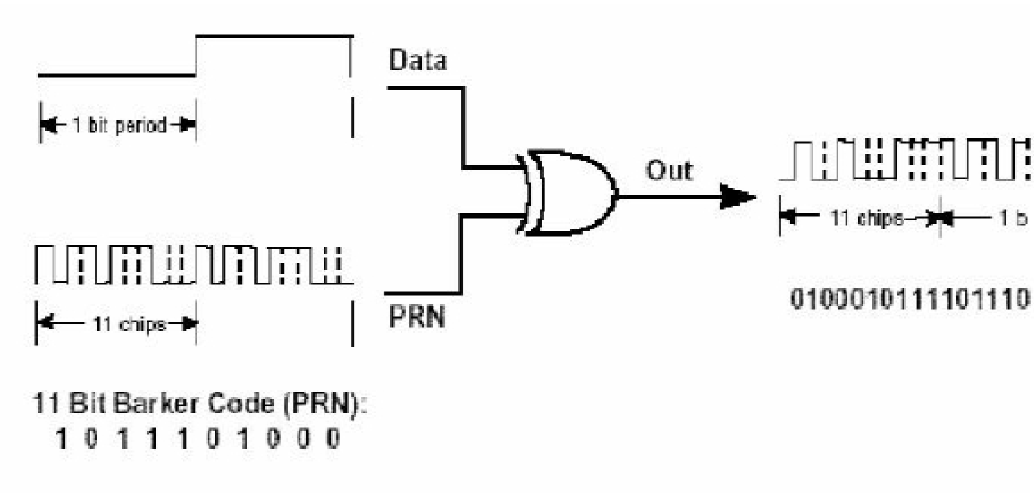
System Specifications

From the point of view of “Signals and Systems”, nodes can be described as systems. For example if adhoc networks are used in cellular industry then the mobile phones are the nodes and they can be thought as the systems in which incoming signal is electromagnetic waves and the output signal is voice. (it can be understood other way round also.) As in the broad sense, adhoc network can be thought as a collection of nodes or terminals so using the above mentioned view we can say that adhoc network is a collection of systems with all the systems invertible.

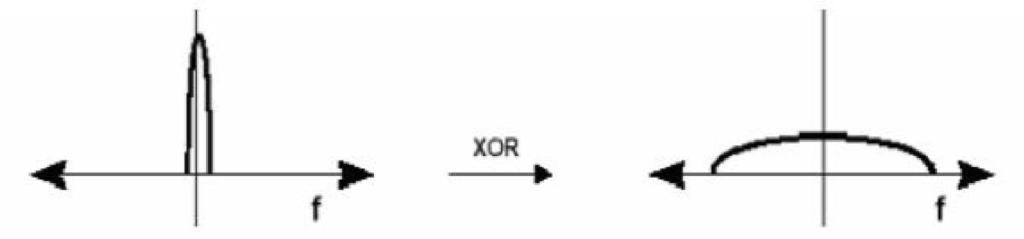


Adhoc Network as a collection of systems (nodes)

Like this we can define CDMA as a system which takes a string of bits as input and gives another string of bits as output. The operation this system performs on the input string using another string of bits (called pseudo random sequence) is XOR, which is shown in the figure below.

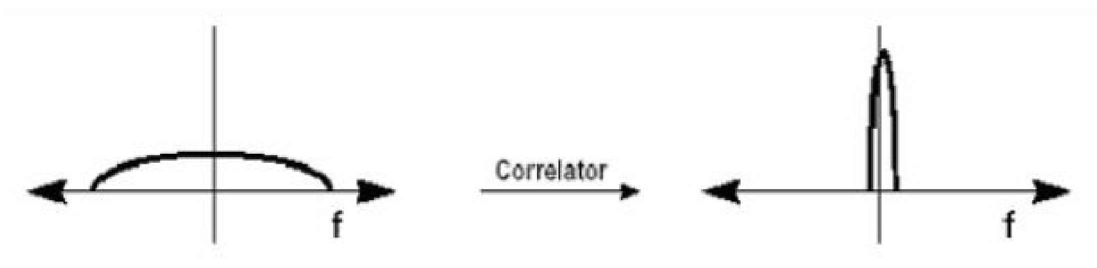


Digital Modulation of Data with PRN Sequence



Effect of PN Sequence on Transmit Spectrum

At the receiving node, the reverse process occurs i.e. the signal is converted into a string of bits and then that string acts as an input to another system(where the inverse operation occurs). The output comes out is another string of bits. In this operation input string is correlated with the same pseudo random sequence, and then the actual string of bits is recovered (the output one in this case) as shown in the figure below.



Received Signal is Correlated with PN to Recover Data and Reject Interference

WORKING AND THEORY BEHIND THE APPLICATION

Power control:

The importance of power control MAC protocol can be explained by taking the case of fixed transmission power based and RTS/CTS based CSMA/CA protocols. Consider the four stations shown in

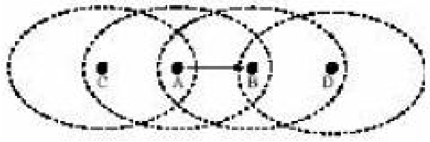
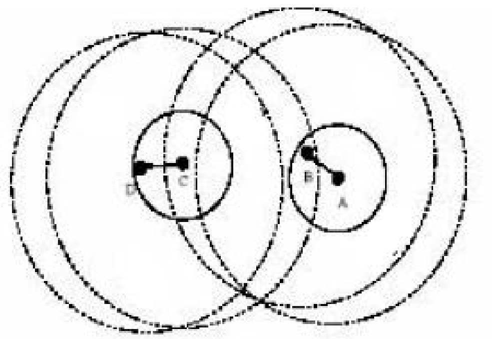


Fig. 1 **CSMA/CA protocol: floor acquisition hidden/exposed terminal**

The transmission range of individual nodes is shown by the dashed lines. Let A wants to transmit data to B. Since C is in the listening range of sender A (C can hear transmissions from station A) but not of receiver B, C is an "Exposed Node". Node D is in the listening range of B (receiver) but not of A (sender). Hence, D is a "Hidden Node". Before data transmission, A senses the channel to see if it is free. Then A sends an Request-to-Send (RTS) to B. If C hears the RTS, it defers transmission until A can hear B's Clear-to-Send (CTS). If node B is not occupied with any other transmission, it responds back with a CTS. On hearing the CTS, node D defers its transmissions until A finishes sending data to B. When C hears a busy carrier, it defers transmission. After B receives data packet correctly, it sends back an ACK to A. The RTS/CTS handshake (as described above) creates a silence zone in which nodes have to defer their transmissions till data transmission is complete. This is a major reason for low throughput in CSMA/CA based protocols.



Motivation for power control and variable transmission range protocols

Suppose A wants to transmit to B and C wants to transmit to D. In the CSMA/CA MAC implementation, if C is sending to D, then A cannot send toB since B would hear the RTS from C and sense the ongoing transmission. However, if C reduced its transmission power such that it would be sufficient to reach destination node D, then it acquires a lesser floor area. So node A can also proceed with its transmission. Such a protocol closely packs the source-destination pairs in the network , which helps in other simultaneous transmission to proceed and thereby improving spatial channel reuse and hence increasing network capacity. . At any radio receiver, whether a data packet is received correctly or not is decided by the ratio of received Bit energy (Eb)-to-Total Interference per unit bandwidth I0 for the packet at the receiver which is called the Signal to-Interference (SIR) ratio. The SIR at the receiving node j is given by

$$\frac{E_b}{I_0} = \left(\frac{W}{R}\right)\left(\frac{h_{ij}\ P_i^t}{\sum_{k\neq i,j} h_{kj}\ P_k^t + N_0}\right)$$

$$= \left(\frac{W}{R}\right)\left(\frac{\text{Received Transmitted Power}}{\text{Total Interference at Receiver}}\right)$$

where , W is the spreading bandwidth, R is the data rate , hi j is the channel gain between node i and j, Pit is the transmission power of node i and N0 is the background thermal noise. The ratio (W/R) is unity for non-spreading systems. For correct reception, received SIR should be greater than minimum threshold level SIRthres. For any ongoing transmission, the interference tolerance at a receiver j is given by rearranging Eqn as

$$IT_j = \left(\frac{h_{ij}\ P_i^t}{SIR_{thres}}\right) - \left(\sum_{k\neq i,j} h_{kj}\ P_k^t + N_0\right)$$

where , ITj is the interference tolerance of node j. So, even if C is transmitting to D, A is transmitting to B and suppose D is able to listen to transmissions from A, receiver D can still decode the packet correctly if A limits its own transmission power level by considering the interference tolerance of node D. Thus, A makes its decision of transmission power level using two limiting factors: power level should be lower bounded by the value needed to reach destination B, and should be upper bounded by value which does not disrupt the ongoing transmission from C to D. This enhancement in the protocol increases the spatial channel reuse and increase spectral efficiency of the network. The basic principle in designing such a protocol would be to find an efficient mechanism that will ensure the transmission of signals with the variable power in order to satisfy the two basic requirements. First the signal must be able to reach the intended destination and second it should not interfere above the interference tolerance level at all the other nodes in the neighborhood. Based on these two information a transmitter should make a local decision on whether it can progress with the transmission or should defer since commencing its transmission can increase the interference level in the neighborhood and thus could potentially corrupt any ongoing transmission. The most difficult part in this scheme is to design the protocol in a completely distributed and self-organizing environment where there is no coordinator (like a base station in cellular network) to keep track of the interference level at each node and the distance of a receiver from a particular source.

Application of CDMA

CDMA is much more efficient as compared to FDMA or TDMA as it increases network capacity five to six times compared to the other two. Along this the power controlled CDMA gives enhanced system throughput as compared to power controlled protocols in non-spreading systems. This is due to the fact that in a non-spreading case, the W/R ratio is unity (see eqn. 1) (as R equals W here) and Eb/I0 is equal to the ratio of the total signal power received divided by the total interference power received due to the other users in the network and the background thermal noise while in the CDMA, the W/R ratio is much greater than unity because of the spreading of signal . So from Eqn. 1, we can see that the received signal-to-interference ratio is multiplied by the factor of W/R. Since minimum required Eb/I0 ratio for correct reception is typically a fixed number depending on the basis of receiver design, the tolerance level of the receiver increases because the signal power is multiplied by the processing gain. Hence, the receiver is able to decode packets which otherwise would not have been received correctly. This allows more number of simultaneous transmissions in single hop area for adhoc network. Thus power controlled CDMA MAC for adhoc networks combines the advantages of both variable transmission range protocols as well as spread spectrum multi-access communication.

Design aspects

1) Excess bandwidth requirements – Are We Through???

CDMA has additional requirement of larger bandwidth because of signal-spreading. So, CDMA-based network requires more resources in terms of bandwidth as compared to non-spread spectrum based network. Hence, they are beneficial only when the gain offered by CDMA compensates the excess bandwidth requirement. **2) Code Assignment – Needs Mental Perspicacity!!!**

Lack of centralized control makes the code assignment challenging. If the network is small then a fixed code can be assigned to every node. But for large networks there should be some way to assign codes dynamically (ensuring reuse of the codes) such that immediate neighbors get unique codes. **Spreading Protocols**

In spread spectrum adhoc networks, the code assignment scheme determines the code to be used for transmission. There are four basic type of code assignment.

a) Common code: Here all nodes are assigned a common code, and hence all transmissions are done on a single spreading code. In this case, multiple simultaneous transmissions are not feasible.

b) Receiver based: In this scheme, each node is assigned a code. The transmitter uses the code of the intended receiver to spread the packet, while an idle terminal constantly monitors its own code. This approach simplifies the receiver's circuitry because the receiver does not have to monitor the whole code set. Collisions may occur at the receiver, when two transmitters transmit at the same time for the same receiver using the receiver's code.

c) Transmitter based: In a transmitter-based spreading protocol, a transmission code is assigned to each terminal, and receivers must be able to monitor the activity on the whole set of PN codes. The advantage of this approach is that collisions cannot happen at the receiver. In addition, broadcast is inherently supported. However, the drawback is that the receiver circuitry is very complex and expensive.

d) Hybrid Schemes: Various hybrids of the above two approaches are also possible where headers of data packets containing destination and source address can be sent in common code or in receiver code and the data in sender's code. This helps receiver to tune to the particular sender code in next time slot. One of the issues, which can be observed in aforementioned schemes, is that receivers and transmitters required to know the spreading code associated with an another node and this association can be known statically or dynamically. In adhoc network's self-configuring mode dynamic approach is suitable as new nodes are expected to join and leave arbitrarily.

4)orthogonal OR non-orthogonal codes??? Orthogonal codes require perfect synchronization among the nodes- something difficult to realize in a perfectly non-centralized domain. Also, multipath effect in cellular network introduces interference that is not orthogonal to the transmitted signal. So, it is wiser to use non-orthogonal codes with a small cross-correlation factor which would eliminate the need for perfect synchronization. But then non-zero cross correlation of the codes will generate certain amount of the MAI in adjacent receivers.

USE OF ADHOC NETWORKS IN REAL LIFE

Military environments:-

Infrastructure typically break down in military areas. In such areas Adhoc networks are the only rescue. In addition to this adhoc networks provide good networking facilities in case of soldiers, tanks, planes for example.

Civilian environments

There are cases where infrastructure needed is either too small or no infrastructure is needed at all. Eg:-Taxi cab network work with the help of adhoc networks without any infrastructure. Unplanned meetings not relying on pre-infrastructural work, depend on adhoc networks.

Personal area networking

The availability of several mobile computing devices and handheld devices have increased the use of WLANs which have typical fearture of adhoc networks. Egs:-cell phone, laptop, ear phone, small aircraft.

Emergency operations:-

In such cases, due to no time for infrastructure setup, adhoc networks prove to be a great help. Egs.:-search-and-rescue, policing and fire fighting

References

1. Mobile Communications:- Jochen Schiller
2. An Engineering Approach to Computer Networking:- S.Keshav
3. Manish Singh, Krishna Paul and Abhay Karandikar, “On Use of Power Control and CDMA to Increase Spectral efficiency in Adhoc Networks.”
4. Alaa Muqattash and Marwan Krunz, “CDMA-Based MAC Protocol for Wireless Adhoc Networks.”

GLOSSARY

Baseband Signal :- a signal whose frequency response is nonzero only for |f|< B (in strict sense), where B is a positive real number.
Multipath fading :-A type of fading caused by signals taking different paths from the transmitter to the receiver and, consequently, interfering with each other. this causes degradation in the quality of the received signal.
Distributed system :-In distributed system, all nodes have same priority and there is no base node unlike the base station in mobile industry.
MAI :- MAI stands for multiple access interference, this is caused due to interference of non-orthogonal signals propagating in the same medium.
Processing Gain :- The ratio of R to the W is defined as the processing gain , here W is the spreading bandwidth, R is the bandwidth of the signal.
Multi-hop Transmission:-which is directed to the destination via various routers.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude and thanks to our course-instructor Prof. V.M. Gadre for his valuable guidance and encouragement. His teaching of Signals and Systems inspired us throughout our work.

Application of SAS In Optical Mouse

by

Shubham Lahoti (02007017)

Vicky Bhojwani (02007019)

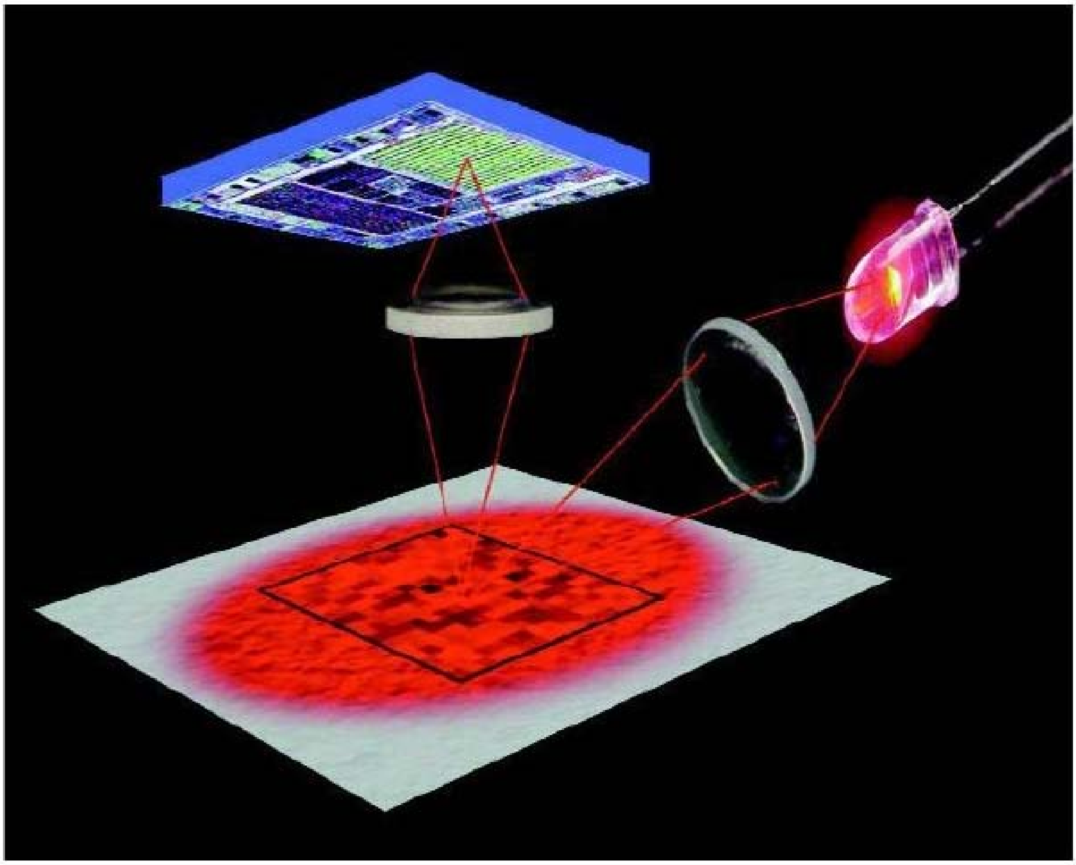
Mukesh Bharti (02007031)

Introduction:

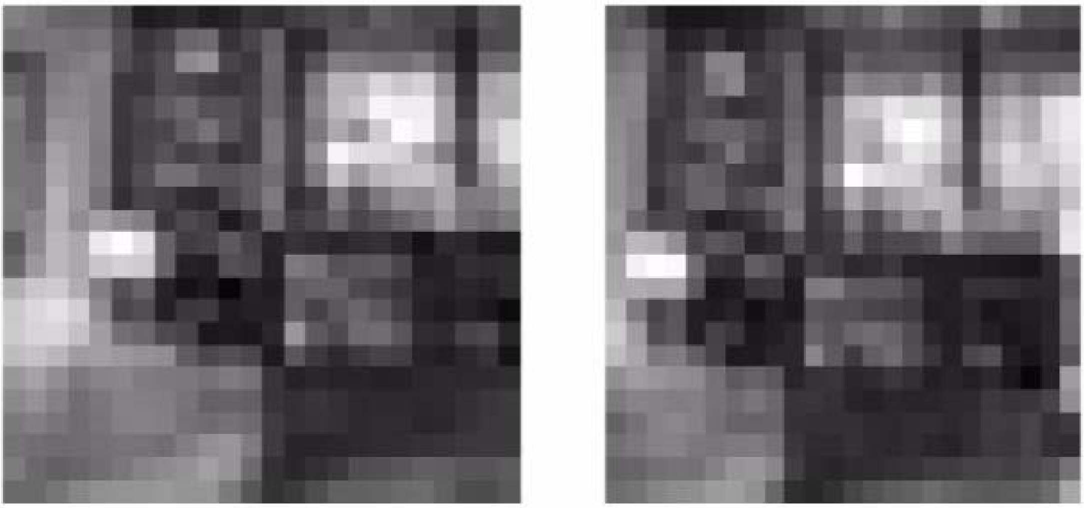
Computer has become a part and parcel of our lives. To start with there was no need to point because computers used crude interfaces like teletype machines or punch cards for data entry. The early text terminals did nothing more than emulate a teletype (using the screen to replace paper), so it was many years (well into the 1960s and early 1970s) before arrow keys were found on most terminals. Full screen editors were the first things to take real advantage of the cursor keys, and they offered humans the first crude way to point. Light pens were used on a variety of machines as a pointing device for many years, and graphics tablets, joy sticks and various other devices were also popular in the 1970s. None of these really took off as the pointing device of choice, however. Need of a simple to use and accurate pointing device lead to the invention of computer mouse (called so because of its resemblance to mouse with its tail).Beginning its life with wheels, mouse was an immediate success. But it was soon realized that the wheel mouse required a lot of maintenance and required a smooth surface for proper operation .Advent of technologies like optical navigation revolutionized the mouse and soon a better version of old wheel mouse "The Optical Mouse" was born.

Working:

Optical mouse has a complete imaging system embedded in it. The mouse is essentially a tiny, high-speed video camera and image processor. Navigation sensors placed underneath the mouse start operating with a slight movement of the mouse. A light-emitting diode (LED) illuminates the surface underneath the mouse. The light from the LED reflects off microscopic textural features of the surface. A plastic lens collects the reflected light and forms an image on the sensor. The image as seen by a naked eye is a black and white picture of a tiny section of the surface. The sensor continuously takes pictures as the mouse moves. The sensor takes pictures quickly -1500 pictures (frames) per second or more -fast enough so that sequential pictures overlap.



Optical mice illuminate an area of the work surface with an LED, to reveal a microscopic pattern of highlights and shadows. These patterns are reflected onto the navigation sensor, which takes pictures at a rate of 1500 images per second or more.The images are then sent to the optical navigation engine (Digital Signal Processor) for processing. The optical navigation engine is the brain of the mouse. It identifies texture or other features in the pictures and tracks their motion. Figure 2 illustrates how this is done.



A. Image at t = 0 ms B. Image at t = 0.67 ms

The Navigation Engine identifies common features in sequential images to determine the direction and amount of mouse movement. Image B was taken while the mouse was moving, a short time after image A. It shows the same features as image A, only shifted down and to the left. Two images were captured sequentially as the mouse was panned to the right and upwards. Much of the same visual material can be recognized in both frames. Through a patented imageprocessing algorithm, the optical navigation engine identifies common features between the two frames and determines the distance between them. This information is then translated into X and Y coordinates to indicate mouse movements. The Central Processing Unit of the computer receives these coordinates and translates the received signals into the motion of the cursor on the monitor.

