

Digital Speech Processing
Prof. S. K. Das Mandal
Centre for Educational Technology
Indian Institute of Technology, Kharagpur

Lecture – 05
Human Speech Production And Source Filter Model

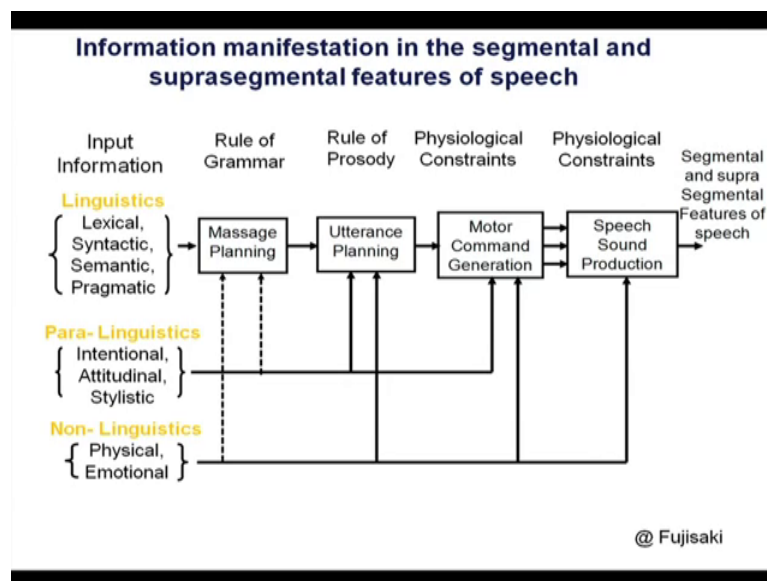
Good morning. So, let us start with that our start with our course that first class we have explained that introduction of speech processing that things, and second class recording part. Now let us start with a human speech production, and it is source filter modelling that. So, since the objective asks is to model the human speech production system.

(Refer Slide Time: 00:42)



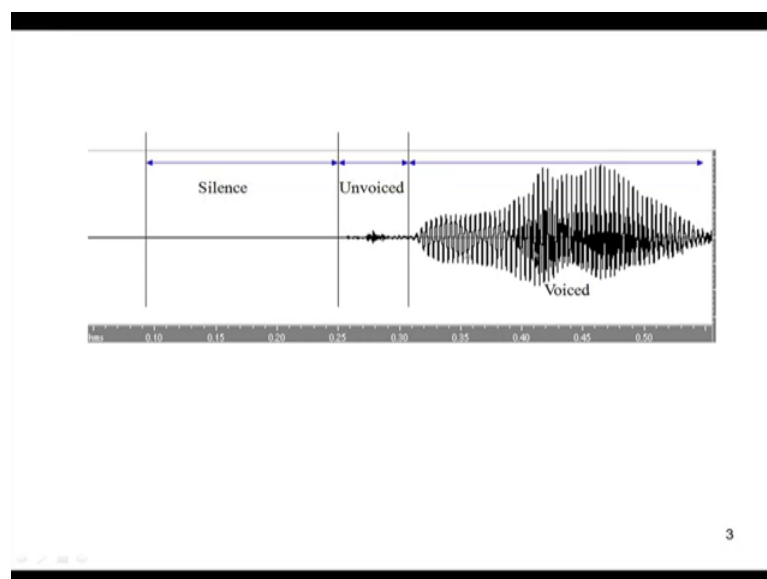
So, let us describe what is human speech production model and how it can be So you can say synthesize using the source filter model. So, as we explained that human speech production systems depends on some steps. First one is the message planning, that involve the linguistic parameter.

(Refer Slide Time: 01:00)



Then rule of prosody that utterance planning then the psychological constraint by the motor command generation then speech sound production, and radiation of the speech from the mouth ok.

(Refer Slide Time: 01:14)



Now, if you record a speech, and if you just display the speech it will look like this. If you see here, this portion I have not spoken anything. This portion let us I started spoken this is called unvoiced sound some noise is sound may be there, then there is voice sound is generated. So, normally in speech I can 3 part one is silence one is voiced unvoiced

another one is a voiced. So, silence means when I am not taking anything nothing is coming out from the mouth is a silence part. Then there may be a unvoiced aspiration will be there, or friction will be there, and there will be a voice sound.

Now, how do I produce this kind of sound? And how do I what is the meaning? How when I communicate a message? What kind of different voice sound and different silence and different unvoiced sounds I have produced? That you have to know that things. So, that is the speech production system all the different kind of sound is produced by the human being. Now if you take the basic definition, somebody will ask, you what is speech?

What is speech? Speech is a composed of a sequence of a sound, speech is nothing but a sound speech is not signal all not kind of things. So, human being use speech to communicate message from one person to another person k. So, it may be transmitted through electronics media on what kind of that is different business. So, speech is a composed or the sequence of sound, which is produced by the human vocal cords.

(Refer Slide Time: 02:50)


Basics Definitions

Speech is composed of a sequence of sounds
Sounds/Phonemes serve as a symbolic representation of information to be shared between humans (or humans and machines)

Arrangement of sounds is governed by rules of language (constraints on sound sequences, word sequences, etc)

Linguistics is the study of the rules of language
Phonetics is the study of the sounds of speech

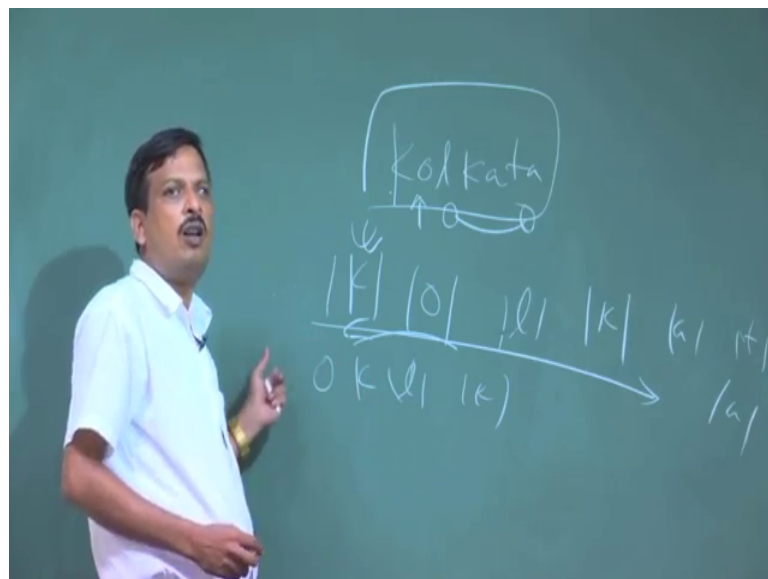
ET60007 © CET, IITKGP



Now, once I say sound, what is sound? What kind of sound? Sound or phonemes serve as a symbolic representation of information to share between human being. If I say I want to communicate a message from one human being to another human being, or human being to machine or machine to machine. In that case what I required a speech is a composition of different sound.

Now, what is sound? Sounds serve a symbolic representation of information which I want to communicate. That information has to be symbolically represent So that it can understand by the listeners. So, arrangement of sound or phoneme sound is called phoneme basic unit of sound is called phoneme. And it should be not haphazardly arranged, it should be arranged in it is some specific manner which governed by the language rule So that a listeners can understand. Suppose if I say Kolkata.

(Refer Slide Time: 04:06)



Kolkata, if I say Kolkata, let us there is a sound kolka [FL] and a. All the sounds are there. So, k is a sounds o is a sounds then la is lo is a sounds kol ka then again ka is a sounds then again a is sounds, then again ta is a sounds then again a is a sound. So, if those sounds are pronounced or composed in a sequence so that then a message the I am saying Kolkata is transmitted from talker to listeners. At those composition of sounds are governed by the language rule. I cannot compose kolka o k la la then k any arbitrary composition will not make any sense, although individual sound has an sense and this composition is governed by the language rule ok.

If I say Japanese, Chinese suppose I do not know Chinese, if some speaker spoken Chinese he is speaking the sounds in a some composition may be some of the sound I know, but I do not or I cannot identify the intended message he want to communicate.

Because we do not know the language rule by which the sequence of sounds are arranged, that is linguistics.

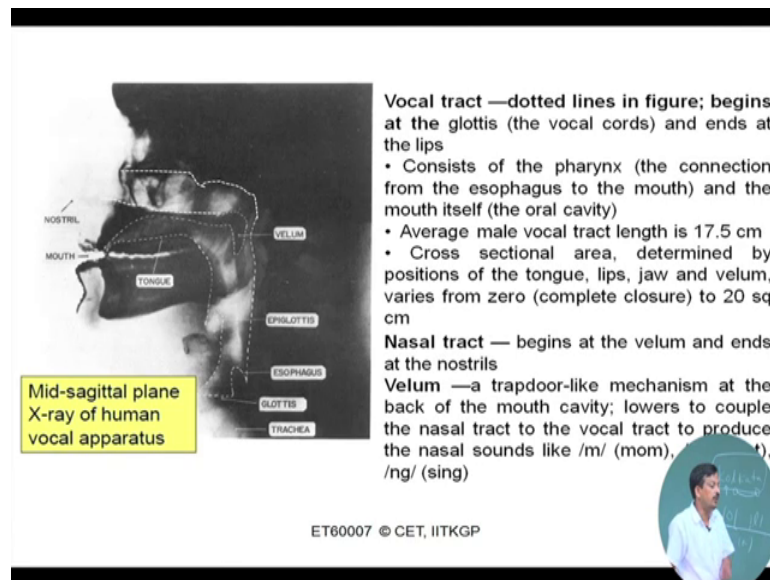
So, linguistics is the study of the rule of such of the rule of those rule, how do I human being produce different sequence of sound to communicate different kinds of message, there may be different kind of things. Different kind of you can say the different kind of rules are applicable for rules may vary from language to language. So, discipline which study those things is called linguistics. Then how the different sound is produced or study of the individual sounds is called phonetics.

Suppose ko this is a individual sound. So, sound of individual sounds that study is called phonetics, how the sounds are governed by the rule language rule for making the valid composition is a study of language that is called linguistics. Here I have explained the details of the phonetics. So, phonetics is a sounds, So it is produced by a human being in acoustics wave form.

So, I will cover the acoustics phonetics and articulatory phonetics once the sound is coming in the air it is acoustics. So, I have to know the acoustics part f the phonetics. And how we produce the different sound using this vocal tract is the articulatory phonetics. So, the phonetics has a 2 part, one is called acoustics phonetics another is called articulatory phonetics. So, I will details cover those things, because we have to understand the sound, but composition of sound it is a study of linguistics.