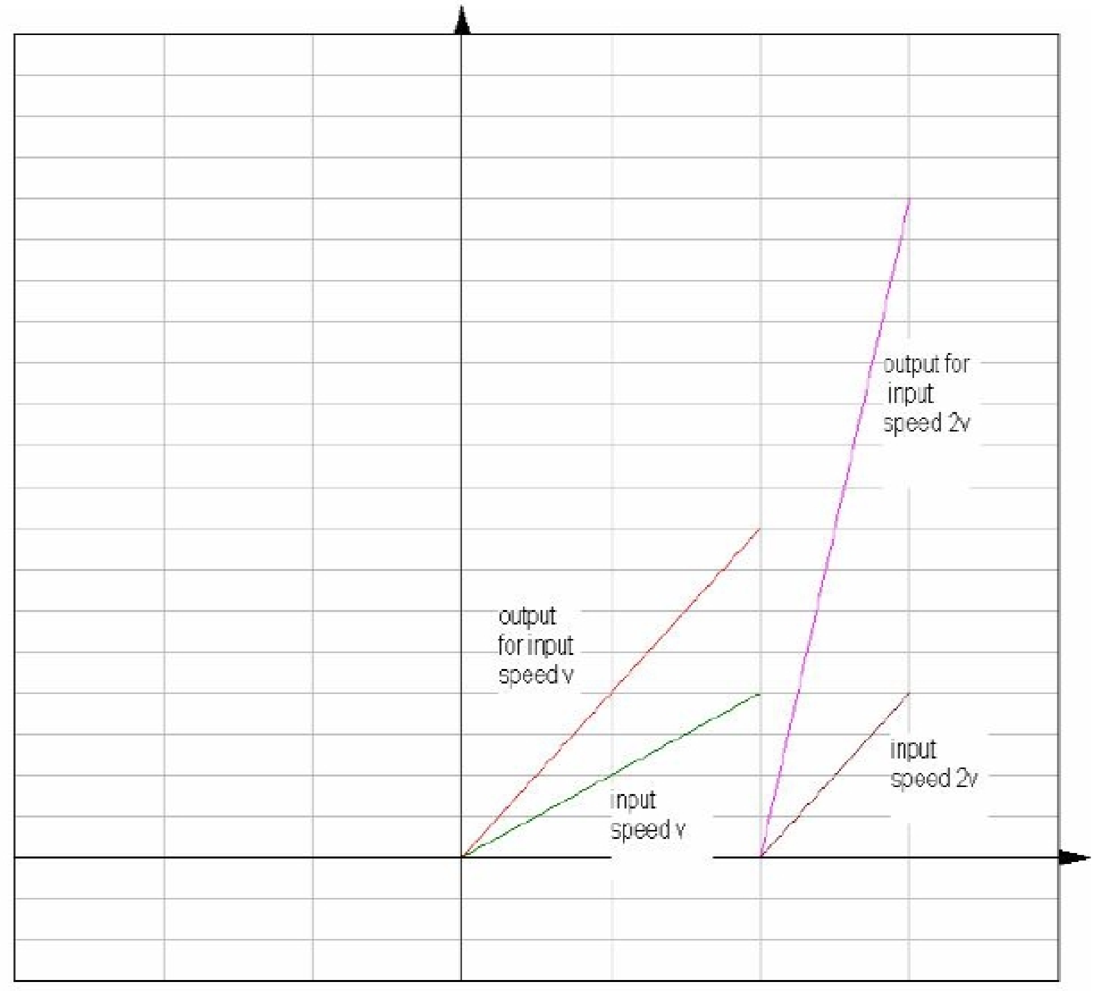
Expectations:

The mouse like any other device is expected to be user friendly i.e. one should be able to take the cursor to any required point on the screen in the least possible time and maximum accuracy with least effort. Now, here we have contradicting properties i.e. for satisfying one of these properties we have to compromise the other. Lets say we increase counts per inch (counts per inch gives the number of measurements taken by the mouse per inch, so increasing counts per inch increases the resolution and decreases the maximum speed at which mouse can be moved) so that we can reach a particular point in the higher accuracy, but since the maximum speed now has been restricted to a lower value time taken to move mouse over larger distances. But if we have low counts per inch then the maximum speed at which mouse can operate is increased but this happens at the cot of resolution. Now we loose the high accuracy we had earlier.

Therefore in the optical mouse the have been adjusted to an optimum level which is the compromise of good speed and high accuracy.

Advantage

Any operator of the mouse takes the cursor to a particular point on the monitor in two basic steps - Taking the cursor in the neighborhood of the point and then- Taking the cursor to the exact location. Here the first operation does not requires a high accuracy hence for this part we move the mouse with a higher speed where as for the second part high accuracy is desired, so in this region we move the mouse slowly. In a mechanical mouse the distance traversed by the cursor on the screen maintains a fixed ratio with the distance moved by the mouse on the surface. Therefore in the case of a mechanical mouse, for the first part of the operation we have to move the mouse over a large distance with a high speed (if the ratio of distance moved by mouse to that moved by the cursor is low), and for second part we have to move the mouse very slowly to obtain high precision (if the ratio of distance moved by mouse to that moved by the cursor is high). Both of the above operations are uncomfortable. Whereas in case of the optical mouse when the mouse is moved fast the cursor moves by larger distance than it would move on moving the mouse slowly between the same two points on the surface(i.e. the speed of the cursor varies non linearly with the speed of mouse).



Shows non linear relation exists between input and output speeds

In the case of optical mouse, for first step of taking the mouse to a given location, one can bring mouse in the neighbourhood of the point by rapidly moving the mouse in the proper direction, (this operation does not requires a high efficiency). So for this part of the operation the cursor moves with much larger speed than the mouse. For the second part, that of taking the cursor to the precise position, we move the mouse slowly, which makes the cursor move slowly to bring it to the desired point. Hence, in case of optical mouse since speed of the cursor depends on the speed of the mouse, the operation of taking the cursor to the required point becomes efficient and easy.

Properties of the system

That was so much for the basic concepts behind the internal working of the optical mouse, but what the user is really concerned with is the mouse as a system and how the cursor on the screen moves with the movement of mouse along a path. So, let us analyze the properties of mouse as a system using the concepts of “signals and systems”. Now we shall consider two very simplified signals which help us determine the basic properties of the system(mouse) for any general input.

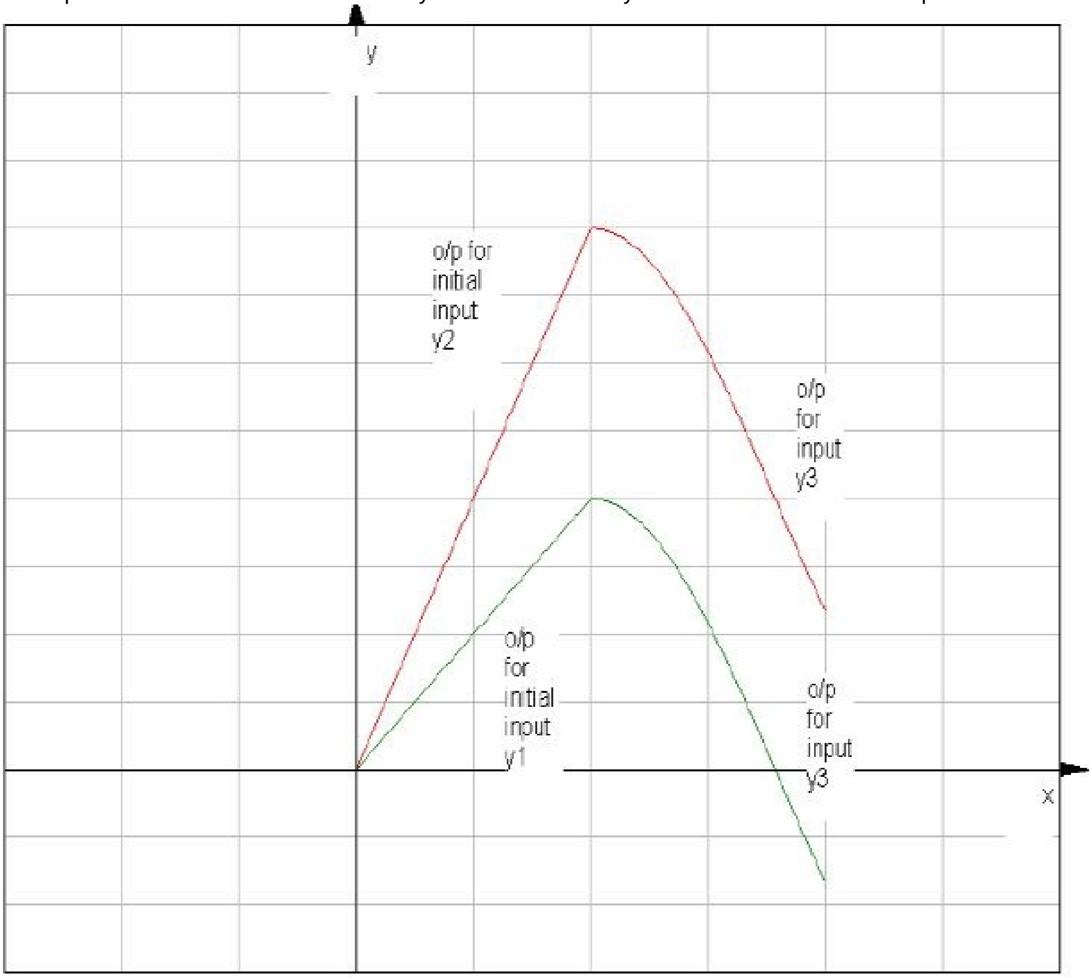
Signal Type A: Consider a simple system in which the optical mouse moves(assume that the mouse is at all the times in contact with the surface) only along a straight line (let us say the x or the horizontal axis) with a constant speed (let speed = v inches /seconds). Attach a set of coordinates to the screen of the monitor with origin at the center, similarly attach a set of coordinates to the surface on which mouse moves. Assume that we start the system at t=0 and at t=0 both, the mouse and the cursor are at the origin. Now for our simple system when mouse moves by a distance d the cursor also moves along the straight line with the same slope in the same direction(x axis in our example) by a distance k*d (where k depends on counts per inch which again depends on the speed of the mouse).

Let us see the properties of the simple system described above:-
Consider the mouse moving along a curve as an input signal to the system, and the output is the cursor moving. Input signal for the system is ux1(t) and the corresponding output is vx1(t) For constant speed we have

vx1(t)= k1* ux1(t)
(where k1 depends on the speed v and is ^{3} 1)
Input signal for the system is ux2(t) and the corresponding output is vx2(t) such that
vx2(t)= k2* ux2(t) (where k2 depends on the speed v and is ^{3} 1)
here k1=k2=k as speeds are same

Memory:

This system can be called memory less in the sense that the shape and direction of the output to a given input does not depend on the past input. But it can also be called a system with memory as the coordinates of the points on the output depend on the previous input.

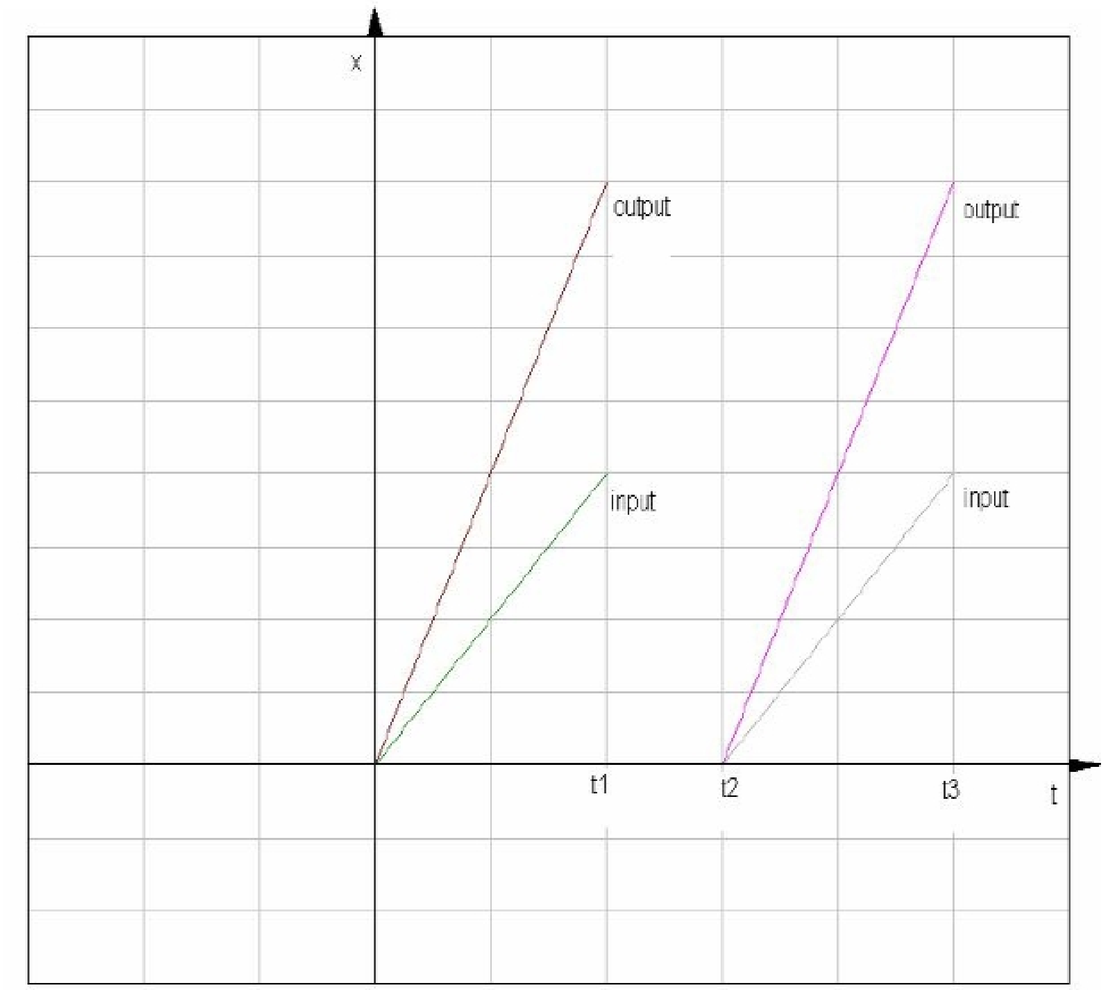


Causality:

The system is causal as the output at a given instant in no way depends on the future input.

Shift Invariant:

This system is shift invariant as when the input is moved on the time axis by t1, the output will also move only on time axis by t1, as output depends only on the input curve and the speed with which it is given. In other words the cursor traces the same curve for a given input irrespective of the time at which the input is given



The input is shifted in time, output remains same, but is shifted in time

Stability:

The system here is stable - The output to any input is always bounded by the boundaries of monitor.

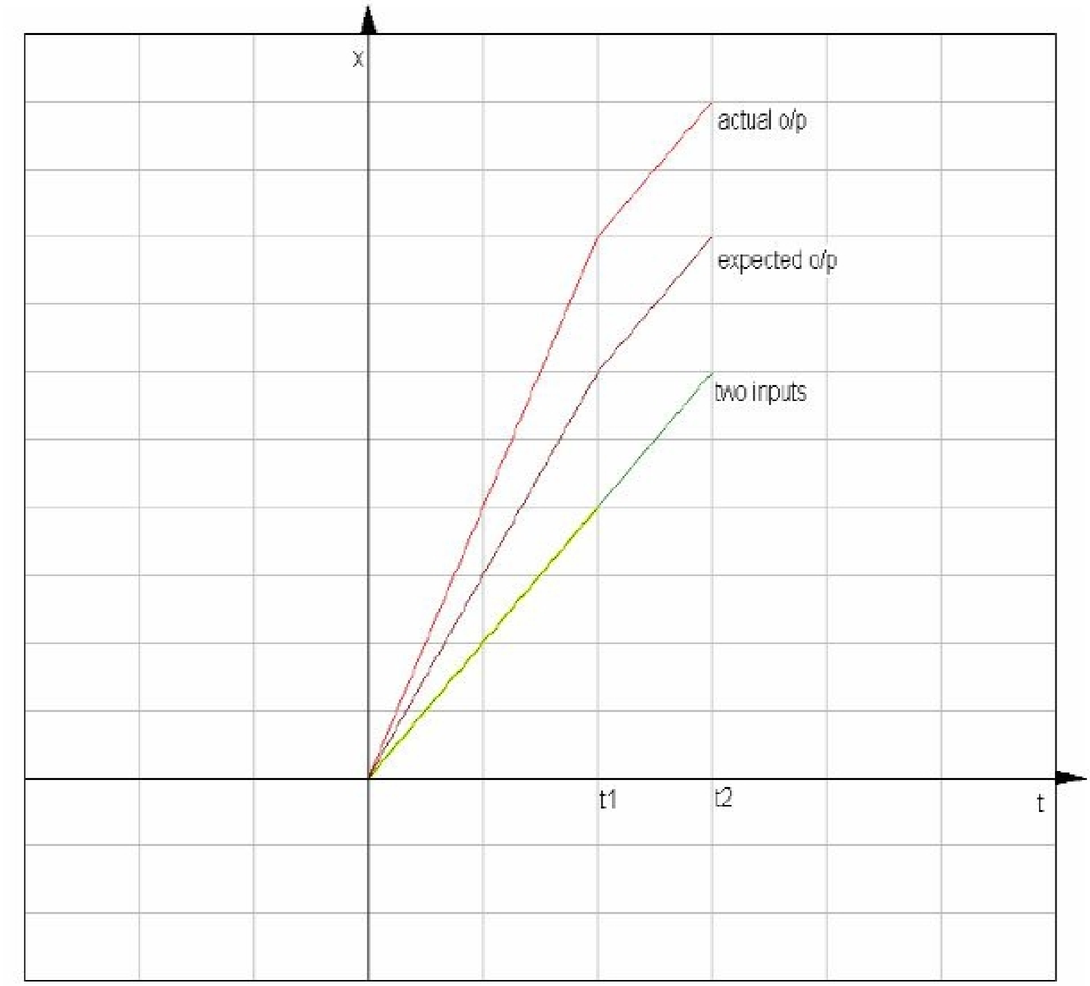
Invertible:

The simple system stated above (input is a straight line with a constant speed = v) is Invertible If we consider all the inputs with same speed then there will be unique output corresponding to a input and for two different inputs the output will not be the same. So the system is one - one. If we define the range to be the set of all the outputs with the cursor moving along a straight line with constant speed, then our system will be onto. So our simple system is one - one onto and hence is invertible.

Linearity:

Additivity:

Now suppose we give the input
ux3(t) = ux1(t) + ux2(t), which gives us the output vx3(t).
ux3(t) = ux1(t) + ux2(t), (0 ux3(t)=ux1(t),(t1£ Under the signal ux3(t) the mouse moves with constant speed of 2v (0
If we increase our domain to the curves which are traced with two different speeds then the input ux3(t) lies in our domain, and now we consider its output vx3(t)=ux3(t)[k3*(u(t)-u(t-t1))+k*(u(t-t1)-u(t-t2))] (where k3 a constant > k)
vx1(t) + vx2(t) = k(ux1(t) + ux2(t))=k*ux3(t)



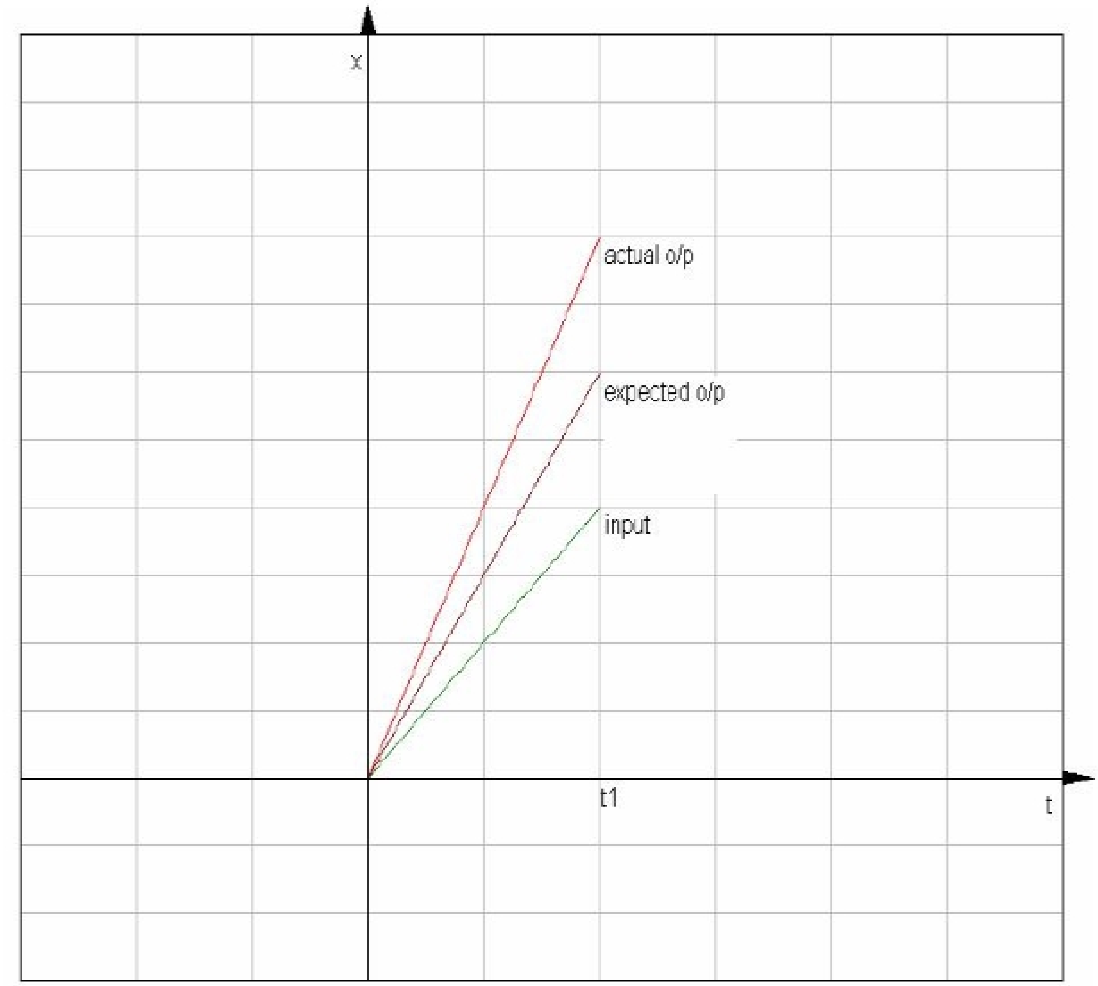
Speed of the cursor varies linearly with the speed of the mouse. So we have $k3 = 2*k$
hence, $vx3(t)=ux3(t)*k*(u(t)-u(t-t1)+u(t)- u(t-t2))$
$vx3(t) \neq k*ux3(t)$

Case II:

Speed of the cursor varies nonlinearly with speed of the mouse. So we have $k3 \neq 2k$
hence, $vx3(t) \neq vx1(t)+vx2(t)$
Therefore the system is not additive. Now, if we define addition of two input signals as one signal followed by another signal ,i.e. we add one signal to the shifted version(the shifted version remains the same as the system is shift invariant) of the other signal such the end point of one signal coincides with starting point of other signal, then the system is additive.

Homogeneity:

Now suppose we give the input $ux3(t) = p * ux1(t)$
Under the signal ux3(t)the mouse moves with constant speed = p*v. So again the curve ux3(t) does not belong to the domain of our simple system as ux3(t) a speed different from v. If we increase our domain to the curves which are traced at any constant speed then the input ux3(t) lies in our domain, and now we consider its output
$vx3(t) = k3*ux3(t)$
$vx3(t) = k3*p*ux1(t)$



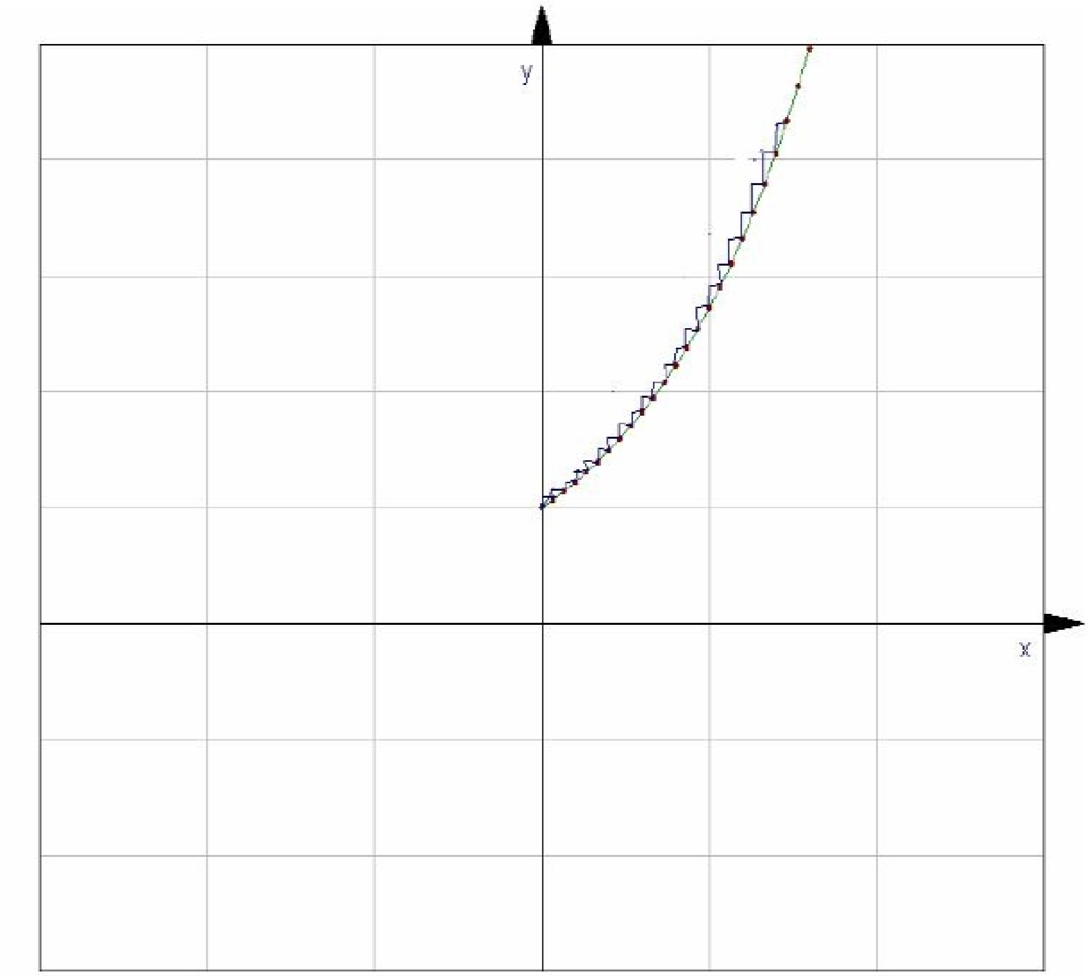
Case I:

Speed of the cursor varies linearly with the speed of the mouse.
So we have k3 = p*k
hence, vx3(t)=ux3(t)*k*p = p*k*p*ux1(t)=p*p*vx1(t)

Case II:

Speed of the cursor varies nonlinearly with speed of the mouse.
so we have k3 ≠ p*k
hence, vx3(t) ≠ p*vx1(t)
therefore system is not homogeneous Hence the system is nether additive nor homogenous. System is non linear.
Similar analysis can be done taking the mouse movement along the y axis, and hence all above properties holds for the movement of the mouse along the y direction.

Any curve traced by the mouse with the constant speed can be approximated as a combination of very large number of small shifted segments of lines parallel to x axis and y axis traversed with constant speed.



The above properties hold for each of these segments, (as our system is shift invariant). So the above analysis holds in general for any curve traced with constant speed.

Signal Type B

Consider a second simple system in which the optical mouse moves (assume that the mouse is at all the times in contact with the surface)only along a straight line (let us say the x or the horizontal axis) and it can move with two different speeds (let, speed1 = v1 inches/second, speed2 = v2 inches /seconds, assume that for any input that mouse can travel either with speed v1 or with speed v2). Attach a set of coordinates to the screen of the monitor with origin at the center, similarly attach a set of coordinates to the surface on which mouse moves. Assume that we start the system at t=0 and at t=0 both, the mouse and the cursor are at the origin. Now for our simple system when mouse moves by a distance d the cursor also moves along the straight line with the same slope in the same direction(x axis in our example) by a distance k*d (where k depends on counts per inch which again depends on the speed of the mouse).

Let us see the properties of the new system described above:- Input signal for the system are ux1(t) and the corresponding output is vx1(t) such that
vx1(t)= k1*ux1(t)
(where k1 depends on the speed v1 and is ³ 1) ux2(t) and the corresponding output is vx2(t)

such that
vx2(t)= k2* ux2(t) (where k2 depends on the speed v2 and is ³ 1)

Memory:

This system, like the previous, can be called memory less in the sense that the shape and direction of the output to a given input does not depend on the past input. But it can also be called a system with memory as the coordinates of the points on the output depend on the previous input.

Causality:

The system is causal as the output at a given instant in no way depends on the future input.

Shift Invariant:

This system is shift invariant as when the input is moved on the time axis by t1, the output will also move only on time axis by t1, as output depends only on the input curve and the speed with which it is given. In other words the cursor traces the same curve for a given input irrespective of the time at which the input is given.

Stability:

The system here is stable - The output to any input is always bounded by the boundaries of monitor.

Invertible:

The system stated above (input is a straight line with a constant speed = v1 or v2) is Invertible For any given output signal traced with a certain speed along a given curve there corresponds a unique input. Hence, our system being one-one onto is invertible.

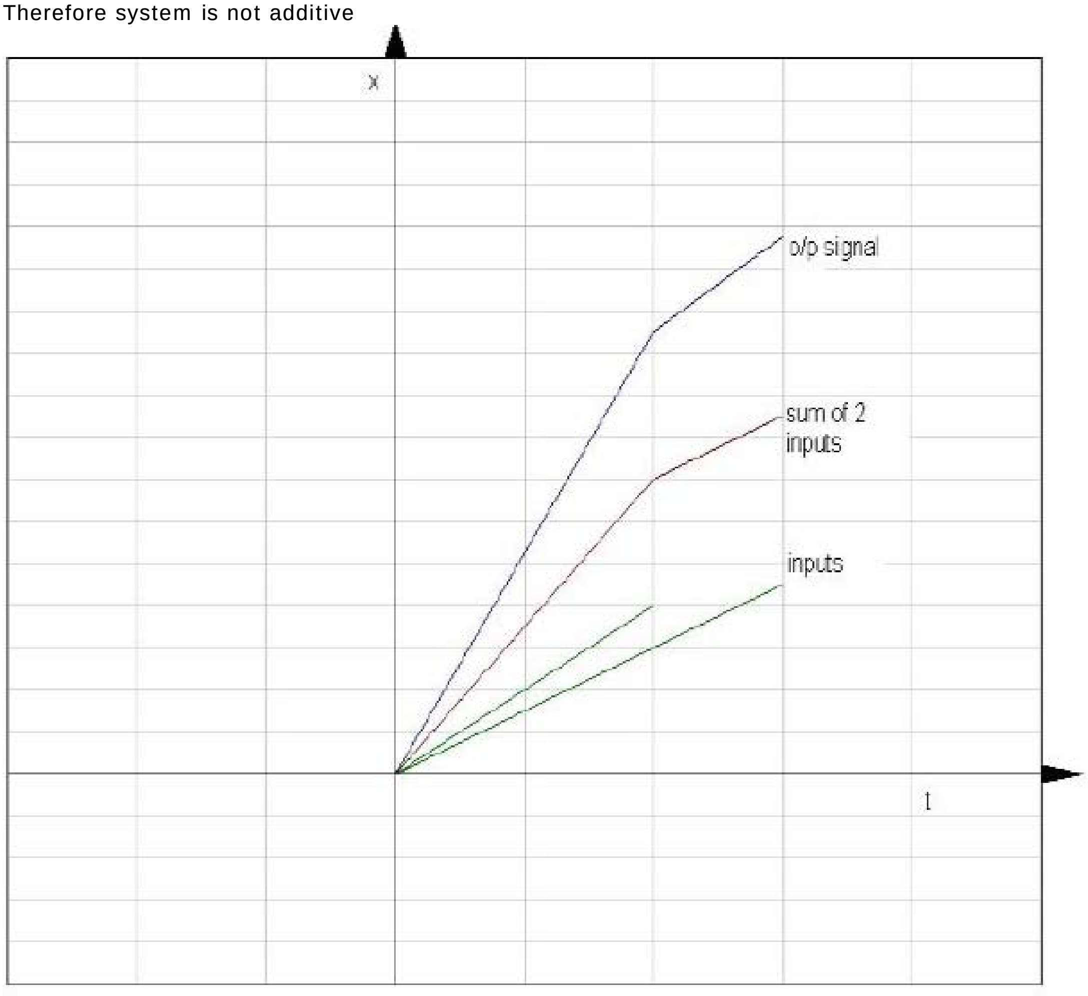
Linearity:

Additivity:
Now suppose we give the input ux3(t) = ux1(t) + ux2(t)
which gives us the output vx3(t).
ux3(t) = ux1(t) + ux2(t),(0 ux3(t) = ux1(t),(t1£t Under the signal ux3(t)the mouse moves with constant speed = v1+v2 (0
system as ux3(t) has changing speed.

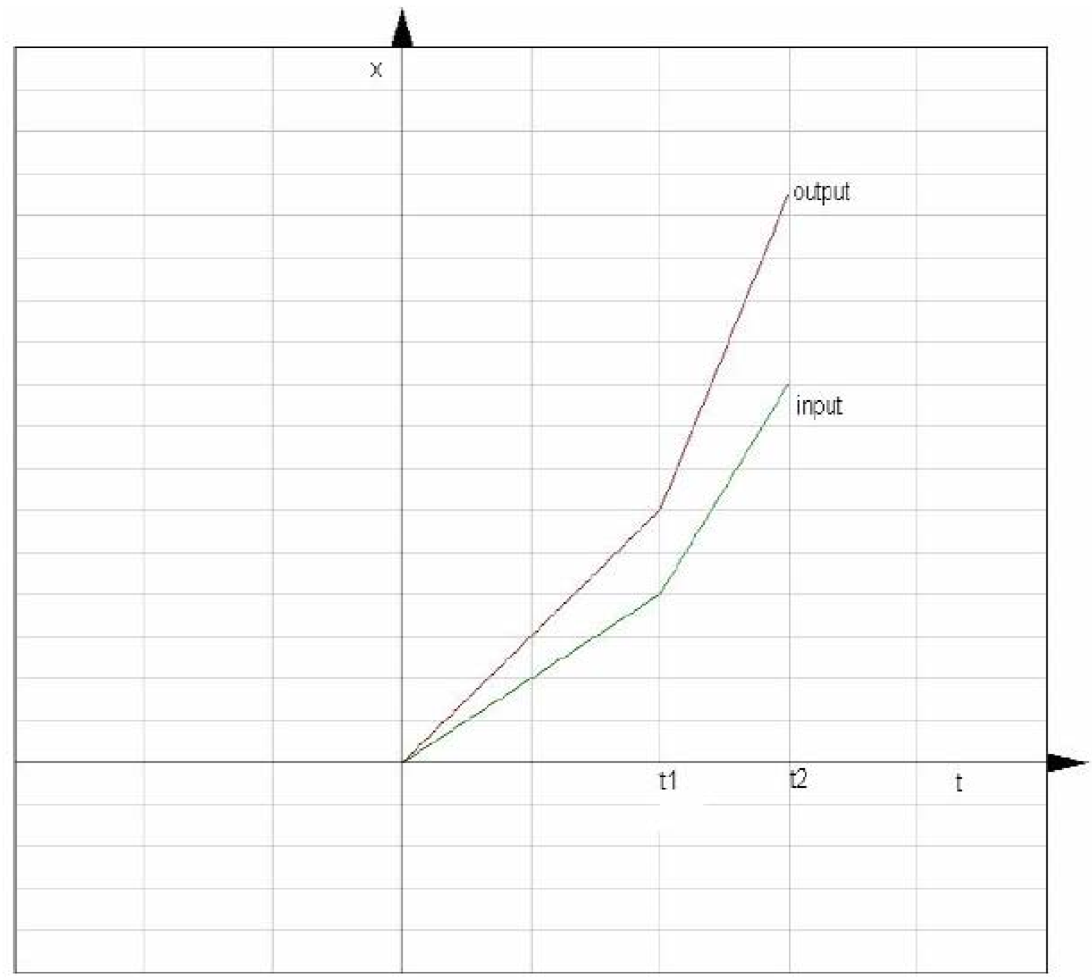
If we increase our domain to the curves which are traced with three different speeds then the input ux3(t) lies in our domain ,and now we consider its output.

vx3(t)=ux3(t)[k3*(u(t)-u(t-t1))+k1*(u(t-t1)-u(t-t2))] (where k3 is a constant > k)
vx1(t) + vx2(t) = (k1*ux1(t) + k2*ux2(t)) ¹ vx3(t)

Therefore system is not additive



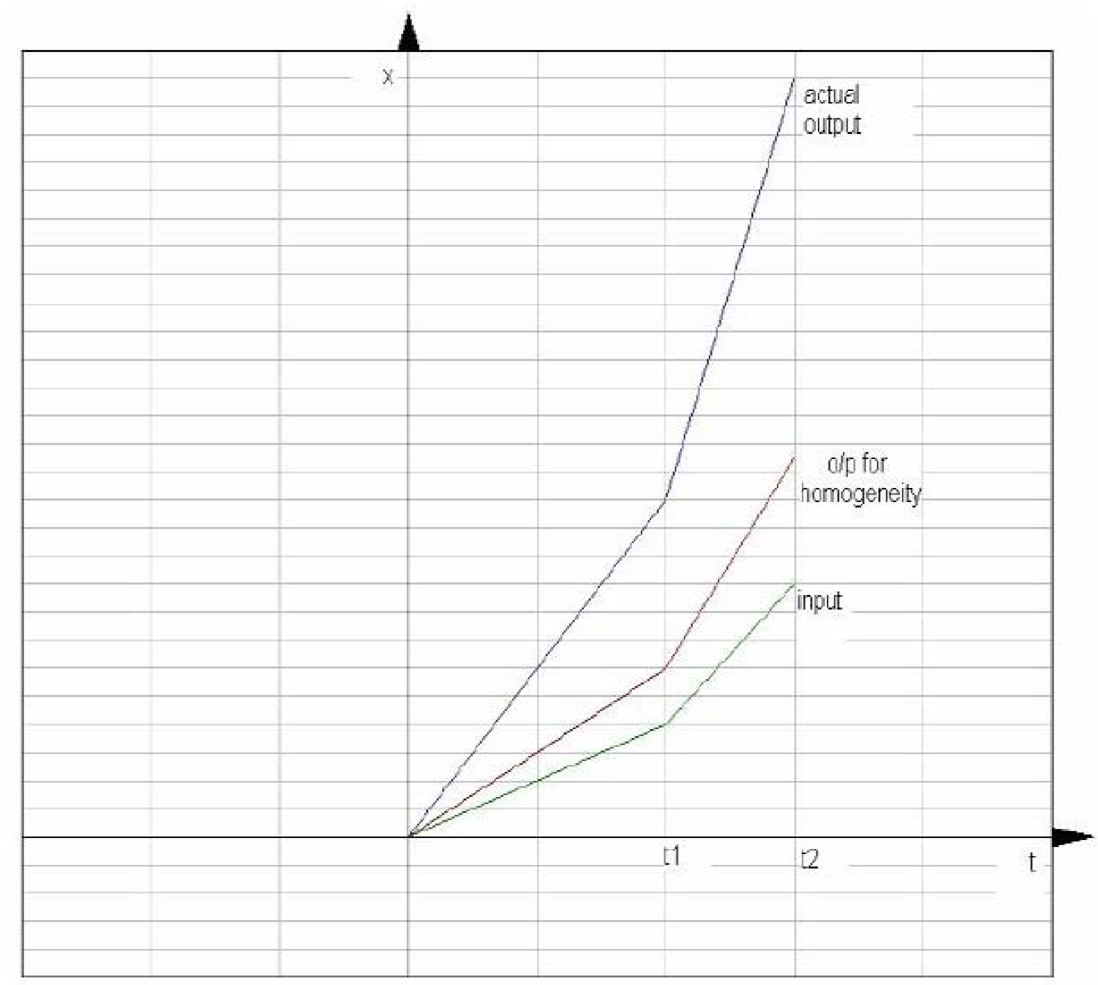
But, if we define addition of two input signals as one signal followed by another signal, i.e. we add one signal to the shifted version (the shifted version remains the same as the system is shift invariant as shown below) of the other signal such the end point of one signal coincides with starting point of other signal, then the system is additive.



Homogeneity:

Now suppose we give the input ux3(t) = p * ux1(t)
Under the signal ux3(t)the mouse moves with constant speed = p*v. So again the curve ux3(t) does not belong to the domain of our simple system as ux3(t) a speed different from v . If we increase our domain to the curves which are traced at any constant speed then the input ux3(t) lies in our domain, and now we consider its output

vx3(t) = k3*ux3(t)
vx3(t) = k3*p*ux1(t)



Case I:

Speed of the cursor varies linearly with the speed of the mouse. So we have
k3 = p* k
hence, vx3(t)=ux3(t)*k*p = p*k*p*ux1(t)=p*p*vx1(t)

Case II:

Speed of the cursor varies nonlinearly with speed of the mouse.
so we have k3 ≠ p*k.

hence, vx3(t) ≠ p*vx1(t)
therefore system is not homogeneous
Hence the system is neither additive nor homogeneous. So, the system in non linear

The above analysis can be applied similarly to the movement along the y axis.
The above analysis can now be generalized over all the permitted speeds for the mouse. Given a general curve which is traced by the mouse at varying speed, can be expressed as the summation of shifted versions of small segments (small displacements along x and y axes) traced with different speeds. Therefore the properties for any general input can be derived from the properties of the two simple cases considered above.

Miscellaneous Properties

What happens if we lift the mouse above the surface (assuming mouse stops working as soon as it loses the contact of the surface)?

Now consider a three dimensional Cartesian coordinate system attached to the surface on which the mouse moves. Now lifting the mouse over the surface means we have a nonzero z coordinate. For this case we can define the system as follows for any non zero value of z the output does not change. Moreover, by moving mouse with nonzero z we shift the entire coordinate system to the point where z again becomes zero(the x, y coordinates in the shifted coordinate system of the point where z again becomes zero are same as those coordinates where the mouse last time had z=0). What happens when the cursor reaches its bound?

The cursor on reaching any of its bound stops responding to any motion of the mouse trying to take cursor beyond the bound, i.e. there is no output to any component of the input signal perpendicular to the bound at the boundary.