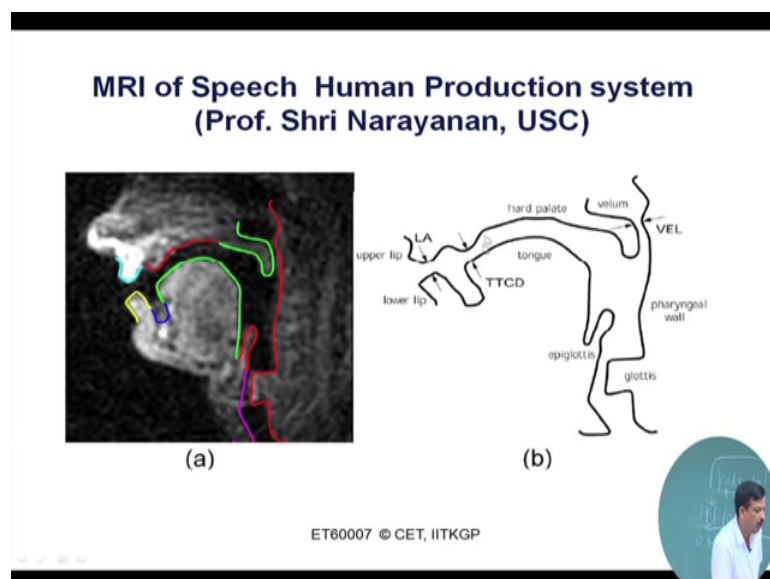
So, there may be a phonology, there may be a language rule grammar, and there may be a some kind of arrangement pragmatic paralinguistics linguistics all kinds of rules will be there ok.

(Refer Slide Time: 08:03)



Now, let us look how the human vocal tract look like this is the x ray of human vocal this whole, whole systems x ray whole systems there is a vocal cords and there is a tube, articulatory tube whole system x ray is look like this.

(Refer Slide Time: 08:19)

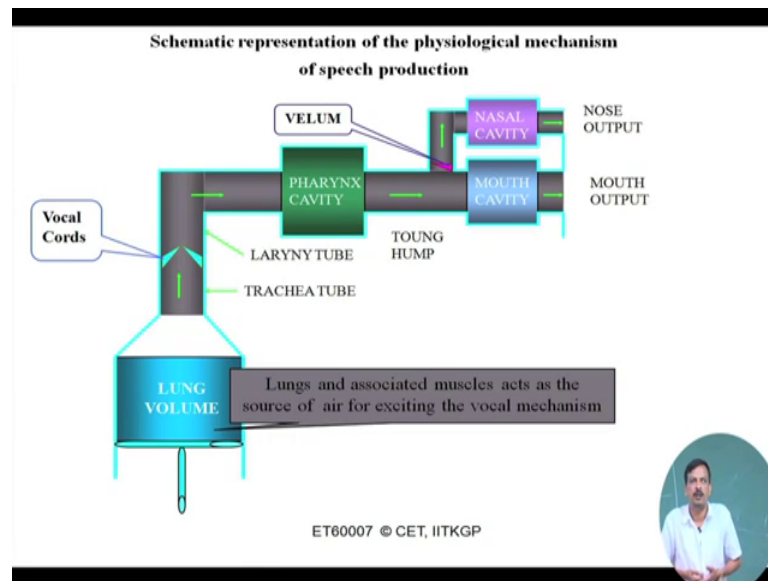


If I take the MRI this is taken from professor shri narayanan. So, if you see that MRI of the human vocal tract which look like this. This is a lip, then there is a tongue, whose lip tip of the tongue is open back portion is closed. Then there will be a vocal cord here. So, if I schematically represent this thing it will look like lip, then tongue tip of the tongue is

open and back with a the bottom is closed bottom is fixed with a lower jaw. Then this is a palate, then there is a velum because we have a nasal cavity, and we have a oral cavity.

So, velum may be closed or velum may be opened. Then there will be a pharyngeal walls this (Refer Time: 09:04) there will be a glottis section ok.

(Refer Slide Time: 09:11)



Now, if I take the schematic representation of the speech production, system it will look like this. How the speech is produced by human being? This is schematic representation. So, there is a lungs. So, what is the source of the sound, if you see forget about the speech, basic source of sound is nothing but a vibration if I want to generate a sound, sound is nothing but a mechanical oscillation. So, it is nothing but a mechanical oscillation.

So, a mechanical body is vibrating, and sound is acoustic waves is generated. Now vibration of mechanical body required external force I required a force to vibrate the body. Now if you see if I have membrane connect a membrane if I pass the air, then thin membrane is closed, and if I pass air in force the membrane will vibrate and sound will be produced. In childhood if you see the kite who is flying in the sky and there is a some membrane is attached and membrane is vibrating in the air and producing a sound, this childhood in India people are used that things.

So, there is a source of you can say the source external force is required. So, during the speech production the lungs when I take the air inhale, and I gradually exhaled that air and produced different sound by constructing the different location of the tube. So, lungs produce or give the air pressure which is required to the production of sound. So, this is the force. And once this lungs press the air to upwards there is a vocal cords.

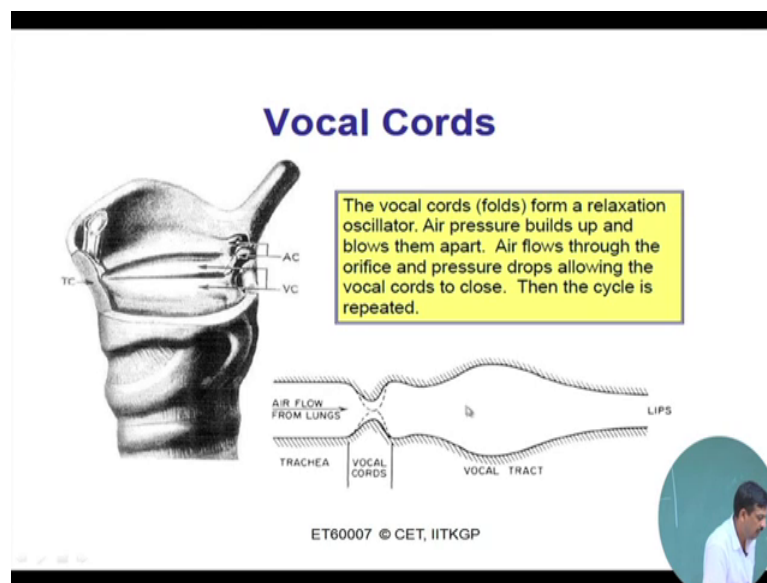
So, vocal cords are generally open when we take inhale. If I want to speak something the vocal cords are has to be closed if it is open no sound. So, sound there is a no sound is coming from the mouth. So, if it is vocal cord is closed then when the air is passing through the vocal cords it is a membrane and if the membrane is exposed to a high velocity of air, that that create the vibration in the membrane. So, that vibration cause the sound and pass through the cavity.

Now, if I see tongue we have a tongue. Tongue separate the cavity in 2 part, one is called back cavity one is the front cavity. And we have a velum if you see the velum either velum can be closed, or velum can be opened. If velum is open, if velum is open and oral cavity is closed then the air will passes through the nose. So, nasal cavity is involved if the velum is closed, the sound is only passed through the mouth that is called oral cavity. So, if there is a that is a 2 kind of sound, either velum may be closed or open if velum is opened. Then the sounds become nasal because nasal cavity involved if the velum is open closed, then nasal cavity is destruct from the cavity then it is called oral sound.

So, if the sound is produced using his mechanism lungs. Create the pressure goes upward if the vocal cords are closed, then the vibration will produce and that vibration create the sound and that sound is modified. Depending on the tube structure when it is passes through the tube. So, I can say in human speech production system has a 2 part. One is the source here, the source of sound is vocal cords. And lungs only the provide the pressure.

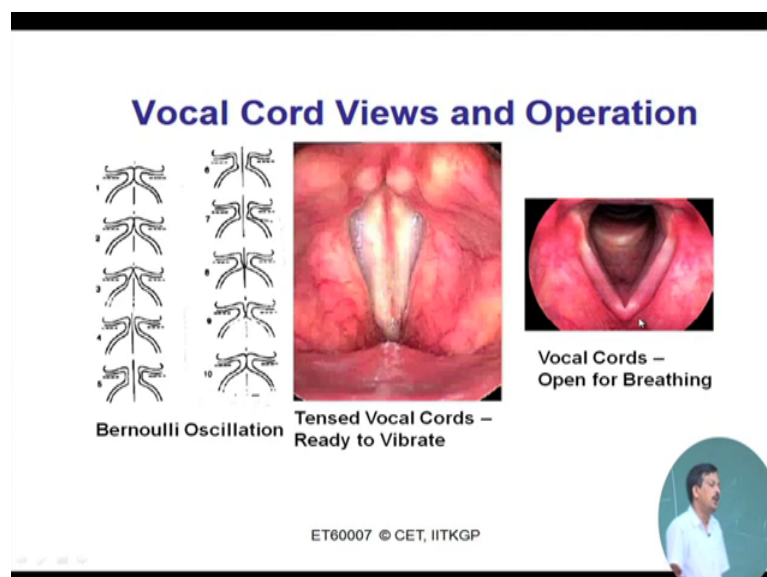
So, one is the source, and once the sound is produced that source is passes through the cavity or you can the tube this tube. And it is modify to create the different count kinds of side sound. Once a source create the sound when it passes through the cavity depending in the structure of this tube, it is modify and it produce different kinds of sound. So, I have a source and I have a tube this is a 2, 2 part of the sound production.

(Refer Slide Time: 13:50)



So now go goes to the source what is how the what is vocal cords how it is look like. If you see the vocal cords is look like this picture. Now if you see that air is coming from the lungs if the vocal cord is closed, then the air the air is flow what flow will be obstructed and the velocity of the air will be more will be defined different because air pressure in here will be increased air pressure will be decreased. Once it is little bit of open the air will passes through this vocal cords and create the vibration ok.

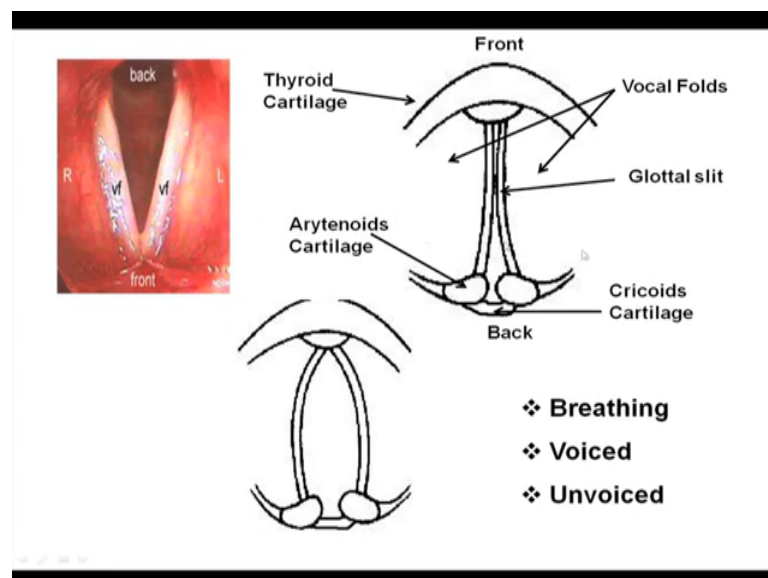
(Refer Slide Time: 14:24)