Examples & Conclusion

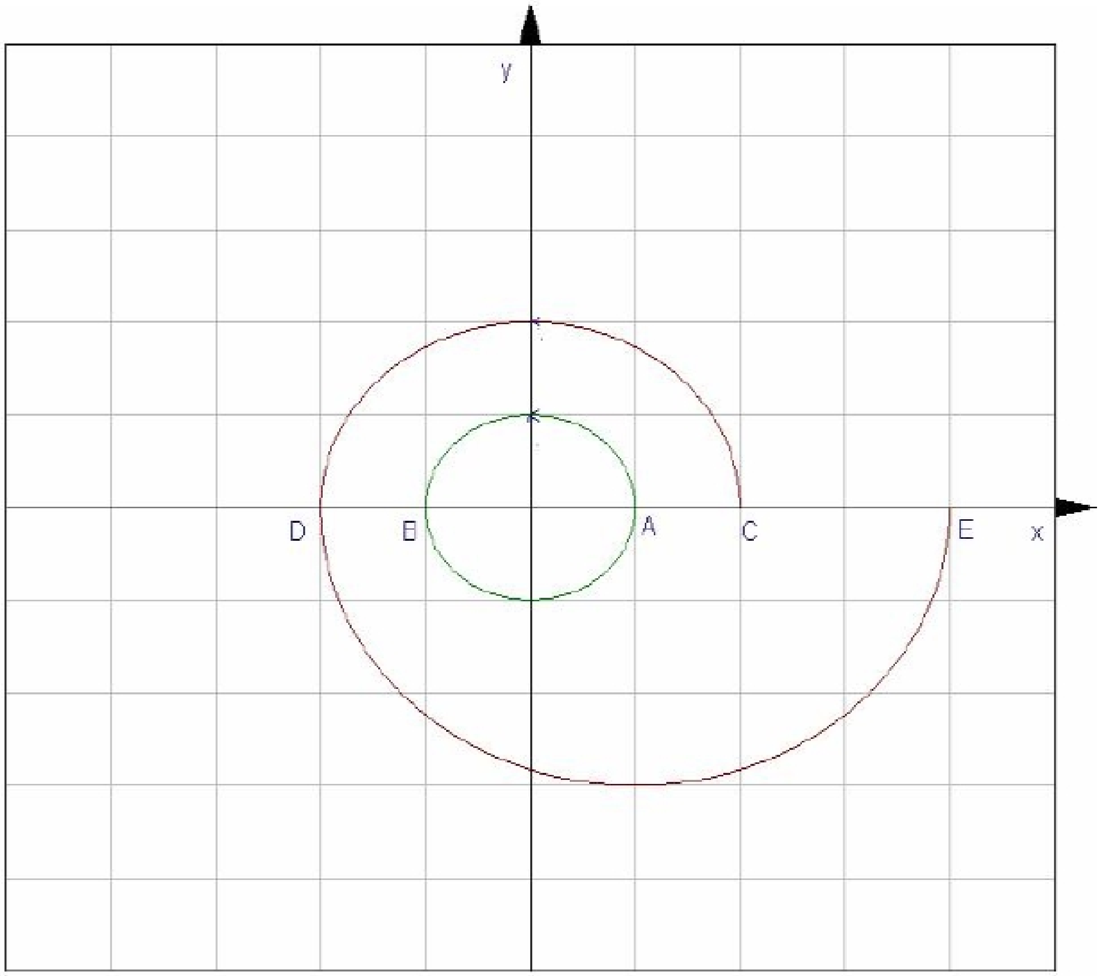


Fig. From A to B speed of mouse is v1 and from B to A speed of mouse is v2 (>v1) Corresponding movement of cursor will be from C to D and D to E.

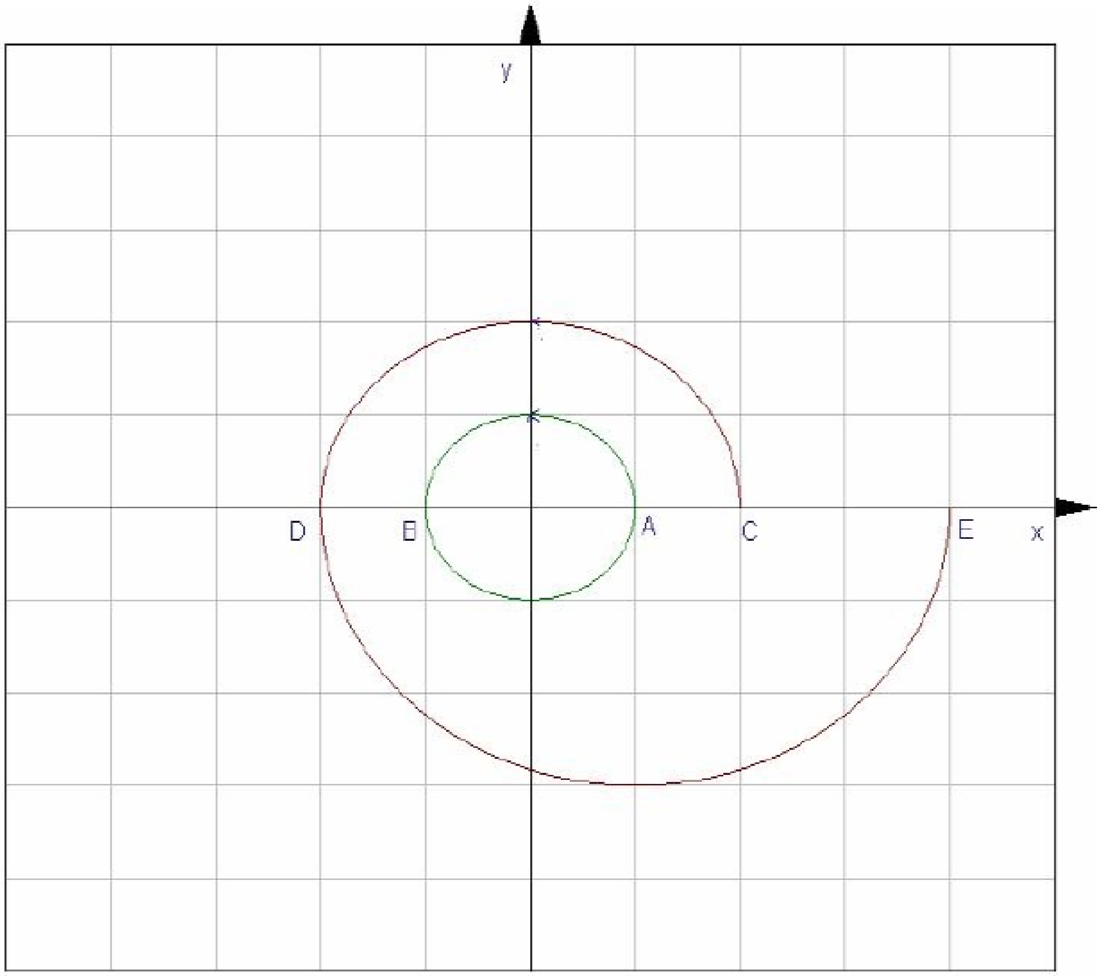


Fig. An input with two different speeds is given. Output as shown in figure.

Conclusion

So, using the concepts of signals and systems we have analyzed the system of optical mouse, and now we can easily tell the properties of the system for a general input. This is one for the many ways the signals and systems are applied in the real world around us.

ENCRYPTION / DECRYPTION SYSTEMS

by

KAUSTUBH JOSHI (02007018)

VIKRANT KULKARNI (02007021)

GOURAV DHAVALE (02D07009)

INTRODUCTION:

Communication is one of the most important aspects of human behaviour. The ability to express our thoughts in a written form is a distinguishing feature of human communication. The need naturally arises to conceal sensitive information from illegitimate eyes. This is where the science of cryptography steps in. Cryptography is the practice and study of encryption and decryption – encoding data so that it can be understood only by specific individuals. A system for encrypting and/or decrypting is a cryptosystem. These usually involve an algorithm for combining the original data (plain text) with one or more keys – numbers or strings of characters known by the sender and/or recipient. The resulting output is called the cipher text. We now proceed to examine this class of objects from a signals and systems point of view. A cryptographic system is defined by an algorithm concerning the cryptographic system.

This system takes string characters (say decimal numbers) as input signal. These characters are taken from a set of finite cardinality. The output signal is the corresponding encrypted string (or decrypted string in case of decryption systems) that need not have same number of characters as the input and whose domain of characters need not be the same. The independent variable for both input & output signals is position of the character in the corresponding string. The characters represent weights attached to the position.

A cryptosystem must be invertible that is, an inverse system must exist for it being practically useful. For, what use is an encoding system when the cipher text can not be decoded to get back the original message? An important issue that comes up is that a third party must not be able to decode an encoded message.

This brings in the notion of security of the system that is the degree of its vulnerability to unauthorized attacks.

While examining an algorithm for security we must take into account the fact that a potential hacker has complete knowledge of the encryption system being used and only does not know the key being used. This is the statement of the Kerckhoff's Principle and usually is an extreme assumption.

When dealing with cryptographic systems it is important to note that the system is represented by an algorithm which typically requires a string ('key') as a priori input.

Thus, each algorithm is actually a collection of 'm' systems (here systems is in the context of this course) where 'm' is the size of the key space of the algorithm.

Consider the simplest of the ciphers – the Caesar (or shift) cipher. Here let the input string be $m_0m_1m_2\ldots m_p$. The algorithm shifts each letter of the alphabet by a fixed number of positions (k) forward. This fixed number is the key. This is equivalent to adding the key to each character of the input string.

That is, if the output is $n_0n_1n_2\ldots n_p$ then $n_i = m_i \oplus_{26} k$.

We realize that this number k can be any number between 0 and 25. Thus the key space cardinality of the Caesar cipher is 26. The Caesar cipher actually corresponds to 26 different cryptographic systems.

Shift invariance in a "Signals and Systems" point of view has no practical significance in cryptosystems. For the sake of completion of the discussion of the properties, it is discussed wherever relevant.

SYMMETRIC AND ASYMMETRIC SYSTEMS

The set of plain texts is P; the set of cipher texts is C. We have key dependent encryption functions $E_k(\bullet): P \rightarrow C$ where K is a key, $K \in \mathcal{K}$ (where $\mathcal{K}$ is the key space) that produce the cipher text $E_k(M)$ of a plain text M. In order to recover the plain text, we require a family of key-dependent decryption functions $D_k(\bullet): C \rightarrow P$ such that $D_k(E_k(M)) = M$ for all $M \in P$ and all keys $K \in \mathcal{K}$.

We call a cryptosystem symmetric if the algorithms $E_k(\bullet)$ and $D_k(\bullet)$ are essentially similar. Symmetric algorithms allow us to derive $D_k(\bullet)$ from $E_k(\bullet)$ in a clearly defined and quickly computed manner. Recalling the Caesar cipher, $n_i = m_i \oplus_{26} k$ we see that to decrypt $n_0n_1n_2\ldots n_p$ we need to apply the algorithm $D_k(\bullet): m_i = n_i \oplus_{26} (-k)$.

This presents a rather difficult prospect for symmetric algorithms. If due to implementation constraints the run time key and the source code of an encryption algorithm are stored on a medium, it can be easily decrypted by competent users. Historically this problem of implementation has caused symmetric algorithms to be vulnerable to attacks.

On the other hand, it is very difficult to derive $D_k(\bullet)$ from $E_k(\bullet)$ for asymmetric algorithms like the RSA. Thus the security of the system

depends on the security and protection of their keys.

Vernam Cipher

Here we analyze the working of the encryption algorithm of the Vernam Cipher (also known as One Time Pad) in the realm of Signals and Systems. It is the only known cipher in the world that can keep messages secret no matter how long an adversary attacks, and no matter what machinery the adversary has.

System Formulation:

For a string of m numbers, a string of m random numbers is generated using a key r which is "large prime number". Here the term "large" is in a sense that it should have as many bits as the message to be transmitted has.

Encrypted output E (i)= (x (i) + k (i))%26

x(i) = Number at the ith position in input string k(i) = Corresponding random number generated

Hence 'm' random numbers + 'm' meaningful numbers give rise to set of m numbers which form the encrypted message. Decrypted output D(i)= (x (i)- k(i))%26

The position of the number in the string is the independent variable for both input and output signals. We define following operations on the signals:

Addition: Defined and meaningful if and only if length of the two strings being added is the same.

Strings a={a(i)} b={b(i)}

Then a+b={[a(i)+b(i)]%26}

Scaling: ca={ca(i)}

Now the properties of Vernam Cipher can be discussed as follows:

1) Linearity:

The system is highly non-linear as it is neither additive nor homogeneous.

This can be shown as follows:

$$E_{a+b}(i) = \{[a(i)+b(i)]\%26 + k(i)\}\%26$$
$$\neq \{a(i)+k(i)\}\%26 + \{b(i)+k(i)\}\%26$$

Similarly

$$E_{ca}(i) = \{[ca(i)]\%26 + k(i)\}\%26$$
$$\neq c\{a(i)+k(i)\}\%26$$

2) Memory: The system is memoryless as the encryption of each character is independent of the previous or next characters.

3) Stability: The system is stable .Here stability is examined in the sense that for a finite input character, the encrypted character in the output is also finite.

4) Causality: The system is non-causal because if a signal is shifted in terms of its position of characters the corresponding output signal is not the shifted version of original output as the random number generated are characteristic of positions.

The difficulty with the Vernam Cipher is its requirement of very large prime number. But this difficulty also renders almost unbreakability to Vernam Cipher. Hence it is used in securing hotline communication between Moscow & Washington.

Hill Cipher

This cipher was used for the encryption of radio call signs in World War Two. Fix a block size N and choose an invertible N ´N matrix K with entries in Z+%26

This matrix K is the key. We rewrite a block of N characters as an $N \times 1$ matrix.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

The encryption key, $E_k(y) = Ky$

The i[th] character will be encoded as

$$E_k(y)_i = \sum_{j=1}^{N} K_{ij} y_j$$

Now, K is invertible. Hence the message can be decrypted by multiplying $E_k(y)$ by K^{-1} since $D_k(E_k(y)) = K^{-1}.(K.y) = y$.

Now the input signal might consist of series of elements such as 'y'. Then output will also be series of vectors like 'ky'. The independent variable for both input & output signals will be position of particular vector in the series.

Properties:

Additivity:

The Hill cipher is additive as per our definition of addition.

Consider two strings $a = \{a(i)\}$ and $b = \{b(i)\}$ where a and b are of equal length N. Let a_k and b_k be the corresponding numbers of the alphabet. Rewriting a and b as $N \times 1$ matrices we get

$$E_k(x) = Kx$$
$$E_k(y) = Ky$$
$$E_k(x+y) = K(x+y) = Kx + Ky = E_k(x) + E_k(y)$$

Homogenity: The Hill cipher is homogeneous.

Ek(x)=Kx

Ek(cx)=K(cx)=cKx=cEk(x)

Linearity: As Hill cipher is homogeneous and additive, it is linear.

Memory: The definition of memory from the time domain is not applicable here.

However we can say that the system has memory in the sense that the encryption of each character is dependent of the other characters because

$$E_k(y)_l = \sum_{j=1}^{N} K_{lj} y_j$$

Stability: String length at the output = String length of the input. For a finite string as input the output is also a finite string of the same length.

Shift Invariance: The Hill Cipher system is Shift – Invariant. This can be shown as follows :- Consider input signal, x ={x(i)} , series of vectors . Let the corresponding output signal be y ={y(i)} .

Then y(i) = Kx(i) .

Now if x(I)->x(i + i0)

Then corresponding y(I) = kx(I) = kx(i + i0) = y(i + i0).

Causality: The system is indeed causal .The output can not be nonzero before the input deviates from zero.

RSA ENCRYPTION METHODOLOGY

The RSA (named after its inventors R.Rivest, A. Shamir and L.Adelman) is generally used in communication applications where the communication channel is generally unsafe. For such applications, RSA is more efficient than DES in two respects:

1) Like DES, network users don't have to share a common key.

2) Dynamic generation of keys for each communication session is not required. Hence RSA is more time and computation efficient.

System Formulation:

RSA cryptosystem is characterized by generation of two “very large” prime numbers p and q (which have typically 512 bits each) which are known only to the user A .Then mathematically the encryption system can be defined as:

$$P_A(M) = M^e \bmod n$$

where

M = message made up of input alphabets

$n = p \times q$

$e =$ odd integer that has no common factor with $p-1$ and $q-1$

Here, the input signal M is supposed to be a series of decimal numbers where m(i) and i is the position of a particular number in the input series. The output signal is PA(M) (known as Public Key) which is also a series of decimal numbers having same length as that of input message. For this system, we define following things:

M1 +M2 ={m1(i)+m2(i)}

{Defined and meaningful if and only if length of the two strings being added is the same} and cM ={cm(i)} for cbelongs to R

Then the properties of the RSA encrypting algorithm can be discussed as follows:

1) Linearity:

We can clearly see that

$$P_A(M_1)+P_A(M_2) \neq P_A(M_1+M_2)$$

Hence the system does not possess the additivity property.

Also

$$P_A(cM) = c^e M^e \bmod n$$
$$\neq c P_A(M)$$

which means the system is not homogeneous.

Clearly, RSA algorithm is non-linear.

Stability of RSA system shall be discussed in terms of whether a series of bounded decimal numbers gives bounded series of decimal numbers as output. From the relation describing the system, it can be noted that for finite mi and e , the corresponding output e mod mi n is finite . Hence RSA algorithm is stable system.

3) Memory: The system is necessarily memoryless as any number in the output string depends only on the corresponding position in the input.

4) Causality: The system equation indicates that as long as input mi does not become nonzero the output can not become nonzero. Hence the RSA cryptosystem is causal.

Note that here; causality is with respect to position of the number in the input string, which is the independent variable in this case. Now we come back to the point of non-breakability of the Vernam Cipher. As explained above, to break into Vernam encrypting, one must find the sequence

RSA DECRYPTION METHODOLOGY

RSA Decryption algorithm is classified as asymmetric cryptosystem because the encryption algorithm and decryption algorithm are not same. System Formulation:

The Decryption algorithm uses same two prime numbers p and q but varies in system equation as follows:

$$S_A(M) = M^d \bmod n$$

where the new parameter d is calculated such that $d \times e \equiv 1 \bmod \{(p-1)(q-1)\}$.

The input for this system is encrypted message M, a string of decimal numbers. The output $S_A(M)$ (known as Secret Key) is again string of same length.

The properties of the both encryption and decryption systems are same as the relations between input and output signals are of the same nature. Cascading of these systems as shown forms the complete “secure communication system”: SA(PA(M)) =M

References

1) Secure Communication Systems

Design, Analysis , and Implementation

Michael R.A. Huth

2) Making,Breaking Codes

An Introduction To Cryptology

Paul Garret

Sampling Theorem

by

Siddharth Chaudhari (02007022)

Nishant Jain (02007024)

Abstract:

1. Initially we describe the system which is a day-to-day application of the Signals and systems concepts and mention what the inputs and outputs of the system is.

2. Next, we describe what are the different stages the input undergoes and what changes undergo the raw input data till the output stage is reached. 3. We then introduce some of the concepts used in the algorithms used to further compress this audio data to smaller sizes. Some of them being the algorithms used for creating mp3,wma,qt files each having their own advantages and unique features.

4. An argument as to why digital audio is better than its analog counterpart and how it has become the inevitable choice follows.

5. Lastly mention the properties of the system with respect to one of the inputs, the audio input. (variation of pressure with time)

Introduction:

Sampling theorem is one of the very basic theorems in the field of digital processing and communication which has gained increasing importance of late because of its many advantages over its analog counterpart. Sampling refers to picking out values of the signal for certain values of the independent variables. It is of utmost importance that this sampling produces a signal from which we can get back the original signal. This puts some constraints on the input signal and this brings in the Sampling Theorem.

Simply put the Sampling Theorem says that when sampling an analog signal the sampling frequency must be greater than twice the highest frequency component of the analog signal to be able to reconstruct the original signal from the output signal completely. Indeed most signals in nature are of analog form. But computing devices can handle only digital signals. Digital signals also occupy less space on a storage device. So digital systems are advantageous from a mechanical point of view. This requirement of converting analog to digital signals and vice versa is fulfilled by the Sampling Theorem.

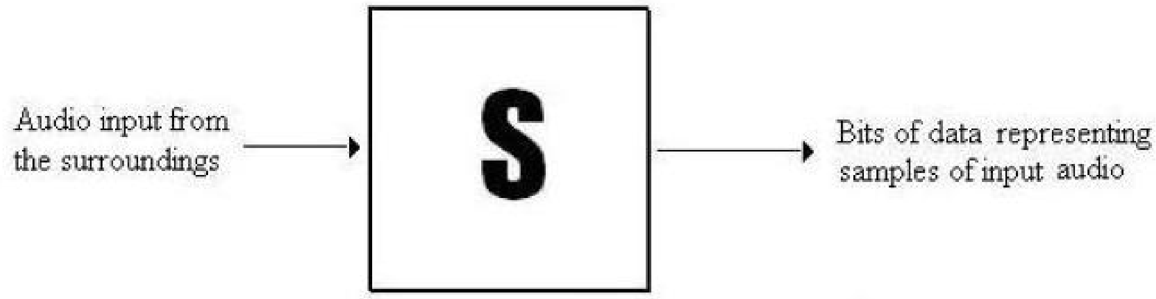
A common example is that of the video recording process. The motion is captured by a video camera (camcorder) which picks out certain instances when to take the frame. The film is then played back at such a rate that for the eyes the sequence of discrete images assumes the form of continuous motion. This theorem has a much higher potential than this simple example with its applications ranging from fields like wireless communication to scanning and photography i.e. wherever analog to digital conversion of signals is necessary. This theorem has proved its importance in completely diverse fields and we propose to examine the application of this theorem in more detail from a Signals and Systems point of view.

PCM recorder

Analog audio signals to a stream of digits representing the analog audio (digital audio). It is an Analog to Digital converter of the most basic kind The conversion of an analog signal to a digital signal is a theoretically simple. What happens is that the analog waveform is sampled periodically, and the samples are digitized (converted into their binary equivalent) one after the other.