So, that is the sound this is the original picture, or you can say that the endoscopic picture of the vocal cord. This is the closed position of the vocal cord this is the open position of the vocal cord. So, this is the vocal cord structures, this is the closed position, this is the open position.

Now, if you see the vocal cords are closed it is not like a this side wave movement open end of the vocal cord is always fixed. And it is closed and open in the other end, and that creates the sound. So now, if I take the schematic structure of this, this will look like. So, different cartilage and closed this either.

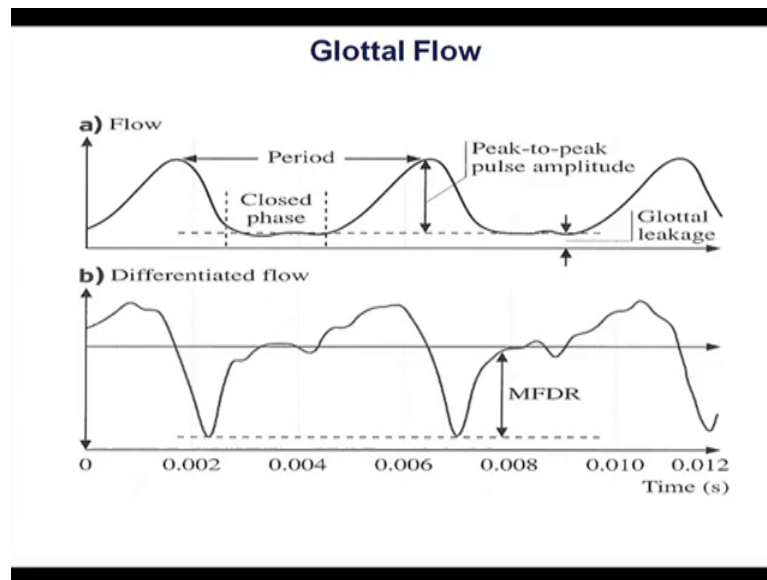
(Refer Slide Time: 15:03)



So, movement of the vocal cords depends on the muscle movement. Now one end is closed, this end vocal cord is closed and this end is the back end is open ok.

Now, once either it can be closed either or it can be totally opened. So, there is no sound breathing, if it is closed worst, if it is slightly it is open, but the tube can produce a noise kind of sound that is called unvoiced sound. So, unvoiced voiced and silence region I have seen in the first slides in here.

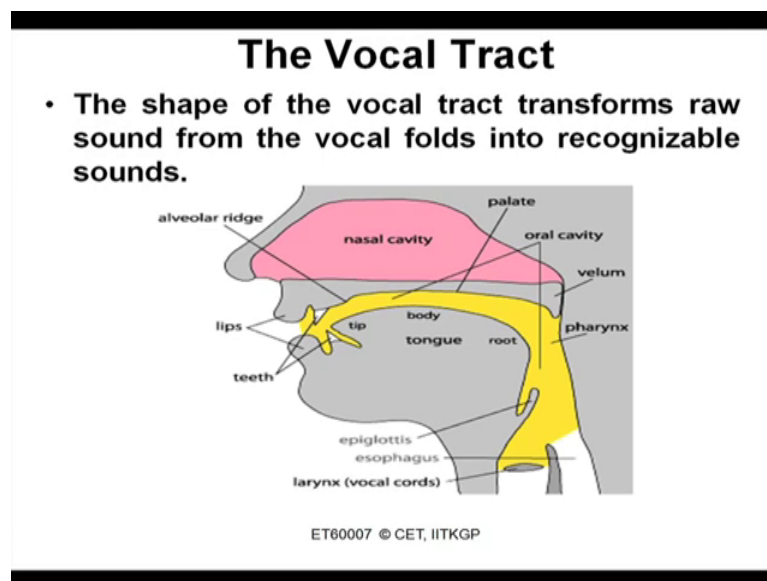
(Refer Slide Time: 15:40)



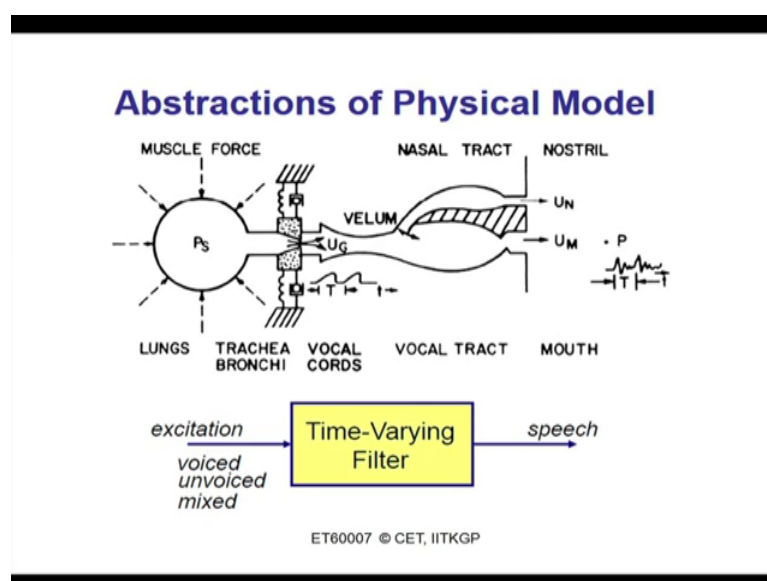
Now, if I see the glottal flow, how the air is flowing in the glottis. So, once the vocal cord is closed, if the vocal cord is closed no air can flow. So, either it is gradually open and closed open and closed open and closed. Once is closed air flow stopped once is gradually open one flow increases, increases, increases, increases and again this decreases, decreases. So, I can say this is the flow diagram air flow is increases, increases, increases and again decreases and closed phase, again it is open and again it is closed. So, I can say either complete open to open is a period or closed to closed is a period and if I take difference of the flow which create, So difference of the flow.

Will create the sound vibration. So, differentiate if I differentiate this thing it will be look like this which create the vocal cord vibration. So, this is the differentiation of the flow now if you see that this tip to tip is a period either closed to closed is a period or open to open is a period.

(Refer Slide Time: 17:00)



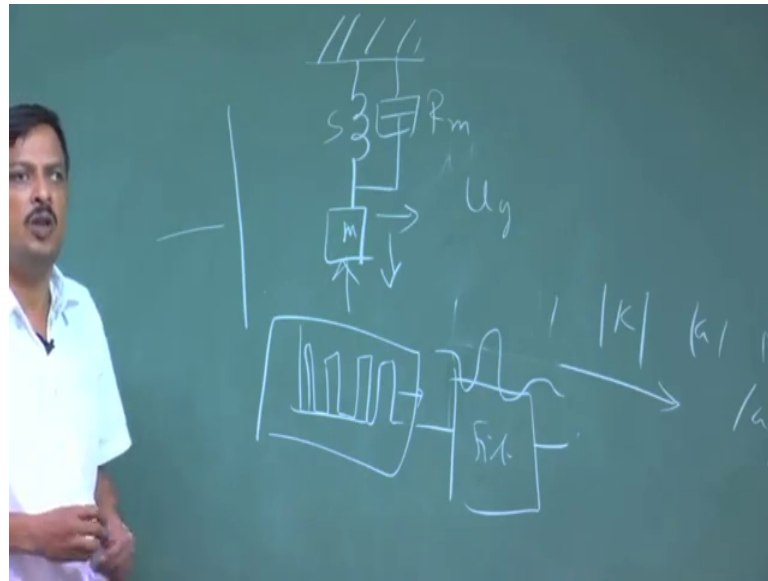
(Refer Slide Time: 17:04)



So, this is source, I can show you the diagram in here this is called engineering model of the source.

So, how do how what is the engineering model. If you see this is the lungs. Now vocal cords any mechanical vibration is nothing any sound production of the sound is nothing but a mechanical vibration.

(Refer Slide Time: 17:24)



So, what is the form of a mechanical oscillator? There is a spring there is a mass and there is a mechanical damping. This is the mass this is the spring constant s and this is the mechanical damping R m ok.

So, air act as a force on this mechanical oscillation. So, I can say 2 vocal cords can be act as a mechanical oscillator, and that oscillate by the force which is coming from the lungs. And that is U_g , if you see the U_g mechanical oscillation produce the acoustics wave. So, acoustics wave has a volume velocity and is a pressure wave. It is has a volume velocity. So, it is a volume velocity U_g particle velocity, and probably you can multiply a for a fixed unit fixed volume is a volume velocity ok.

Now, U_g is the air flow velocity volume velocity. Will come through this different vocal tract shape. So now, if you see that whole chart, the source is nothing but a mechanical oscillator, it create the vibration. Once it mechanical oscillator it is create the vibration lest it has create the vibration is come up impulse. So, source can be modelled using a impulse source. Now this has to be passed through this tube, this is a tube.

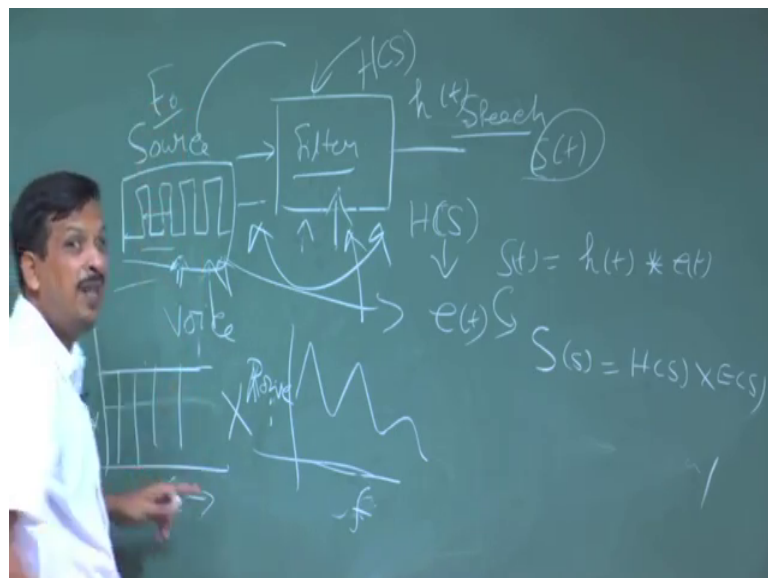
Now, tube can form different structure using the tongue lip and velum movement. Now if you see I can move the tongue in upward I can back off the tongue can be raised, tongue tip can be touches in the upper palate. So, when you speak if you see the upper palate is fixed. Lower palate upper this upper portion is fixed upper palate is fixed. Now in lower jaw upper jaw is fixed, and lower jaw is moving if you see. So, I can move the either lip,

I can close the lip I can open the lip or I can move the tongue to divide the cavity in different structures.