Input : Analog audio(variation in the pressure of surrounding air with time), Sampling rate and Bit-depth(sampling resolution)(most commonly used is 44.1kHz 16bit for audio cd which can be changed as per requirements, for example some telephones use 10kHz sampling whereas)

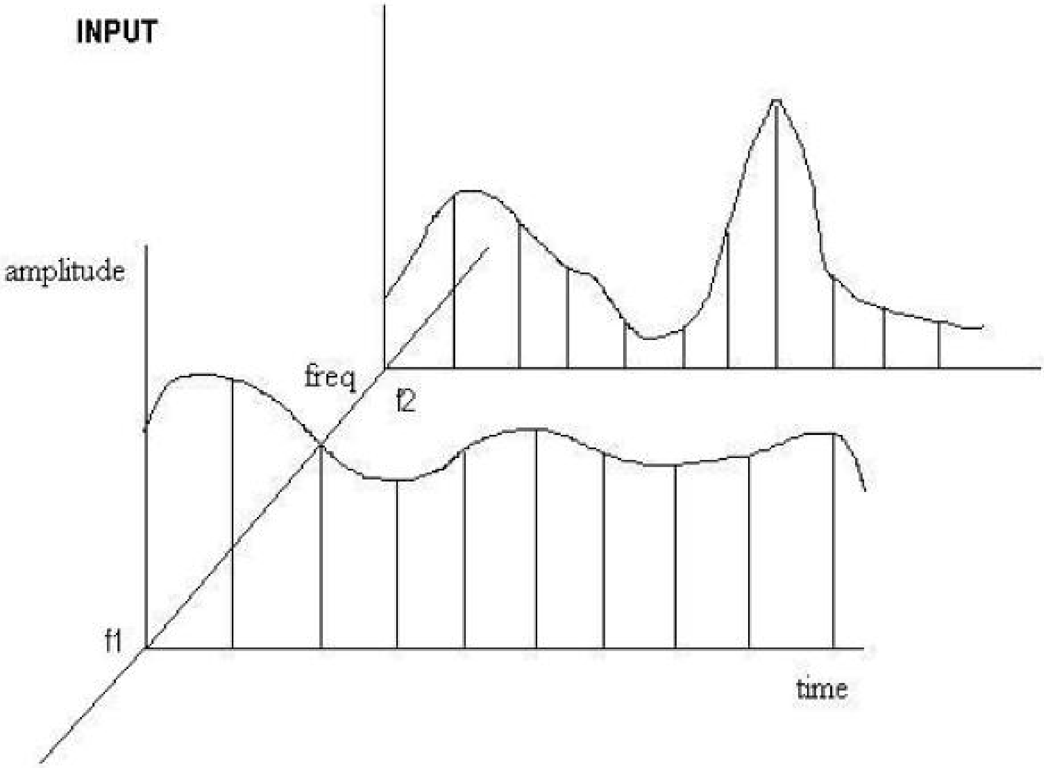
Output : Digital signals representing audio (music files typically .wav Microsoft .aiff Apple Mac) These are the inputs for further compression using algorithms like those used for .mp3, .au, .qt, .wma .ra etc)



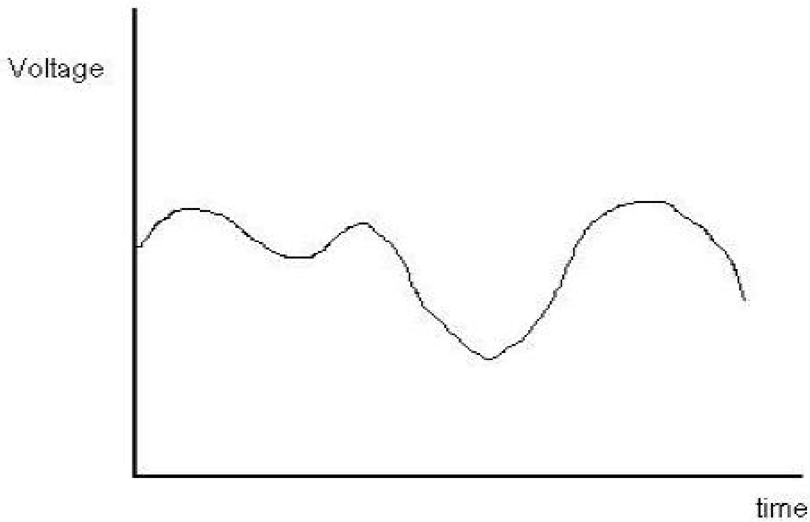
Explanation of how it works…

First stage : Microphone changes audio that is the pressure it senses into representative analog voltage waveforms of suitable value thus acting like a system which passes on the information of the audio heard to the sampling device (system).Note that we assume that the output of the microphone (the voltage waveform) varies linearly with the air pressure.

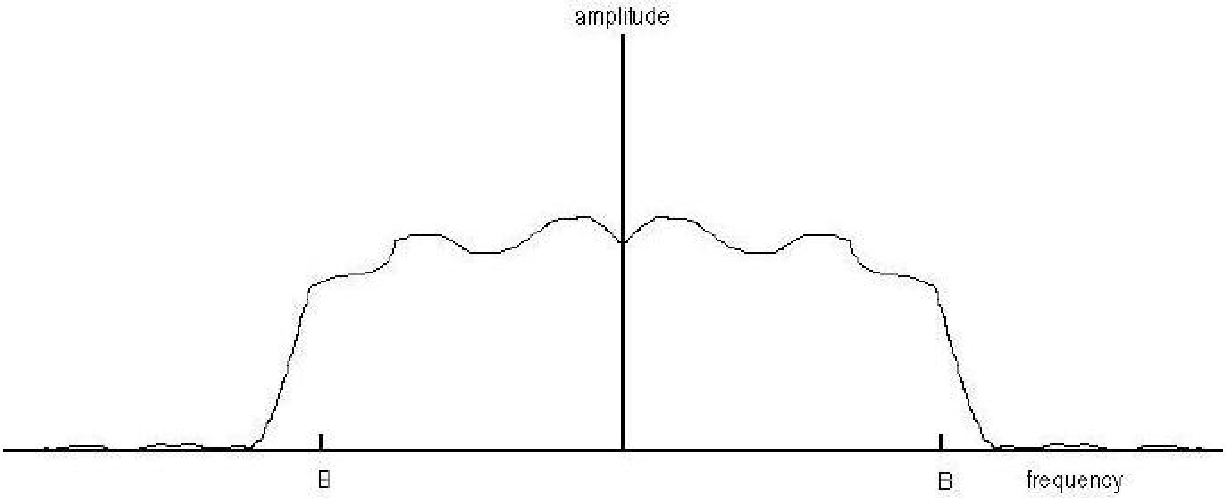
Thus an example of the input to the microphone can be represented like this.



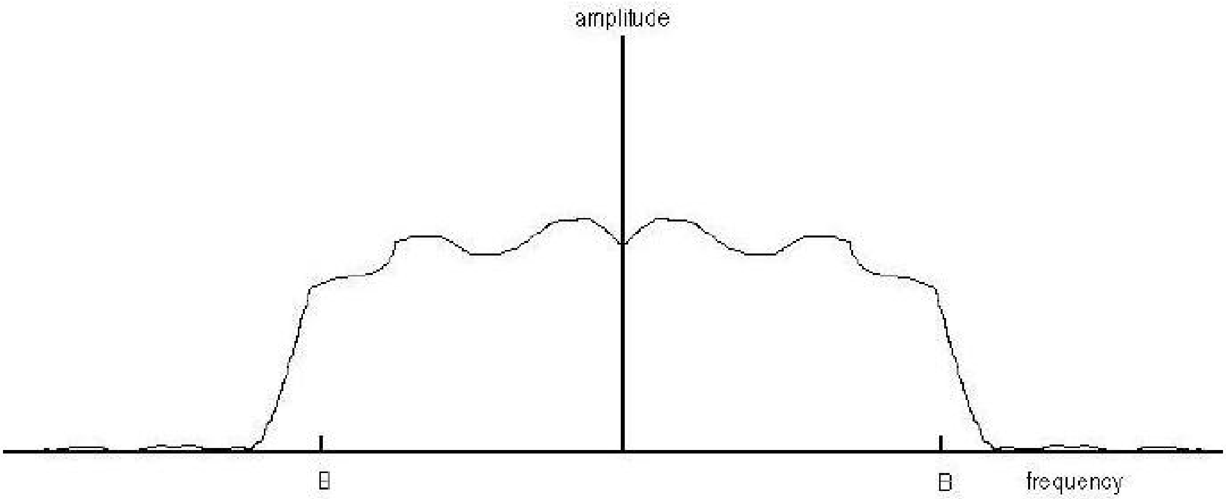
And the corresponding output of the microphone would be as shown below.



Second stage : A low pass filter (ideal one assumed) allows only frequencies of the audible range though so that no aliasing effects take place due to any unwanted higher frequencies) In practice a complex ‘brick wall’ filter with sharp ‘edges’ is used to filter the input analog video. The Fourier transform of this signal after passing through the filter is something like the figure shown below.



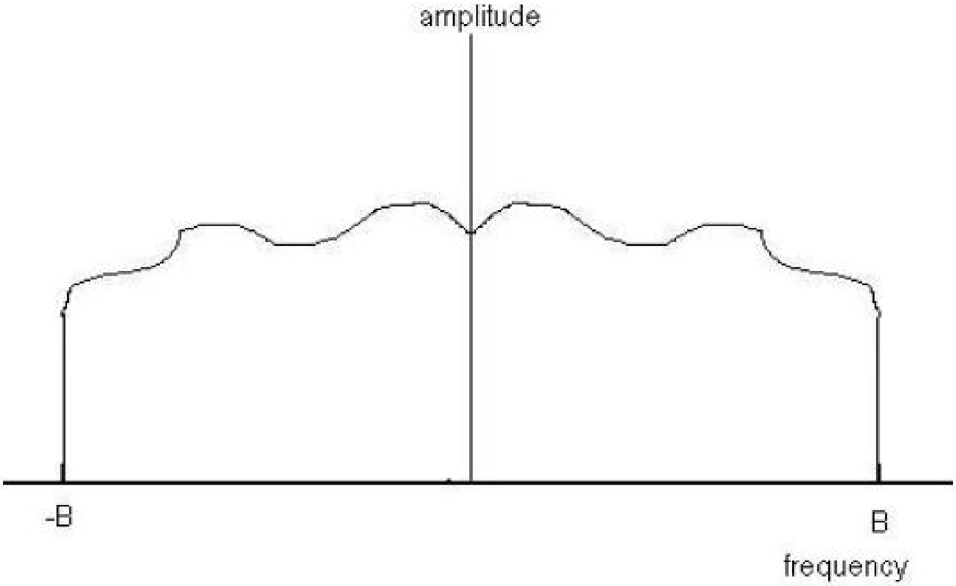
The analog waveform which now is to be sampled (The Voltage waveform) which is the output is shown below (Note this is not similar to the one shown above; it is a different example)(Voltage versus time)



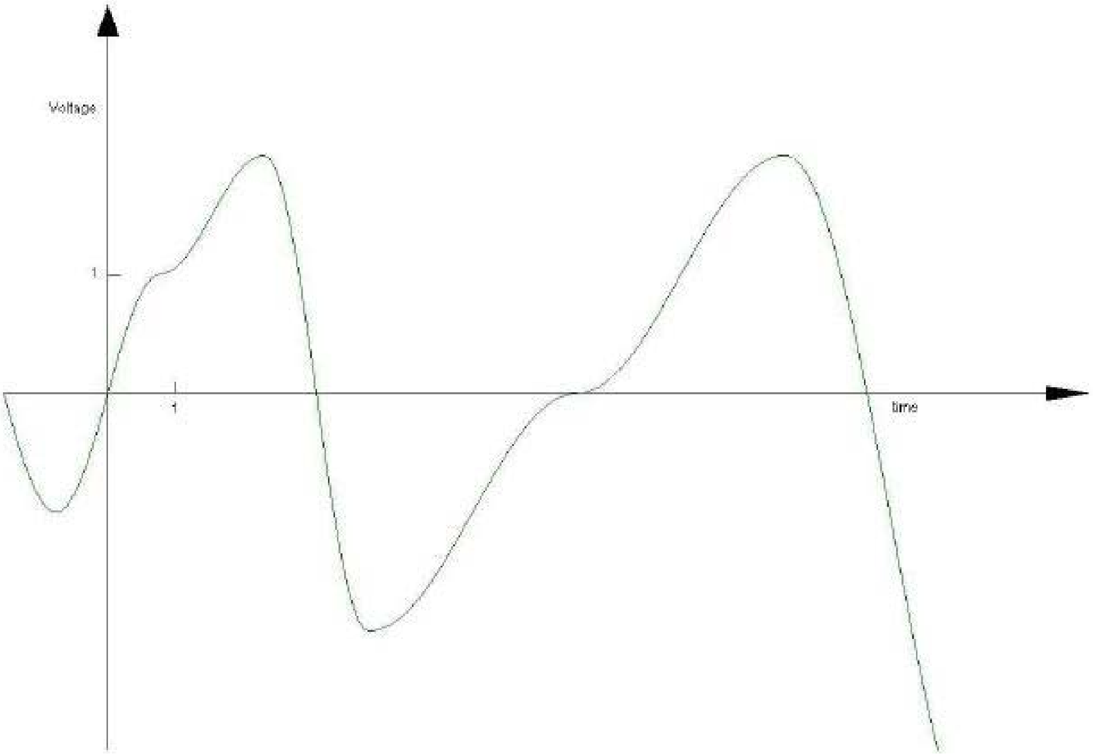
Third stage: Sampling of the voltage waveform takes place according to the input of sampling frequency and number of bits (bit-depth) which we allocate to represent the voltage level. How the rate and bit allocation of sampling varies the quality of sound is that the higher the sampling frequency the better the original sound wave is replicated and the higher the bit-depth the finer is the distinction between the different sounds. After a time the difference is noticeable only to the trained ear and hence further increase in rate of sampling or bit-depth is not required. This is due to the limited ability of the human senses to recognize sound intensities and difference between frequencies.

In the following figures the red lines are the values of the output under a zero-order hold approximation that is the value is retained until the next sampling point in time comes in. (note that the mechanism responsible for the hold is not a part of our system but is depicted here to make understanding the system easy) Some of the sampled waveforms are redrawn below.

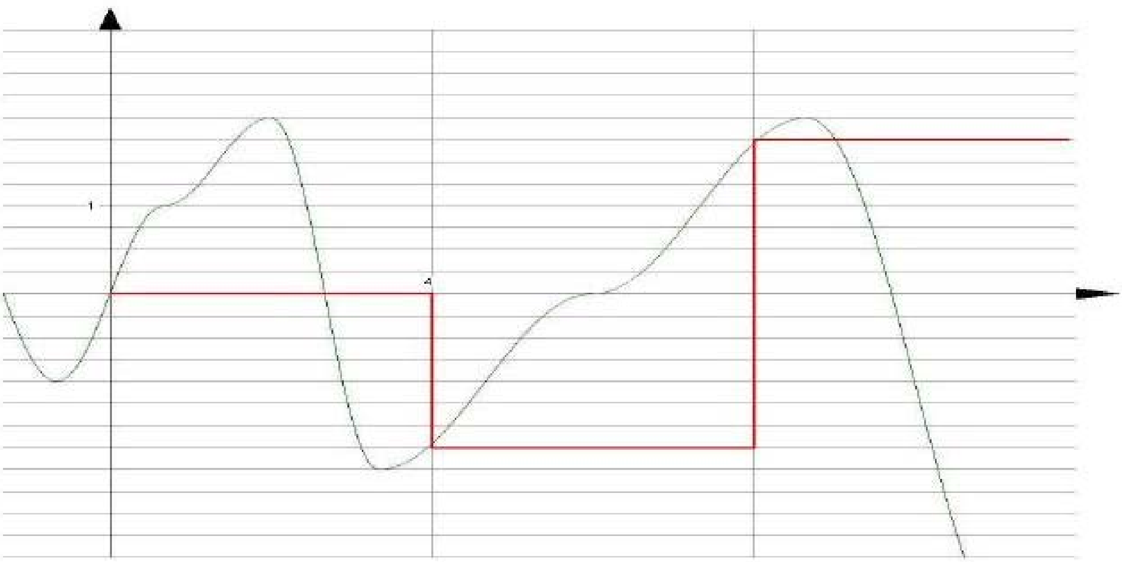
Let sampling rate and bit depth be say once every ‘4*n’ units and bit-depth be ‘m’ digits respectively for the following sampled output.



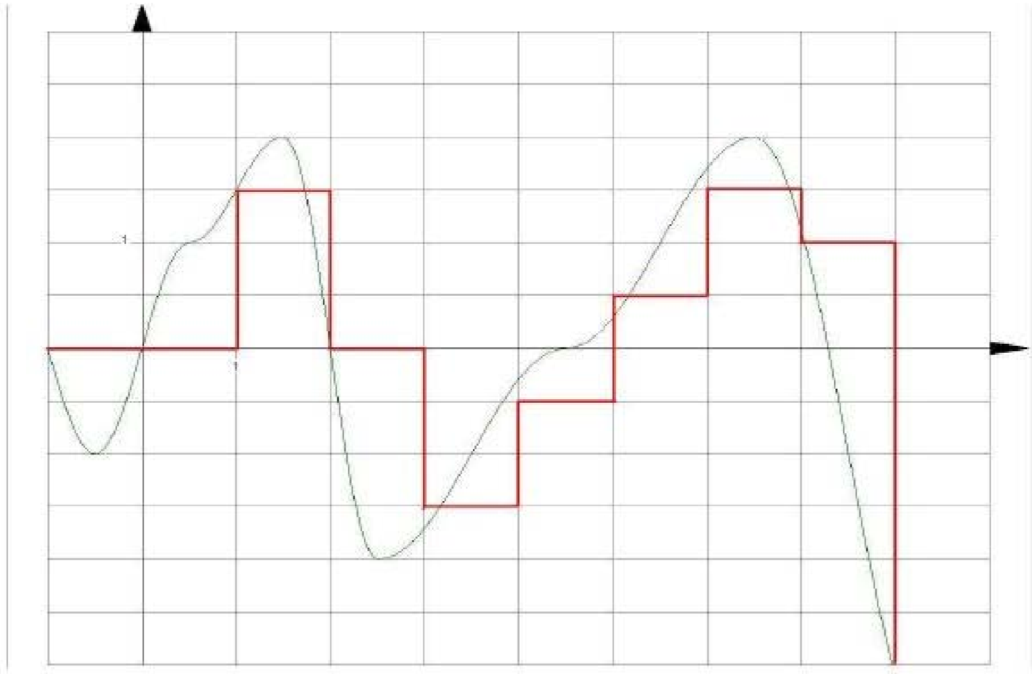
In the following figure the values for sampling rate and bit-depth would be ‘n’ and ‘m-1’ digits respectively.



In the following figure the values for sampling rate and bit-depth would be ‘n’ and ‘m+1’ digits respectively.



In the following figure the values for sampling rate and bit-depth would be ‘n’ and ‘m+1’ digits respectively.



As we observe higher sampling frequencies result in better approximations of the original signal. And as mentioned, let us say the human ear in this example can't differentiate between two sounds if they are less than '2*n' unit Hz apart (while having the same intensity) and less than '2^m' units apart in intensity. Then as we notice there is no need for sampling at frequencies higher than '2*n' and for allocating more than 'm' digits for the bit-depth. Of course higher values than these are used to make the sound seem even better an approximation though we really can't tell the difference.

The analog filters can only but emulate the ideal ones we are constricted to sampling at higher frequencies since this ensures better phase (zero or linear) and magnitude (almost constant in the region of audible frequencies) filter since they are situated close to the origin and the response being much flatter there. Thus higher sampling frequencies are used to obtain better quality of sound. For example, DVD has a sampling rate of 96kHz while using 20 or 24 as the bit-dept, whereas production quality uses 192kHz giving much better sound due to better approximations of the original signal.

The commonly used sampling rate and bit-depth is 16-bit 44.1kHz sampling used in the .wav or .aiff file types. Though theoretically we can adjust the bit-depth and sampling rate to be anything we wish, the only fallout being that the audio quality would be depend accordingly. As mentioned earlier this is the input to the compression algorithms which further manage to minimize this data (termed lossy compression as compared to lossless compression that we are dealing with) by use of powerful and complex methods which also manage to recover the original data without much loss as compared to the reduction in data which they manage.

Peek into the actual algorithms : Since the data storage required for such audio is quite high we use further compression of this data by ignoring some and representing the rest in more concise manner. This process involves algorithms using complex mathematical calculations with respect to intensity of sound, repetitive sound heard again and again during a recording, variable rate of coding (www.wavetrace.com www.mp3-tech.com)etc which aims to minimize the data required to store a song, all driven by download (transfer) time taken.

Why digital audio : Analog audio has many obvious shortcomings. For example making a copy introduces noise and unwanted signals every time due to mechanical contact. The storage mechanism (tapes, etc.) have limited performance and life that is undergo degradation. Analog audio has uneven frequency response. Error correcting data cannot be added and data lost/damaged cannot be replaced.

Besides everything nowadays is going digital, and it is the obvious choice too since it is much longer lasting, has better performance parameters like flat audio response, higher dynamic range (96 dB as opposed to maximum of 80 dB for analog) etc. Besides since all technologies now use digital format of all available material it has become inevitable too.

Playback is achieved by approximation of the sampled output using various techniques like sample and hold technique, linear approximation, 2nd order approximation etc. Now we look at the properties of the system under consideration. To repeat, the system has the following inputs and output. Audio i/p , Sampling rate, Bit-Depth :: Sampling going on at specific intervals of time, binary digits as output.

Properties of the system

Linearity : No
The sum of two audio signals(vectors with amplitude and frequency of vibration), that is the air pressure created due to them is added linearly and hence the equivalent pressure is a like the pressure created due to the 'sum' of the two signals An example: Two signals A*sin(2pa*t) and B*sin(2pb*t) are signals with frequencies 'a' and 'b' and intensities 'A' and 'B' respectively produce certain variations in the air pressure with time Pa(t) and Pb(t) respectively then the equivalent pressure sensed by the system will be a sum equal to Pa(t) + Pb(t) with the net result being that the system gives a linear output since the part of the system is just represents the analog voltage data in digital form keeping its value unchanged or maybe scaling it by a factor at the most.

If the microphone output does not linearly varying with the air pressure then the above won't hold though the system excluding the microphone would be linear again assuming that the sum of two input pressures would not be above the dynamic uppe limit in which case clipping (a phenomenon in which all the values above the highest are clipped to the highest value) takes place.

Causality : Yes

Clearly no input pressure variation till time t=0 the voltage waveform of the microphone is zero for all t<0 since there is no excitation. Thus no nonzero digits are given out at the output end since values of all the samples is zero. Also this may be said to be so since all real systems are causal.

Time Invariance : Yes Since the value of the output binary number depends only on the value of current input the system is time-invariant.

Stability : Yes Since the value of the output signal is just a scaled version of the input pressure of the air which in turn is always bounded, the system is a stable one.

Memory : No Also the sampler gives out only the output at only a given instant in time the system is memoryless. However systems cascaded to this one use zero order hold, linear approximation etcetera then the system(inclusive of the approximating system) no longer remains memoryless.

Invertibility : No Since the data of the values of the original signal between the sampling points is no longer known, the original waveform cannot be reconstruced however good the sampling mechanism be and hence the system cannot be invertible.

With respect to the other inputs the system, the sampling rate and the bit-depth is Causal, Time Invariant and Memoryless. The other two properties have no significance with respect to these inputs.

References:

Fundamentals of Digital audio : P. Jeffrey Bloom and Guy W. McNally
http://www-ccrma.stanford.edu/courses/192b/lectures/2/2.html
http://www.tc.umn.edu/~erick205/Papers/paper.html
http://www.teamcombooks.com/mp3handbook/MP3_Handbook.htm

Digital Eye

by

Ashish Bhat (02007025)

Sandhya Hegde (02d07027)

Abstract:

Image Processing and a simple “Digital Eye” Images are a description of how a parameter varies over a surface. A standard visual image results from light intensity variations across a plane. Other parameters that can be used to form an image are temperature, x-ray emission from a galaxy, ground motion during an earthquake etc. Digital Images are a result of quantization of surface dimensions and the intensity of parameter used to form the image.

Image Processing refers to the various techniques used to improve the quality of image. It involves the two basic principles of Convolution and Fourier analysis.