So, I can say when this vibration passes through a different kinds of tube structure can produce different kind of sound, cascading tube. So, if I consider that things then I can say this is the source of the sound production. And that source sound is passes through a different structure of the cavity and different time to produce the different kind of sound sequence. So, I can say that source is a impulse response that can be passes through a filter depending on the filter structures different sound can be generated.

So, in engineering model or human production system if I see it is as source filter model.

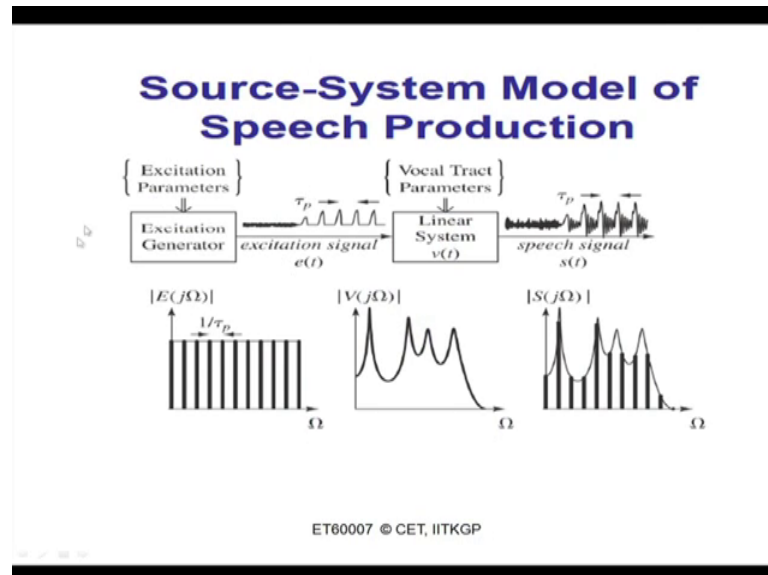
(Refer Slide Time: 20:52)



So, vocal cord is producing the source which is nothing but a impulse, either sound is present or sound is absent either source can be present or source can be absent. Then it is passes through a filter, this is time varying filter. Different time structure of the tube will different to produce different kind of sound, let us this is s n or this is speech. So, digital circuits I will come later on. So, I can say whole human speech production system, can be modelled using a source filter model. So, in can say I can study the source or I can study the filter, combine the both I will get the speech. So, who produce different kind of sound, depending on the structure of the filter. Source either source vocal cord can vibrate or vocal cord without vibration. So, either source can be present, that is voiced sound either source can be absent, that is called unvoiced sound. Or voiced sound is

produced on it passed through the filter it can produce the different kind of voiced sound different kind of voiced speech ok.

(Refer Slide Time: 22:39)



Now, if I see the I will come this one again I will come when I talk about the tube modelling, but here I will explain. So, excitation So impulse is nothing but a excitation parameters, excitation is generated and that excitation part of the going to the linear time varying system filter and I get the speech. So, this filter I have to model this excitation part I have to model. Now if I say what is the properties of this source, if you see if it is a impulse then it has a period impulse is a period. So, which is the fundamental frequency of the source human being also has a fundamental the speech is a quasi periodic signal it is not exactly periodic. So, human being also produce different kind of this source fundamental frequency depending on the requirement.

When you speak, we never speak in a single fundamental frequency, that will become to mechanical to hearing. So, fundamental frequency are gradually moving to provide the melody on the speech. Think about singing. Filter produce the different sound for content purpose, but melody is defined as a source itself. If you singing think about the singing, the control of the fundamental frequency sa re ga ma the all the notes are nothing but the ruminant of the fundamental frequency. So, those controlling of the fundamental frequency is generated by the source.

So, movement of the vocal cords. I will discuss how this vocal cord movement, how the vocal cord produce different fundamental frequency, to by varying the tension of the vocal cord and position of the vocal cord. So, that is I will discuss in later class in during the prosody modelling.

So, movement of the vocal cord tension can change by 2 types, one is called forward and backward movement one is vertical movement. So, if the source of fundamental frequency is the property. So, fundamental frequency is a property of the source, and filter responsible for producing different kind of sound. So, if you know that digital filter, how the digital filter is characterised depending on the frequency response ok.

So, I if I say the digital filter is represented by h_s you know the Laplace transform h_s or you can say the h_n whatever h_s represent the frequency response of the filter. So, if the frequency response of the filter is look like this, and the source is passing or the impulse response contained all the fundamental frequency, all the frequency let us same height and this is the you can say time and this is the frequency sorry, the this is the this sorry this is the frequency and this is the power. This is the frequency, this is the power.

So, as per the signal processing model, when the source is passes though a filter source signal will convolved with the filter signal to produce this speech. In frequency domain convolution is nothing but a multiplication. So, if I say in signal literature of signal processing if I if I apply here, if I say source is nothing but a e_t and filter is nothing but a h_t . So, output speech is s_t . So, s_t is nothing but a h_t convolve with e_t , convolved with e_t . If I take the Laplace transform in frequency domain or z transform let us take the Laplace transform then s , s Laplace domain is nothing but a h_s multiplied by s .

So, frequency response of the filter, will be multiplied with the frequency response of the source and produce the speech. So, different sound is responsible by the filter, source is responsible either there will be a voicing or there will be a no voicing. Either there will be a voicing signal or there will be a no voicing. So, response. So, this is nothing but a multiplication ok.

But source provide the fundamental frequency of the speech. So, in case of singing when you say sa note [FL]. So, actually you are controlling the fundamental frequency of the vocal cords. So, the controlling of the fundamental frequency which is I will talk about

the prosody modelling part, I will details mathematical explanation has given that how this fundamental frequency of the vocal cords can be changed.

Now, if you see the singers are practicing in morning to touching up the some or you can see those comment like that, you are you are upper note is note that clear lower note is clear. What is meaning is that the changing of the fundamental frequency, from lower to high or high to low. So, upper note and lower note that changing or controlling of the this vocal cords is importance.

So, you are practicing controlling of the vocal cords by (Refer Time: 28:31). So, that you can touch the perfect note during the speaking also, we change this vocal cord fundamental frequency with respect to time to provide the melody on the speech ok.

(Refer Slide Time: 28:49)

Women and Men

- The acoustics of male and female vowels differ reliably along two different dimensions:
 1. Sound **Source**
 2. Sound **Filter**
- Source-- F_0 : Depends on length of vocal folds
 - Shorter in women** \Rightarrow higher average F_0
 - Longer in men** \Rightarrow lower average F_0
- **Filter--Formants**: Depend on length of vocal tract
 - shorter in women** \Rightarrow higher formant frequencies
 - longer in men** \Rightarrow lower formant frequencies

Now, some fact if you see. So, sound source and sound as a filter. So, acoustic the acoustics of male and female vowel differ reliably along 2 different dimensions.

So, sound can be this s t is depends on the source and depends on the filter. Source if it is voiced sound source is exists and source provide the fundamental frequency, and filter verify that source and produce different voice sounds, that may be vowel or that may be a voice consonant also. So, it is not necessary it will be vowel it may be consonant voice consonant also.

So, for force modify by the filter producing different kind of sound. So, fundamental frequency or F_0 sometimes we say pitch F_0 is sometimes defined as pitch, but pitch is a perceptual elements in parameter F_0 is the physical dimension of the parameter F_0 can be measured, but pitch is perceptual pitch cannot be measured. So, F_0 of the source depend on the source and different sounds is depend by the filter. So, if I say the human being different kind human will a male versus female male versus female.

If you see F_0 depend on the vocal cord length, vocal cord mass those 2 things if you see the vocal cord is closed, this end if the length of the vocal is increases what will happen? The fundamental frequency will be decreases. If the mass of the vocal cord is increases fundamental frequency will be decreases, if you see the drum huge membrane drum. Mass length is area is increases fundamental frequency is decreases, if you take a short membrane sound dung fundamental frequency will be increases.

So, for a woman since the vocal cord length is shorter, their fundamental frequency is higher. If you see the woman singer they say the yours best scale is be sharp, what do you mean by be sharp? Average fundamental frequency in the level of the average fundamental frequency. So, if it is woman then the higher the F_0 f_0 average F_0 will be very high not compared to male, but for a male since the vocal cord length is longer the fundamental frequency is shorter means average is lower.

So, average fundamental frequency for a male is lower than the average fundamental frequency of female then filter. So, filter depends on the characteristics of it is frequency response. Now if I say $h(s)$ is a transfer function of the filter if you are electronics students you know electronics or electrical students or you know signal processing, then $h(s)$ is nothing but a system functions. So, that systems depend on what system has a property some pole and 0. So, from there I will define that formant frequency in the next class. So, that is called formant or resonance frequency based on the resonance frequency different sound is produced.

I will explain in the next class, what do you mean by pole and resonance frequency. So, those is for speech it is called formant. So, in case of woman since the length of the vocal cord is shorter the formant frequency is little bit of higher, longer in case of man. So, formant frequency is little bit of lower. So, that is difference between the male and female speech. If it is child the fundamental frequency much more higher because of the

length of the vocal cord is much shorter, and the formant frequency will be higher because the length of the vocal cord is shorter. So, I can say that formant frequency depends on length of the vocal tract, and also other some properties we will discuss later on when you module this local tract.

So, if the length is increases, the formant frequency will be different. So, I can say if average if I can say personalised the speech means my vocal cord or vocal this length of this vocal tract or you can say length of this tube, may not be exactly equal to other person. So, if it is not equal then I should have a formant structures which will be different from the other people. So, there may be some information exists from which I can identify the person. So, I can exploit that information, to recognise the biometry or biometric signature of the person.

Similarly, source vocal cords my vocal cords length and weight will be different from somebody else. So, there may be a some parameter exists here also by which I can identify the persons. So, biometric signature may be exists in the speech either in filter portion or in source portion. Or if I say speech is also depends on the composition of the sound sometime this composition pattern also provide some information which can be exploit to detect the person. So, next class I will discuss what is formant, how it is extracted, those things will be discussed ok.

Thank you.