Digital Eye: A Simple model of the Human eye

We will consider a system that is a simplified model of human eye. The input to this system will be a digital image and the output would be a digital image as perceived by the digital eye. We shall then compare the characteristics of our model with those of the human eye, and analyze its properties as a system:

- Linearity
- Causality
- Memory
- Shift Invariance
- Stability and convergence

We shall also consider some real-life situations and show how our system behaves in a manner similar to the human eye.

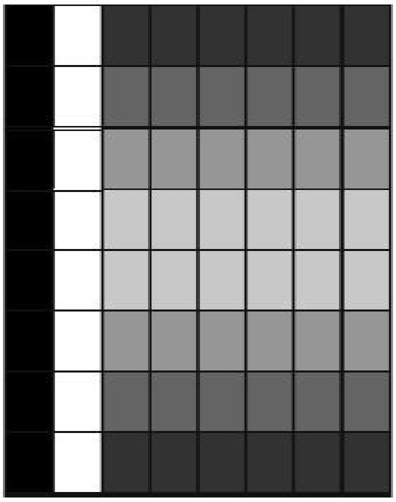
The System

A digital image represents the quantized variation of a parameter over the spatial domain (usually in two dimensions). A typical image may consist of thousands of **samples** arranged in an array of hundreds of rows and columns. Each sample is called a **pixel** or a picture element.

0	255	50	50	50	50	50	50
0	255	100	100	100	100	100	100
0	255	150	150	150	150	150	150
0	255	200	200	200	200	200	200
0	255	200	200	200	200	200	200
0	255	150	150	150	150	150	150
0	255	100	100	100	100	100	100
0	255	50	50	50	50	50	50

Each **pixel** is a number between 0 and 255, directly proportional to the magnitude of the sampled independent parameter. To convert this set of pixels into a **visual image**, the value of the pixel is converted into a **grayscale.**

In a **grayscale**, the value of 0 is **black,** the value of 255 is **white** and intermediate numbers are shades of **gray.**



It is common for 256 **gray levels** (quantization levels) to be used in image processing, corresponding to a single byte per pixel. There are **two** very important reasons for this :

- A single byte is convenient for data management, since this is how computers usually store data.
- A brightness step size of 1/256 (0.39%) is smaller than the eye can perceive. **An image presented to a human observer will not be improved by using more than 256 levels.**

Fig 1(a) represents the pixel view of an image with magnitude of a pixel varying between 0 and 255. Fig 1(b) is the grayscale view interpretable by the human eye.

The Digital Eye: A model of the human visual system

System Input : Input to our system is a moving or static digital image
System Output : Output is the image as perceived by the 'eye'

Let **x[u, v, n]** be the current image input to the system. Here **u** and **v** are the co-ordinates of the pixels in the image and **n** is discrete time unit.
Let the dimensions of the image be $\mathbf{U_{MAX} \times V_{MAX}}$

We model our digital eye as follows:

y[u, v, n]	**=**	$\mathbf{k_1 * x[u, v, n] + k_2 * x_M[u, v, n] + k_3 * y[u, v, n-1]}$
$\mathbf{x_M[u, v, n]}$	**=**	$\mathbf{(1/9) * \sum\sum x[u-i, v-j, n] \qquad i, j = -1,0,1}$

Where $\mathbf{k_1 + k_2 + k_3 = 1} \qquad \mathbf{k_1, k_2, k_3 > 0}$
$\mathbf{0 \le x[u, v, n] \le 255}$
$\mathbf{0 \le u < U_{MAX}}$
$\mathbf{0 \le v < V_{MAX}}$

We are making the following simplifications in our model:

1) The input is assumed to be already 'sampled', i.e. **discrete.**
2) Many natural phenomena pertaining to the eye are not discussed, like focusing, resolution, etc.
3) Here, **x[n] is considered to be solely the intensity of the image at time n.**

The output at time **n** comprises of three distinct parts:

1) The input pixel x[u, v] at time **n.**

2) The mean value of the input pixels at and immediately around (u, v) i.e.

$$x_M[u, v, n] = (x[u-1, v-1, n] + x[u-1, v, n] + x[u-1, v+1, n] + x[u, v-1, n] + x[u, v, n] + x[u, v+1, n] + x[u+1, v-1, n] + x[u+1, v, n] + x[u+1, v+1, n]) / 9$$

3) The output at time **n-1.**

Thus, the system has an implicit definition in the form of a **difference equation.**

As an example of a realistic system we may fix the values of k_1 , k_2 , k_3 as follows:

$$\mathbf{k_1 = 0.7}$$
$$\mathbf{k_2 = 0.1}$$
$$\mathbf{k_3 = 0.2}$$

This makes our system ~

$$\boxed{\mathbf{y[u, v, n] = 0.7*x[u, v, n] + 0.1*x_M[u, v, n] + 0.2*y[u, v, n-1]}}$$

We will use this equation to analyze the behavior of the system.

Homogeneity

As a starting condition let $y[u, v, -1] = 0$

Consider some input $\mathbf{x[u, v, n]}$ and its output $\mathbf{y_1[u, v, n]}$
$y_1[u, v, 0] = k_1*x[u, v, 0] + k_2*x_M[u, v, 0] + k_3*0$
$y_1[u, v, 1] = k_1*x[u, v, 1] + k_2*x_M[u, v, 1] + k_3*y_1[u, v, 0]$
and so on...

Now consider input $\mathbf{k*x[u, v, n]}$
As $\mathbf{x[u, v, n]}$ is multiplied by $\mathbf{k}$, the mean function $\mathbf{x_M[u, v, n]}$ also gets multiplied by **k.**
Its output $\mathbf{y_2[u, v, n]}$ will be
$y_2[u, v, 0] = k_1*k*x[u, v, 0] + k_2*k*x_M[u, v, 0] + k_3*0$
$y_2[u, v, 1] = k_1*k*x[u, v, 1] + k_2*k*x_M[u, v, 1] + k_3*y_2[u, v, 0]$
and so on...

Therefore
$y_2[u, v, 0] = k*y_1[u, v, 0]$
$y_2[u, v, 1] = k*y_1[u, v, 1]$

Thus we conclude by induction, that $y_2[u, v, n] = k*y_1[u, v, n] \quad n = 0, 1..$

Thus the system is homogenous.

This property can be interpreted as follows:
"If the intensity of input image increases by a factor of k, the intensity of the image perceived by the eye also increases by a facto of k".

Additivity

As a starting condition let $y[u, v, -1] = 0$

Consider some input $\mathbf{x_1[u, v, n]}$ and let $\mathbf{x_{M1}[u, v, n]}$ be its mean function
Its output $\mathbf{y_1[u, v, n]}$ will be
$y_1[u, v, 0] = k_1*x_1[u, v, 0] + k_2*x_{M1}[u, v, 0] + k_3*0$
$y_1[u, v, 1] = k_1*x_1[u, v, 1] + k_2*x_{M1}[u, v, 1] + k_3*y_1[u, v, 0]$
and so on...

Consider another input $\mathbf{x_2[u, v, n]}$ and let $\mathbf{x_{M2}[u, v, n]}$ be its mean functior

Its output $\mathbf{y_2[u, v, n]}$ will be
$y_2[u, v, 0] = k_1*x_2[u, v, 0] + k_2*x_{M2}[u, v, 0] + k_3*0$
$y_2[u, v, 1] = k_1*x_2[u, v, 1] + k_2*x_{M2}[u, v, 1] + k_3*y_2[u, v, 0]$
and so on...

Now suppose the input to the system is $\mathbf{x_1[u, v, n] + x_2[u, v, n]}$
The corresponding mean function will be $\mathbf{x_{M1}[u, v, n] + x_{M2}[u, v, n]}$
Let its output be $\mathbf{y[u, v, n]}$. Therefore
$y[u, v, 0] = k_1*(x_1[u, v, 0] + x_2[u, v, 0]) + k_2*(x_{M1}[u, v, 0] + x_{M2}[u, v, 0])$
$+ k_3*0$
$y[u, v, 1] = k_1*(x_1[u, v, 1] + x_2[u, v, 1]) + k_2*(x_{M1}[u, v, 1] + x_{M2}[u, v, 1])$
$+ k_3*y[u, v, 0]$

Therefore
$y[u, v, 0] = y_1[u, v, 0] + y_2[u, v, 0]$
$y[u, v, 1] = y_1[u, v, 1] + y_2[u, v, 2]$

and so on...

By induction, we conclude that
$y[u, v, n] = y_1[u, v, n] + y_2[u, v, n]$ $\qquad n = 0, 1, 2...$

Thus the system is additive.

However addition of two image signal has no physical relevance. So this property of the system does not translate into any property of the eye.

- **Linearity**

As the system is both homogenous and additive,
The system is Linear.

Thus if
$\mathbf{x_1[u, v, n]}$ → **System** → $\mathbf{y_1[u, v, n]}$
$\mathbf{x_2[u, v, n]}$ → **System** → $\mathbf{y_2[u, v, n]}$

We have
$\mathbf{a*x_1[u, v, n] + b*x_2[u, v, n]}$ → **System**
→ $\mathbf{a*y_1[u, v, n] + b*y_2[u, v, n]}$

However care must be taken so that the value of
$\mathbf{a*y_1[u, v, n] + b*y_2[u, v, n]}$ does not exceed 255. Otherwise the data of the digital image becomes corrupt.

To overcome this problem we can define out system as:

$y[u, v, n]$	=	$MIN(255, k_1*x[u, v, n] + k_2*x_M[u, v, n] + k_3*y[u, v, n-1])$
$x_M[u, v, n]$	=	$(1/9) * \sum\sum x[u-i, v-j, n]$ $\quad$ $i, j = -1,0,1$

Where $k_1 + k_2 + k_3 = 1$ $\quad$ $k_1, k_2, k_3 > 0$
$0 \leq x[u, v, n] \leq 255$
$0 \leq u < U_{MAX}$
$0 \leq v < V_{MAX}$

In that case the system becomes non-linear, but the data of the digital image is always valid. In case the intensity of an pixel becomes more than 255, it is limited to the value of 255. This produces some distortion in the image.

However such superposition of signals does not have much of a meaning when considering images as signals. The input to the system is always a single image.

Causality

As the output at discrete time [**n**] depends only on the input at discrete time **n** and the output at discrete time [**n-1**] the system is Causal i.e. The output at any time depends only on values of the input at present time and values in past, but not any future values of the input.

Memory

The system has memory because the output value at discreet time [**n**] depends on values of inputs at [**n-1**], [**n-2**], ... , [**0**].

Shift Invariance

Technically, our model is not shift–invariant with respect to time since each output depends on ALL past inputs to the system.

Hence, the same picture viewed at two different points in time may appear different, if the image viewed just before were so.
But due to the very small value of k_3, in practice, the system functions as shift invariant over larger periods of time .

Stability

$y[u, v, n] = k_1*x[u, v, n] + k_2*x_M[u, v, n] + k_3*y[u, v, n-1]$

Now as $0 \le x[u, v, n] \le 255 \rightarrow 0 \le x_M[u, v, n] \le 255$

As $k_1 + k_2 + k_3 = 1$ and $k_1, k_2, k_3 > 0$
$0 < k_1 < 1$
$0 < k_2 < 1$
$0 < k_3 < 1$

As a starting condition let $y[u, v, -1] = 0$

Therefore
$0 \le y[u, v, 0] \le k_1*255 + k_2*255 + k_3*0$
$0 \le y[u, v, 0] \le (k_1 + k_2)*255$
$0 \le y[u, v, 0] \le 255$

$0 \le y[u, v, 1] \le k_1*255 + k_2*255 + k_3* y[u, v, 0]$
$0 \le y[u, v, 1] \le k_1*255 + k_2*255 + k_3* 255$
$0 \le y[u, v, 1] \le (k_1 + k_2 + k_3)*255$
$0 \le y[u, v, 1] \le 255$

and so on..

Thus, by Induction, $0 \le y[u, v, n] \le 255$

Therefore for a bounded input the output of the system is also bounded.
So the system is stable by definition.

Stability interpreted as convergence

Now we consider the stability of this system in another sense. Suppose the input changes and remains constant for a while, we know from the equation that the output will keep changing. But will it converge or keep on varying?

As a starting condition let $y[u, v, -1] = y_0[u, v]$ (some image)

Now let the input become $x_0[u, v]$ and remain constant.
Let $x_{M0}[u, v]$ be its mean function which will also be a constant.
Let $k_1*x_0[u, v] + k_2*x_{M0}[u, v] = x_1[u, v]$

Therefore

$$\begin{aligned} y[u, v, 0] &= k_1*x_0[u, v] + k_2*x_{M0}[u, v] + k_3*y_0[u, v] \\ &= x_1[u, v] + k_3*y_0[u, v] \end{aligned}$$

$$\begin{aligned} y[u, v, 1] &= k_1*x_0[u, v] + k_2*x_{M0}[u, v] + k_3*y[u, v, 0] \\ &= x_1[u, v] (1 + k_3) + k_3^2*y_0[u, v] \end{aligned}$$

$$y[u, v, 2] = x_1[u, v] (1 + k_3 + k_3^2) + k_3^3*y_0[u, v]$$

and

$$y[u, v, n] = x_1[u, v] (1 + k_3 + k_3^2 + \ldots + k_3^n) + k_3^{(n+1)}*y_0[u, v]$$

Therefore

$$y[u, v, n] = x_1[u, v] (1 + k_3 + k_3^2 + k_3^3 + \ldots + k_3^n) + k_3^{n+1}*y_0[u, v]$$

Because $k_3 < 1$ as n becomes large we get

$$\begin{aligned} y[u, v, n] &= x_1[u, v] / (1 - k_3) \\ &= (k_1*x_0[u, v] + k_2*x_{M0}[u, v]) / (1 - k_3) \end{aligned}$$

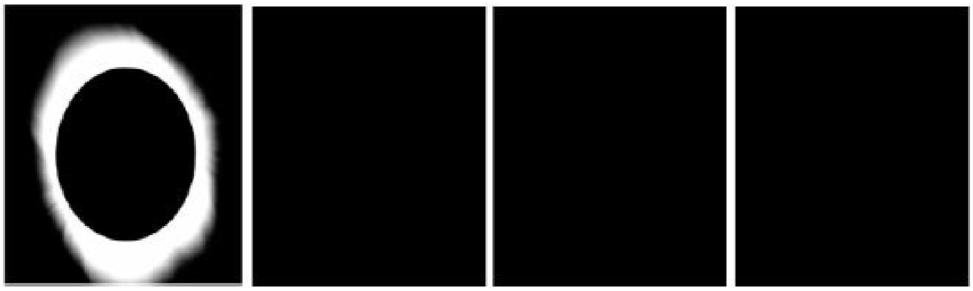
Thus if after varying, the input image becomes constant, the output image converges to an image which depends solely on the constant input image and not the previous input images.

Behavior of the model in real-life

Case I – Image Retention

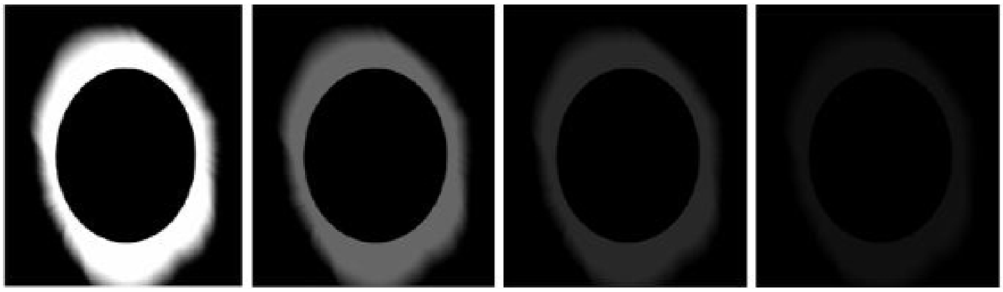
We have observed in real life that many a times when we see an extremely intense image, it stays on for a few split seconds even after we, say, close our eyes. With this example, we wish to demonstrate how our system imitates the human eye.

Let us consider the following inputs at n= 0, 1, 2, and 3.



x[0] **x[1]** **x[2]** **x[3]**

The outputs as given by our system equation is:



y[0] **y[1]** **y[2]** **y[4]**

There is an exponential decay of intensity of the image with respect to time as it fades from the human vision.

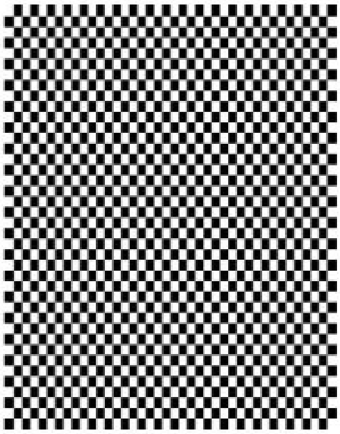
Case II – Making a grayscale

White + Black = Gray

Printed images are divided into two categories: **grayscale** and **halftone**. Each pixel in a halftone image is formed from many individual *dots*, with each dot being completely black or completely white. Shades of gray are produced by alternating various numbers of these black and white dots. Each pixel would correspond to an array of 16 by 16 printable dots. Black pixels are formed by making all of these 256 dots black. Likewise, white pixels are formed making all of these 256 dots white. Mid-gray has one-half of the dots white and one-half black. Since the individual dots are too small to be seen when viewed at a normal distance, the eye is fooled into thinking a grayscale has been formed.

Our system incorporates this feature of the eye by including the mean term in the system equation. However this effect is not observable at normal pixel level because the weight of the mean factor is very low. But as we go to the sub-pixel levels, the weight of the mean factor increases and the value of the pixel approaches the MEAN of all the sub-pixels near it.

If we set $k_1 = 0$, $k_2 = 1$, $k_3 = 0$ in the system equations, we get the following:



→ SYSTEM →

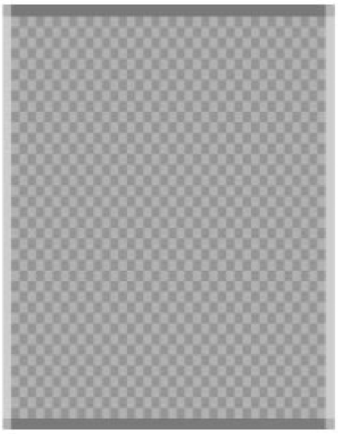


Image processing Methods

Image Processing is the technique of making an image "clearer"
So that its data is easily interpretable by the human eye. It is also used to treat images to make them more visually appealing. The most important methods involved are **Convolution** and **Fourier Transformation.**

Techniques of image Processing :

- **Linear filtering**

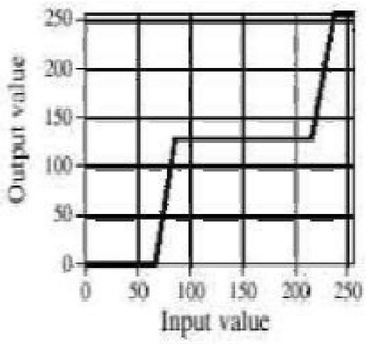
This can improve images in many ways: sharpening the edges of objects, reducing random noise, correcting for unequal illumination, deconvolution to correct for blur and motion, etc. These procedures are carried out by convolving the original image with an appropriate filter kernel, producing the filtered image.

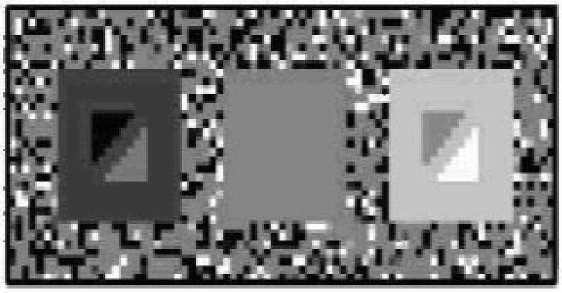
- **Grayscale Transformation**

An image must have the proper brightness and contrast for easy viewing. Brightness refers to the overall lightness or darkness of the image. Contrast is the difference in brightness between objects or regions. For example, a white rabbit running across a snowy field has poor contrast, while a black dog against the same white background has good contrast.
As an example:

Before and after Gray Scale transformation . Source **: www.dspguide.com**

Grayscale transformation is achieved by increasing the contrast between bright and dark regions (by enhancing their respective brightnesses and darknesses) . Thus it acts as a basic high pass filter. For example : two threshold levels are established and the continuous intensity function is discretised with respect to these levels:





Edge Enhancement

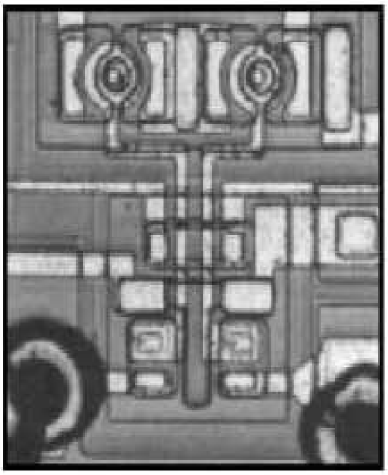
Edge enhancement uses the method of convolution to detect edges and make the image more distinct. (this is also called sharpening).

Two- Variable Fourier Transform (Importance of phase)

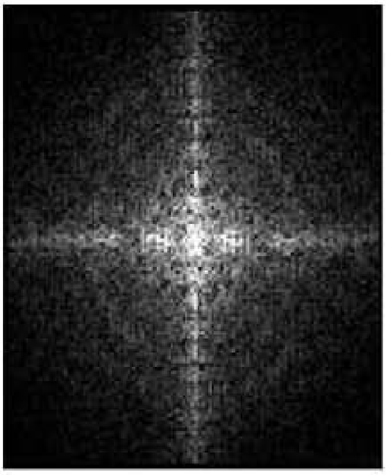
A 2 variable signal x[u1,u2] can be thought of as having a Fourier transform X[jw1,jw2]. This represents a decomposition of the image into complex exponential components that capture the spatial variations of x[u1,u2] at different frequencies in each of the 2 coordinate directions.
The most important visual information is contained in the EDGES and regions of HIGH CONTRAST. Intuitively, regions of minimum and maximum intensities in a picture are places where complex exponentials of different frequencies are in phase.

Thus, the phase of the picture captures much of the information about the image, in particular , the edges.
If the phase information of an image gets corrupted , the image can thus change beyond recognition !!

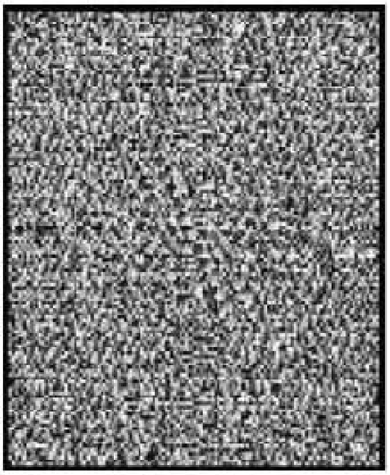
Source: **www.dspguide.com**



$x[u_1,u_2]$



magnitude $X[w_1,w_2]$



phase $X[w_1,w_2]$

RGB coding

Color is added to digital images by using three numbers for each pixel, representing the intensity of the three primary colors: red, green and blue. Mixing these three colors generates all possible colors that the human eye can perceive. A single byte is frequently used to store each of the color intensities, allowing the image to capture a total of 256×256×256 = 16.8 million different colors.

Acknowledgements

Thanks to Prof. V. M. Gadre for his constant encouragement and for giving us the opportunity to create this document.

References

[1] Oppenheim, Wilsky and Nawab, *Signals and Systems*, 5th ed.
[2] Steven W. Smith, *Digital Signal Processing*, 2nd ed.
[3] www.dspguide.com

Pupil Light Reflex of the Human Eye

by

Nilesh Meshram (02007027)

Anand Gajbhiye (02007030)

Viren Diwan (02007032)

Abstract:

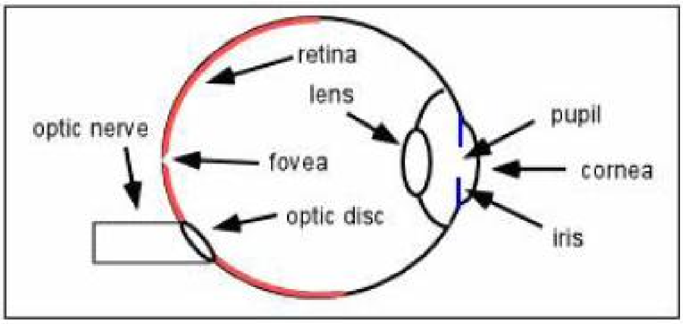
The eye is an extraordinary molecular computer. Even though the structure and the working of the eye is very complex the functioning of the eye can be attributed to the interaction of the pupil, iris and lens. Since the topic chosen by us earlier was the computational eye where we intended to study the computation of visual information from light waves passing through eye, we now will restrict to "the pupil light reflex". The amount of light entering the eye is determined by the size of pupil aperture. In dim light it's best to take in as much light as possible, whereas in bright light it's best to take in less light so as not to damage fragile structures inside the eye.

Introduction:

Working of an eye: Light, reflecting from an object, let's say a ball, first strikes the cornea, which is a small transparent area in the outer covering of the eye. Light then passes through the pupil, which is a hole that is surrounded by the iris, the color in our eyes. The iris expand or contract in response to light, so that the iris is small and the pupil large in dim or dark light and vice-versa in bright light. The light passes through the pupil to the lens which reverses the image so that an upside down image is displayed on the back of our eye to the retina where the signal is converted so that the brain can understand it. The pupil, iris, and lens have already manipulated the light waves that make up this image, and our visual system actively works on the light/image in many ways, hence, what we actually "perceive" is very much an augmented version of the "real" visual stimulus.

Motivation:

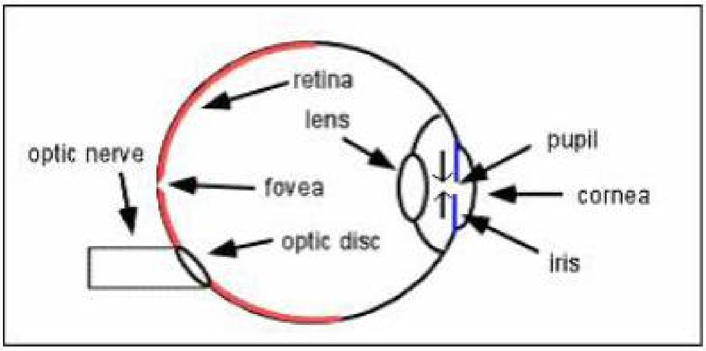
The eyes are our windows to the outside world. These eyes enable us to see this world God has created for us. So we decided to explore something about this wonderful gift of the God in context with the subject we are learning.



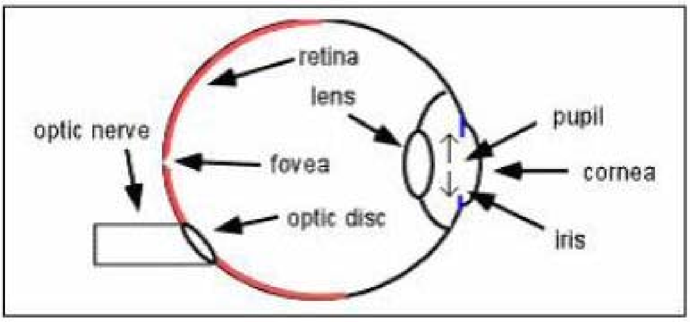
❖ The system: Pupil

The system: Pupil

As stated earlier, Size of pupil aperture determines the amount of light allowed to the retina. m When light intensity is higher than desired, the pupil light reflex causes a contraction to reduce the amount of admitted radiation.



When light intensity is lower, aperture enlarges to allow more light to the retina.

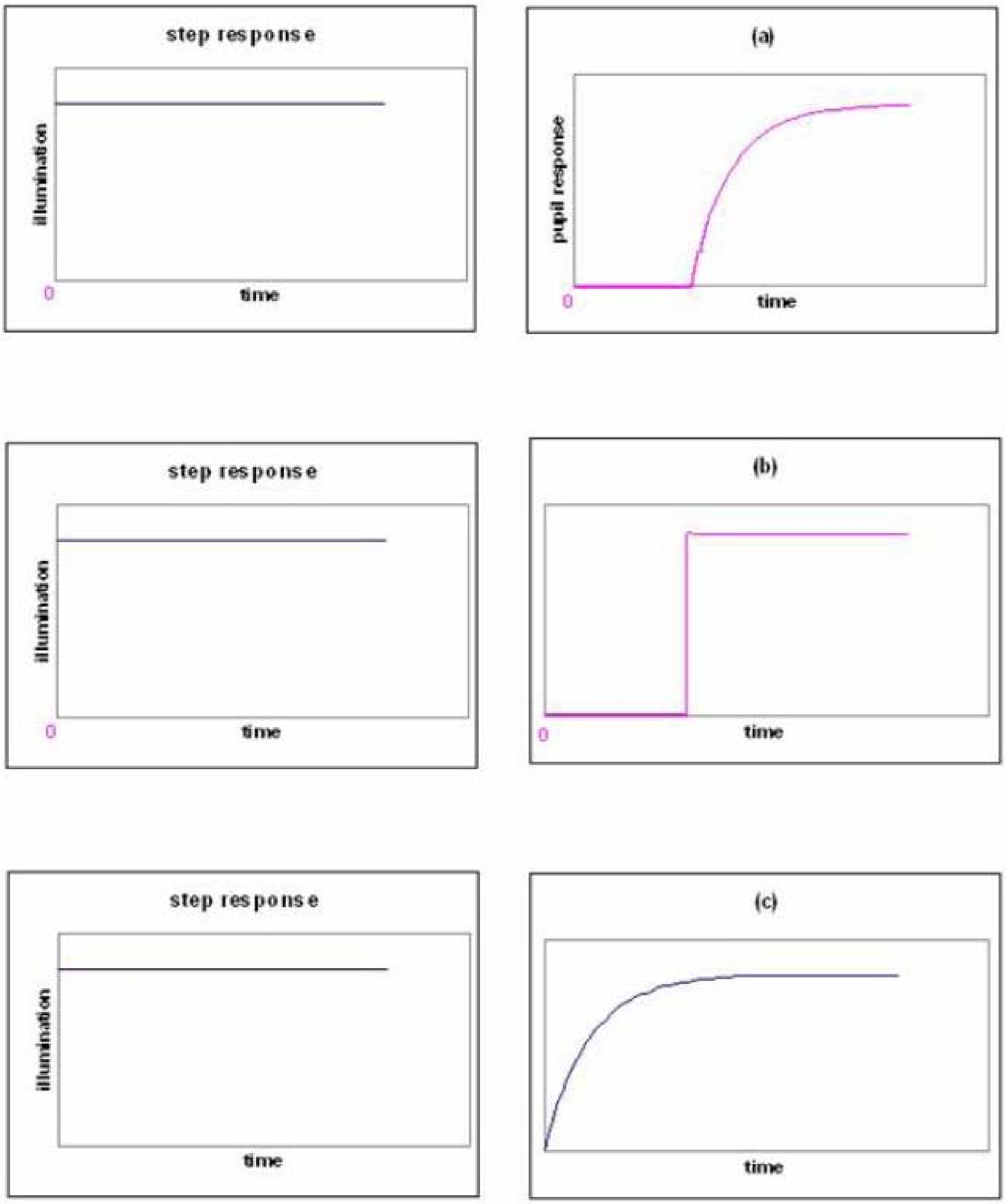


Suppose the level of illumination is suddenly changed. Let the illumination be unit step function of time.
$u(t) = 0 \quad t < 0$
$= 1 \quad t > 0$

This demands a change in the pupil size. The pupil cannot change its size instantaneously. But it approximates it with an exponential function.

Thus, pupil response to step response of change in illumination consists of two parts:
· Exponential delay: As explained above, signal shape is changed from the step function to exponential function.
· Pure delay: The signal is shifted and there is no shape change in the step function. The pure delay occurs because of the time taken by the signal for to and fro conduction on nerve fiber from eye to brain. Time for each synapse varies from 0.3 to several milliseconds. Synapse is the junction between two neurons.

The response of the pupil to the step change in illumination. (a)The response is a combination of a pure delayed response and an exponential delayed response. Taking the response apart (b)a pure delay and (c) an exponential lag.

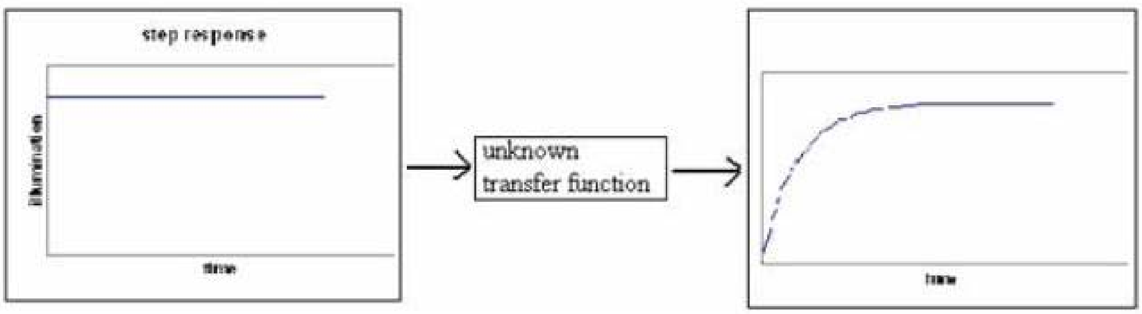


Transfer function for exponential delay component:

Let the input function given to the pupil be a unit step function.
Input: Unit step function
u(t) = 0 t < 0
= 1 t > 0

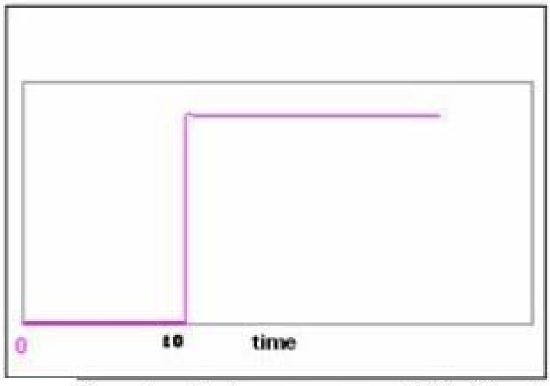
Therefore as discussed earlier, the response of the pupil is an exponentially delayed output response.
Output response=
t=0 output response=0
when t=large output response=1
Laplace transform of input =
Laplace transform of output =
The transfer function is the ratio Laplace transforms for the input and output. Hence,the transfer function is given by,

Thus the transfer function which characterizes the output response to the input function is of the form where k is called gain constant and a is called the speed of an exponential response.

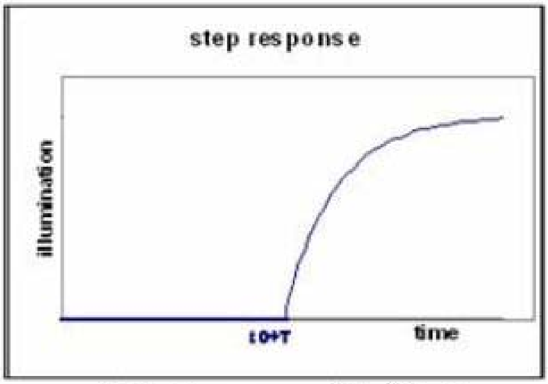


Propeties of the System

Shift-invariance: If the input unit step response is shifted by t0 output is also shifted by t0. Hence the system is shift invariant.



Input unit step response shifted by t_0



Output response shifted by t_0

Causality: A system is causal if the output at any time depends only on inputs at present time and the past, i.e. the system output does not anticipate the future values of the input. Here the system is non-anticipative, hence it is causal.

Stability: A system is stable when a bounded input gives a bounded output.

For $0 < x(t) < Mx <$ infinity
we have $y(t) <$ infinity where $My = Mx$
As derived earlier, transfer function $H(s) = k/(s+a)$.Hence $|H(s)| < k/a$ and $Y(s)$ is bounded for bounded inputs.

Linearity: Output response function for input step response $x(t)=c$ is given by
$y(t) = 0 \quad t < T$
$= c\,[1-\exp(-at)] \quad t > T$
$x1(t) = c1$
$y1(t) = 0 \quad t < T$
$= c1\,[1-\exp(-at)] \quad t > T$
$x2(t) = c2$
$y2(t) = 0 \quad t < T$
$= c2\,[1-\exp(-a't)] \quad t > T$

Additivity:
$x3(t) = x1(t) + x2(t) = c1 + c2$
$y3(t) = 0 \quad t < T$
$= c3\,[1-\exp(-a''t)] \quad t > T$

Homogeneity:
$x4(t) = k\,x1(t)$
$y4(t) = 0 \quad t < T$
$= c4\,[1-\exp(-a'''t)] \quad t > T$
If rate of exponential response ($a = a'=a''=a'''$) is constant then the system is additive and homogenous. Hence it will be linear with $c3 = c1 + c2$ and $c4= kc1$.

Else if the rate of exponential response a is a function of x(t) system will be neither additive nor homogenous. Hence it will be nonlinear.
n Memory: The system is memoryless if the output for each value of the independent variable depends only on the input at the same time. For system under consideration input step response $x(t)=a$,

Output response function is given by
$y(t) = 0 \quad t < T$ $=a[1-\exp(-kt)] \quad t > T$
Here the output depends on x(t) as well as t hence the system has memory.

References:

Web link:http://web.umr.edu/~psyworld/eye.htm
Books: Signals and Systems, Alan V. Oppenheim and Alan Willsky

SPEECH SIGNAL PROCESSING

Abstract:

The application presentation deals with bringing out the importance of basic concepts of Signals and Systems in the field of speech communication and it's contribution to the technological advances.

Introduction

Speech processing typically involves a basic representation of a speech signal in a digital domain which requires limiting the band width of the signal, sampling it at a certain corresponding rate and storing each sample with an adequate resolution.

But our focus in the field of speech processing is in communication. Speech can be represented in terms of a signal carrying some message content or information.

Speech signals can be thought of signals in both continuous and discrete domain.

• Speech signals are produced by our vocal cords and are transmitted from speaker to listener due to the pressure waves that are propagated through air resulting in the form of continuous or analog signals.
• Speech when stored as digital information in computers are represented in the form of discrete signals i.e. using integers as the independent variable.
The basic ideas behind speech communication system are
1. Transmission
2. Storage and
3. Processing

Generally all disciplines of signal analysis involve the following steps :-
• Production of the signal from an information source like Speech delivery
• Signal representation in natural domains
• Signal representation in transform domains to get better understanding of the concepts which are not observable in natural domain. Speech can be represented as a waveform or in terms of bits for digital storage.
• Finally extraction and utilization of the information. Example :- In the field of speaker verification the identity of the speaker is the information.

In conclusion, some of the aims in the field of speech processing is efficient speech transmission, economical digital storage and improved communications.

System Specification

Actually, one cannot define the properties of a system in the form of first principles as they vary from system to system used in different areas of applications.

For example, the properties of man-to-machine communication systems may be different from those of man-to-man and machine-to-man communication systems.

Hence, our area of focus is a general communication system.
LINEARITY :- In general, we would expect the systems to be linear in order to produce effects such as amplification and attenuation.

STABILITY :- for any bounded input, we would expect a bounded output in the field of communication.

INVERTIBILITY :- A communication system has to be invertible because it is ridiculous to transmit a signal in such a way that no one can recover the original signal.

MEMORY :- As far as man-to-man communication is considered, the system should be memoryless . But in case of man-to-machine communication systems like speaker verification systems(security systems) ,it should have memory because the identity of the speaker has to be verified from the stored data.

CAUSALITY :- In our day-to-day life, we can consider communication to be causal since the system does not look into future.

TIME-INVARIANCE :-Clearly ,if we delay the input signal ,the output signal would also be delayed by the same amount.

Working

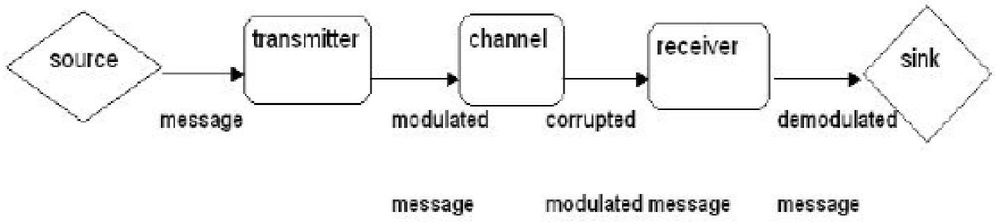


FIGURE :: Structure of a general communication system

In general, a communication system works as seen above.
The source generates a signal such as speech, which is a time-domain signal.
The transmitter is a system designed in such a way that the original signal can be transmitted through the channel and reach it's destination. Since the signal s(t) is weak(low frequency) it can't be transmitted as such, it has to be modulated with a carrier wave first .
Channel is the medium through which the signal is transmitted. One cannot avoid the channel though the signal being transmitted will be corrupted by noise, distorted and attenuated.
The receiver would serve as the inverse system to the transmitter. However, because of the channel, the receiver must do it's best to reproduce the original signal.
The received message is passed to the information sink that some how makes use of the message.

Theory behind the application

• The general process of embedding an information varying signal into a second signal is modulation.

• Extracting the information varying signal is known as demodulation.

Modulation methods rely on the concepts of

1. Amplitude modulation

2. Frequency modulation

3. Phase modulation

• Let x(t) be the information varying speech signal and c(t) be the carrier signal. Then ,the modulated signal y(t) is given by y(t)=x(t)c(t)

A speech signal typically has a low frequency range(200hz to 4khz).

Modulation is the key to transmit a voice signal x(t)at different range of frequencies.

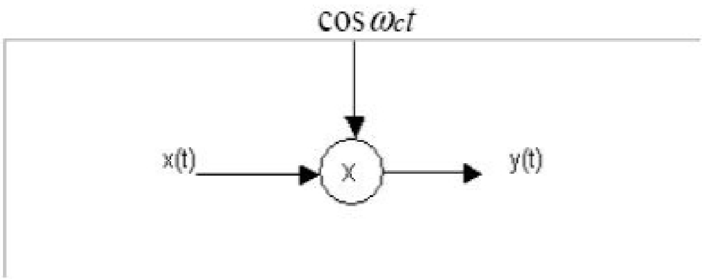


FIGURE :: amplitude modulation with sinusoidal carrier

Assuming θ_c =0 , where θ_c is the phase of carrier signal,

$$C(j\omega)=\pi[\delta(\omega-\omega_c)+\delta(\omega+\omega_c)]$$

$$\Rightarrow Y(j\omega)=\frac{1}{2\pi}\left(X(j\omega)*C(j\omega)\right)$$

$$\therefore\ Y(j\omega)=\frac{1}{2}\left(X(j\omega-j\omega_c)+X(j\omega+j\omega_c)\right)$$

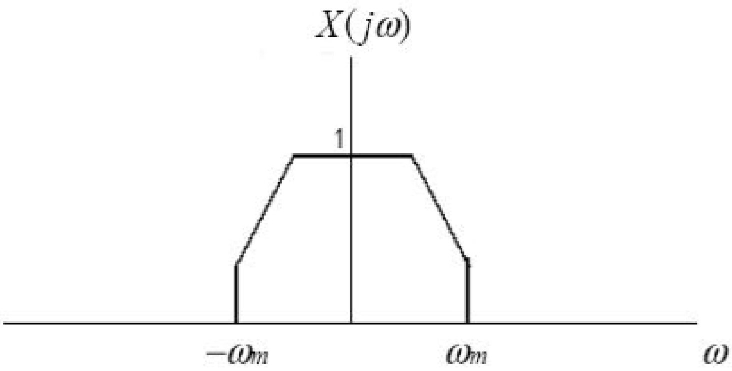


FIGURE :: spectrum of x(t)

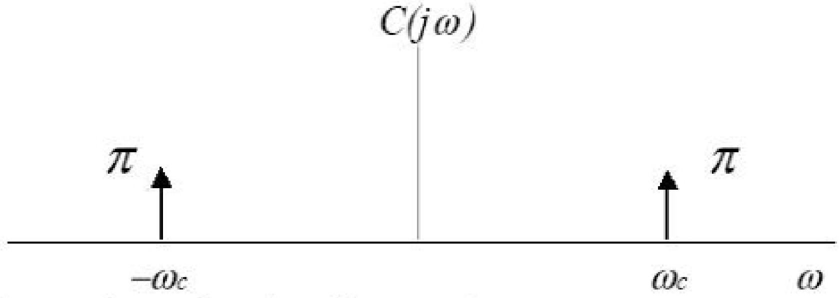


FIGURE :: spectrum of carrier c(t)= $\cos \omega_c t$

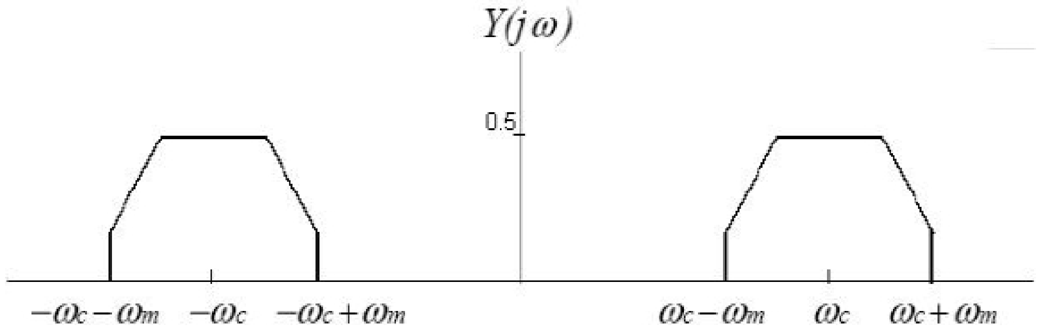


FIGURE :: spectrum of amplitude-modulated signal

- **Demodulation** :-The modulated signal is demodulated at the receiver end in order to recover the original signal.
  At the receiver end ,the modulated signal is multiplied by the same carrier wave and then, the resulting signal is applied to a low-pass filter.

$$y(t) = x(t)\cos\omega_c t$$
$$and \quad w(t) = y(t)\cos\omega_c t$$

The spectra of y(t) and w(t) are as given. x(t) can be recovered from w(t) by applying an ideal low pass filter with a gain of 2 and a cut off frequency that is greater than ω_m and less than $2\omega_c$ - ω_m .

w(t)=x(t) $\cos^2 \omega_c t$

$$\left(\text{Using } \cos^2 \omega_c t = \frac{1}{2} + \frac{1}{2}\cos 2\omega_c t \right)$$
$$w(t) = \frac{1}{2}x(t) + \frac{1}{2}x(t)\cos 2\omega_c t$$

Applying a low pass filter to w(t) we can recover the signal x (t) with half the amplitude of that of the original signal.

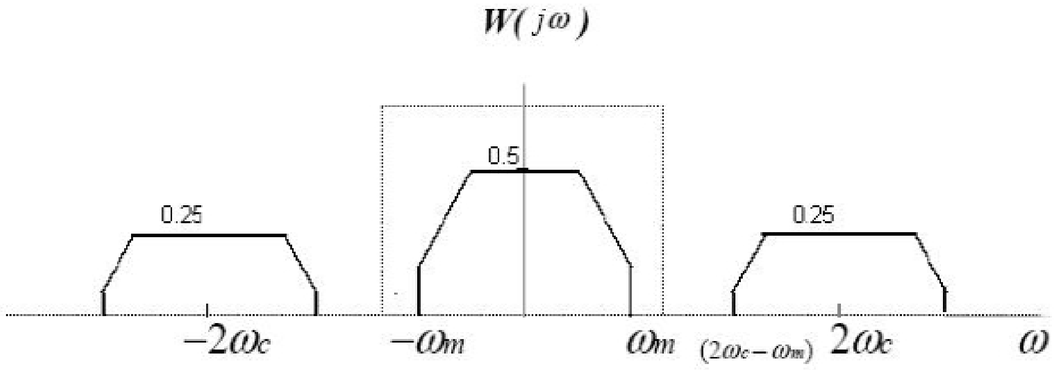


FIGURE :: spectrum of carrier signal multiplied by the carrier. The dashed line indicates of a low pass filter to extract the demodulated signal.

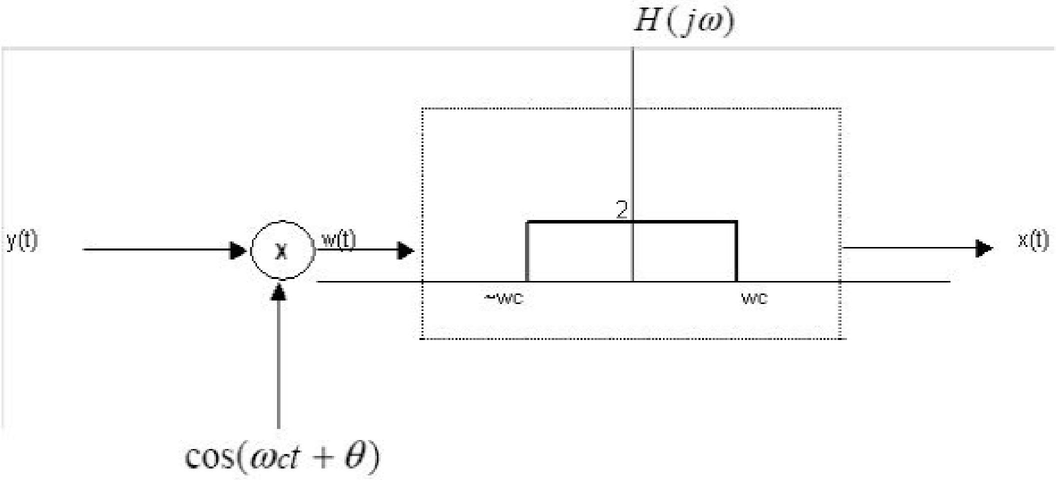


FIGURE :: The demodulation system

Application in real life

Some of the applications of speech communications in real life are Digital transmission and storage of speech.

- **Speech synthesis systems.**

A computer voice response system is basically an all digital automatic service which responds with the desired information by voice commands of the speaker.

- **Speaker recognition systems**

The acoustic speech wave form is converted to a written equivalent of the message information in order to preserve the features of the speech signal that are relevant to speaker identity .The area of speech recognition can be classified into two sub areas

1. Speaker verification: For speaker verification, one's identity is claimed by the user and the decision to accept or reject the claimed identity is taken by the verification system strictly.

2. Speaker identification: Speaker identification is to choose the speaker whose reference patterns are closest to the sample speech patterns. But the above systems can be affected by the following factors
Type of speech
Number of speakers
Type of speakers
Speaking environment
Transmission system etc.

Apart from the above applications there are many more ideas of signals and systems involved in the area of speech processing like :

- **Spectrograms**

In order to design systems for transmitting speech we use spectrograms. Spectrogram is a visualization method or tool to look at frequency content of a signal .It is a plot of frequency-energy envelope versus time. From a spectrogram of a speech signal we can find the energy variation involved in the spectrum. Generally, a human speech contains energy between 0 to 5khz. Thus effective speech transmission systems can be designed with the help of spectrograms which can cope up with the signal's energy variation and thereby enhancing the quality of the signal. For example, in telephone systems which act as band pass filters passing energy between 200hz to

3.2khz, it is difficult to distinguish few high frequency sounds due removal of high frequency energy.

- **Correlation function**
 is a measure of similarities between two signals. Thus it can be used in the field of security systems for speech verification. Cross correlation between two discreet time signals x[n] and y[n] is defined as

$$R_{xy}[l]=\sum_{n=-\infty}^{\infty}(x[n]y[n-l])$$

 where n is the sample index and l is the lag or time shift between the two signals. Speech signals are not stationary, so we are interested in similarities between signals only for a short period of time say less than 30 ms .The cross correlation is computed only over an interval of time samples and for only a few time delays.
- **Linear prediction techniques**
 This method involves representation of samples x[n]in terms of the past samples x[n-k] where k=1,......,p. Therefore

$$x[n]=\sum_{k=1}^{p}a_k x[n-k]$$

 where p is the number of past samples of x[n]which we wish to examine.

Acknowledgement

We would like to thank Prof. V.M.Gadre for his motivation in doing this presentation.

References

• "Digital processing of speech signals"
by Lawrence R. Rabiner
Ronald W. Schafer

• "Signals and systems"
by Alan V. Oppenheim,
Alan S. Willsky
S. Hamid Nawab

• "Speech communications"
by O'Shaughnessy

• "Speech and audio signal processing"
by Ben Gold
Nelson Morgan

TELEVISION SIGNAL BROADCASTING IN SATELLITE

by

SHANTANU RAJWADE

SHANTANU REGE

DEEPAK RAMINEEDI

Introduction:

Traditional forms of communication i.e. using cables, require physical connectivity of the transmitter and receiver. Besides, this is also hindered due to geographical limitations. eg. Having a trans-Atlantic cable to connect the North America and Europe would be economically and physically unviable. Similarly, communication with ships at sea, aeroplanes or for that matter any moving object, would pose a problem for this form of communication. All these problems are easily overcome if we make optimal use of natural resource that is freely available to us – SPACE.

Using Geo-stationary satellites which orbit 35,700 kilometres above the earth's surface give us the freedom to communicate without above limitations and also has the unique advantage that the cost of communication is independent of distance. Today they are used in a wide variety of applications ranging from television broadcasting, mobile telephony, radiometry and weather forecasting.
In our future discussion we focus our attention on use of satellites in television broadcasting.

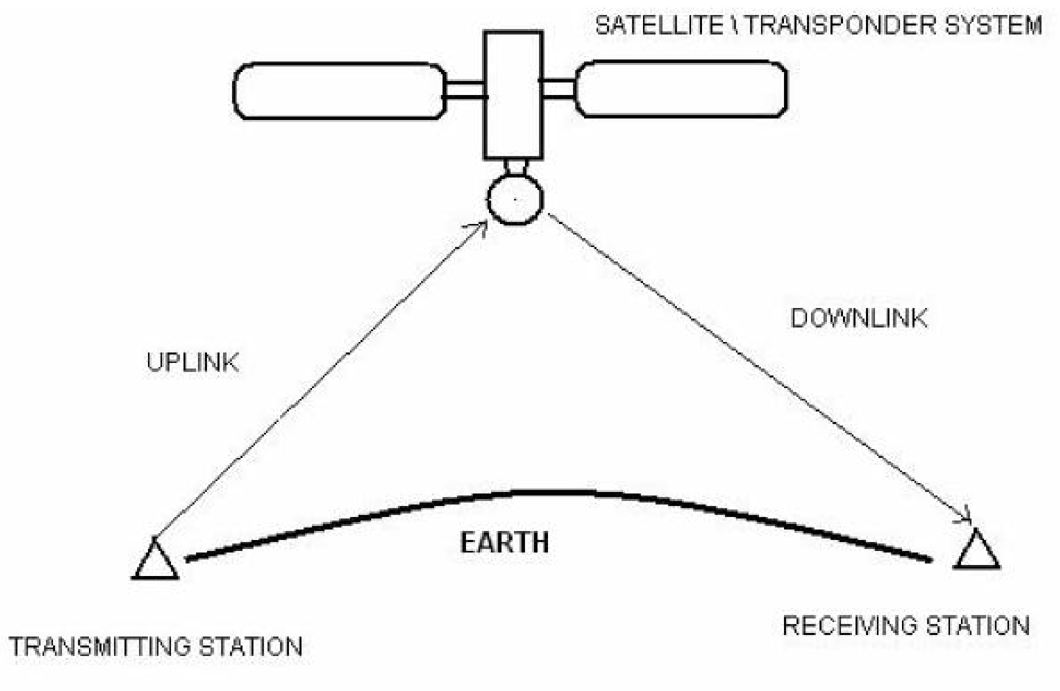
BASIC EARTH STATION - SATELLITE LINK MODEL

Broadcasters uplink scrambled or encrypted signals(audio and video) to a satellite at a particular frequency. The signals are encrypted to prevent their unauthorized reception before retransmission.

The signal received by the satellite is rather weak. Therefore, it is first amplified by the receiver antenna, filtered and further amplified by a low noise Tunnel Diode Amplifier (TDA). The output of the satellite is the frequency displaced and processed version of the input. This frequency translator is required to prevent interference from the high power satellite input to the output. All the tasks mentioned above can be thought of as being performed by a single device within the satellite namely, the TRANSPONDER.

After traveling 22,000 miles to a ground-based antenna, the signals are again very weak and must be amplified. Therefore, satellite "dishes" focus the signals onto the actual antenna. The signals from the antenna are then fed to a "low-noise block," or LNB, amplifier which amplifies signal and converts them to a lower frequency. The lower the power of the satellite, the larger the antenna required to focus the signals. A C-Band satellite, with power ranging between 10 and 17 watts per transponder, typically has an antenna between 5 and 10 feet in diameter; whereas a highpowered Ku-Band satellite, with a range of 100 to 200 watts per transponder, only requires an antenna 18 inches in diameter. The downlink path of the signal comprises of a receiving station, and a network of descramblers\decoders.

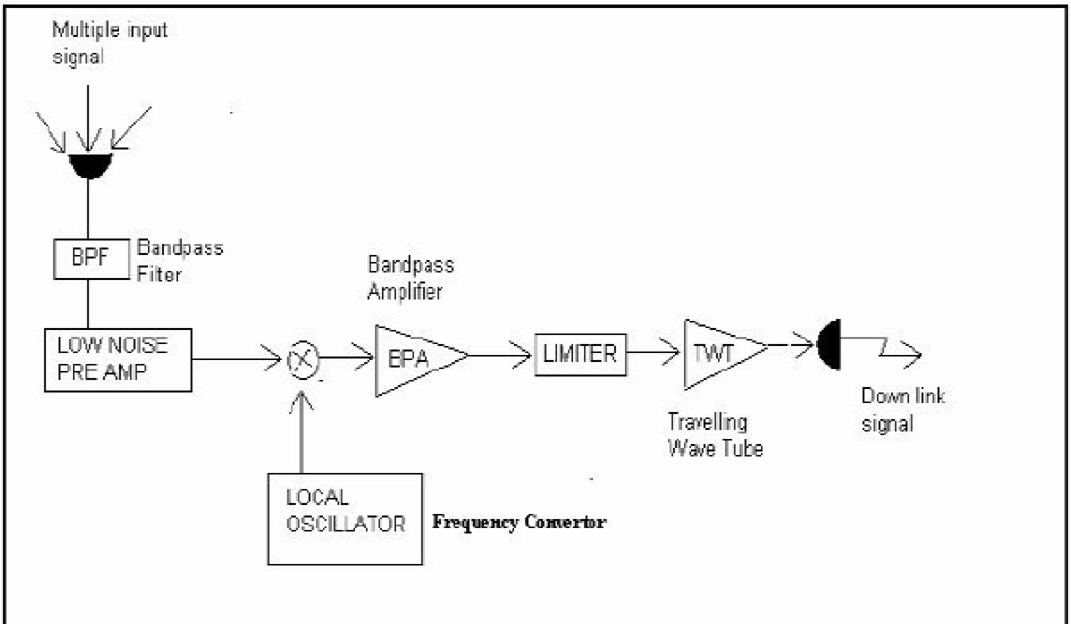
Each integrated receiverdecoder (IRD) is assigned a unique address code. The authorization center uses this code to turn individual decoders on or off and even selectively control large groups of decoders, thus enabling them to receive the signal.



Transponders

The transponder is essentially a receiver which receives the signal transmitted from the earth by the uplink, amplifies it and retransmits it with the downlink, with a different frequency. Thus the word "Transponder" is formed by combining the two words- TRANSmitter and resPONDER. Most satellites have anything between 10 to 30 transponders of different bandwidth on board.

Transponders can be either active or passive. A passive transponder allows a computer or robot to identify an object. Magnetic labels, such as those on credit cards and store items, are common examples. A passive transponder must be used with an active sensor that decodes and transcribes the data the transponder contains. The transponder unit can be physically tiny, and its information can be sensed up to several feet away. Simple active transponders are employed in location, identification, and navigation systems for commercial and private aircraft. An example is an RFID (radio-frequency identification) device that transmits a coded signal when it receives a request from a monitoring or control point. The transponder output signal is tracked, so the position of the transponder can be constantly monitored. The input (receiver) and output (transmitter) frequencies are pre-assigned. Transponders of this type can operate over distances of thousands of miles.



Simplified Transponder Block Diagram showing a single channel transponder

Modulation

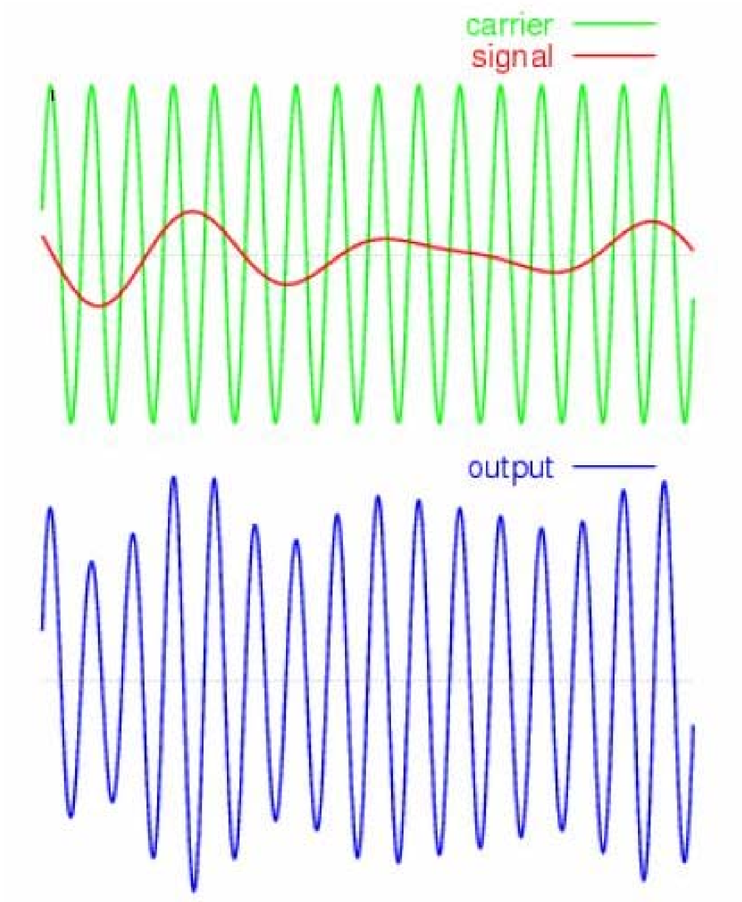
In the similar way as the human speech quickly loses its strength and becomes inaudible with the distance, so would become the transmitted signals. Also, the uplink and downlink signals may interfere with each other. This problem can be overcome by amplifying the signal but it also amplifies the noise that is present. The other solution would be to superimpose the information to be transmitted onto a higher frequency carrier wave. This process is commonly called as modulating the carrier. The signal which is being modulated is called the base band signal. The method used to modulate a signal is called Modulation. In other words, the base band signal is encoded by the carrier signal. It must however be noted that the initial modulation of the signal happens at the transmitting station and the final demodulation of the downlink signal occurs at the receiving station.

MODULATION TECHNIQUES

There are a variety of modulation techniques developed over the years, broadly they are classified into 6 categories, namely, amplitude, frequency, phase, delta, differential and pulse code. Here, we discuss the 2 most important of the above techniques, i.e. frequency and amplitude modulation.

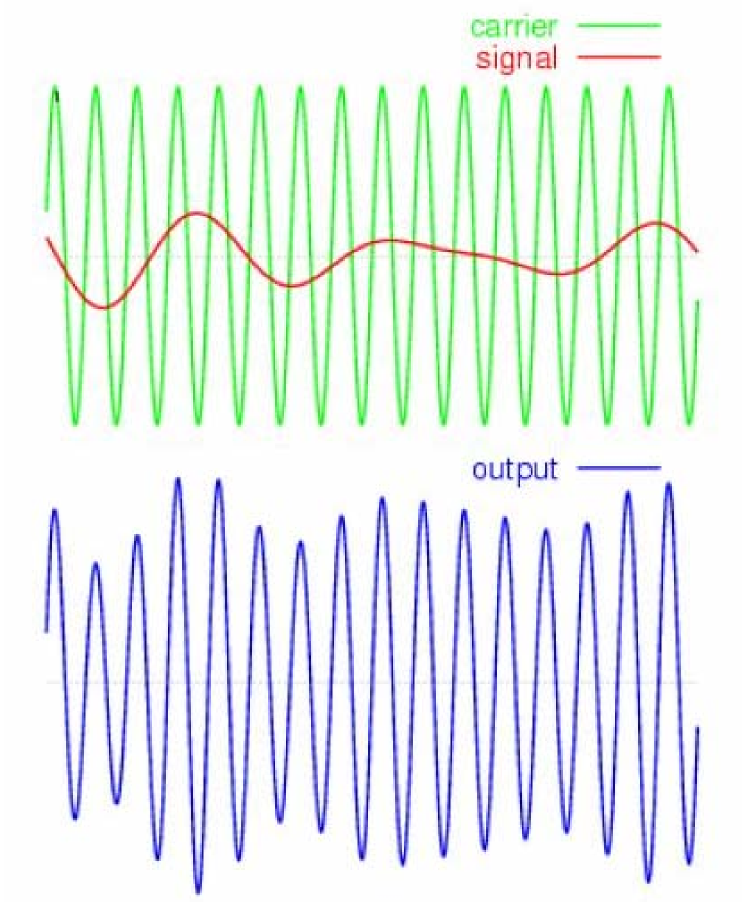
1. Amplitude Modulation :

In this technique, we vary the instantaneous amplitude of the signal by the combination with a high frequency carrier wave. The highest frequency of the modulating data is less than 10% of the carrier frequency. Instantaneous amplitude varies depending on the instantaneous amplitude of the modulating data.



2. Frequency Modulation :

In this technique we vary the instantaneous frequency of the wave.



Until now, frequency modulation methods are extensively used in all communication satellite systems. But this is changing quite rapidly as current satellites are replaced from their analog predecessors by employing digital transmission.

Advantage of Digital over Analog Satellite communication

1. Increased capacity in multi-access mode (using Time Division Multiple Access)
2. Economical advantages through increased capacity, more flexible operation and reduced production cost
3. More robust to interference
4. Compatibility : Since a digital bit string is processed as such independently as whether the source information is a colour TV signal, analog voice or a digital data.
5. Freedom to add new facilities and services
6. Transmission policy is almost independent of distance and network topologies
7. Direct low-cost interconnect with terrestrial microwave, cable and optical fibre systems.

SATELLITE V/S CABLE TV

Satellite TV offers a distinct technological advantage over cable TV systems: cable operations are essentially satellite operations in that cable networks receive their programming via satellite but then re-transmit the signals through the cable companies' trunk lines, which degrade the audio and video signals to a greater or lesser degree on the way to the customer’s home. Satellite TV, on the other hand, eliminates the intervening wiring and provides a direct link from the satellite to the customer’s antenna in a 100% digital video and audio feed.