Indian Institute of Technology Kanpur

National Programme on Technology Enhanced Learning (NPTEL)

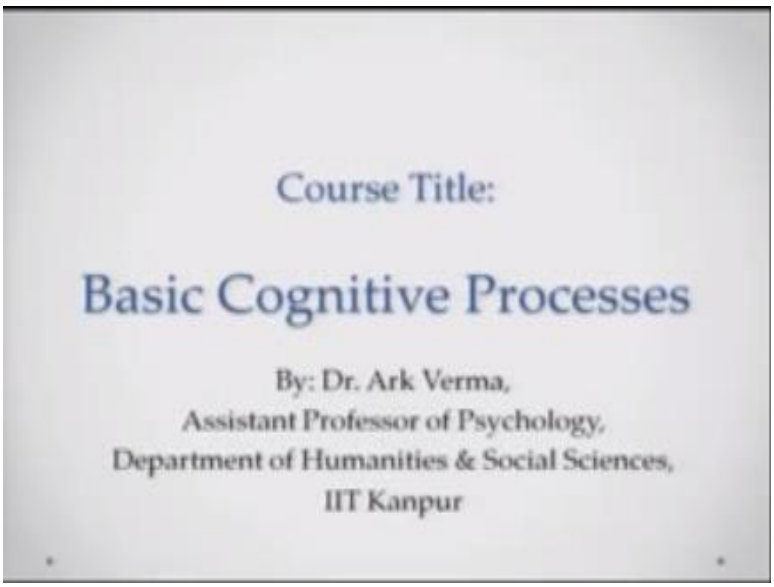
Course Title
Basic Cognitive Processes

Lecture- 23
Audiotory Preception

by
Prof. Ark Verma
Department of Humanities and Social Science
Indian Institute of Technology Kanpur

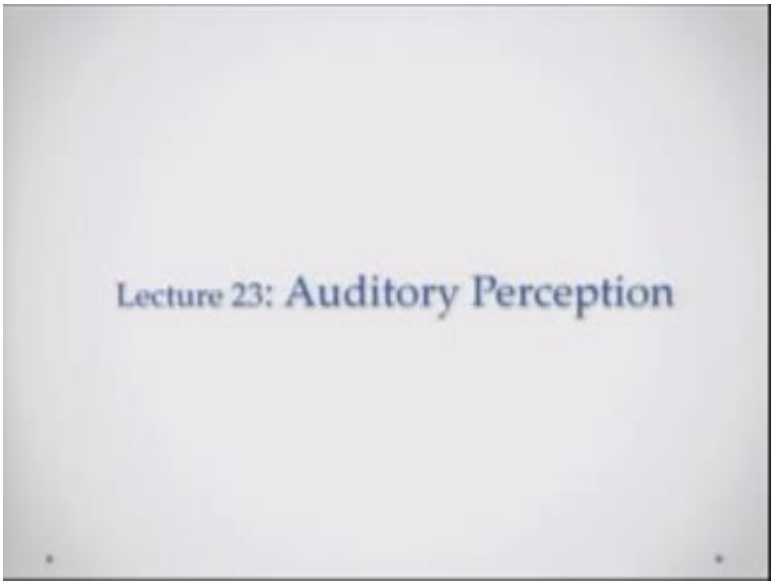
Hello and welcome to the course basic cognitive processes I am dr. Ark Verma from IIT Kanpur.

(Refer Slide Time: 00:17)



Today I will be talking about auditory perception.

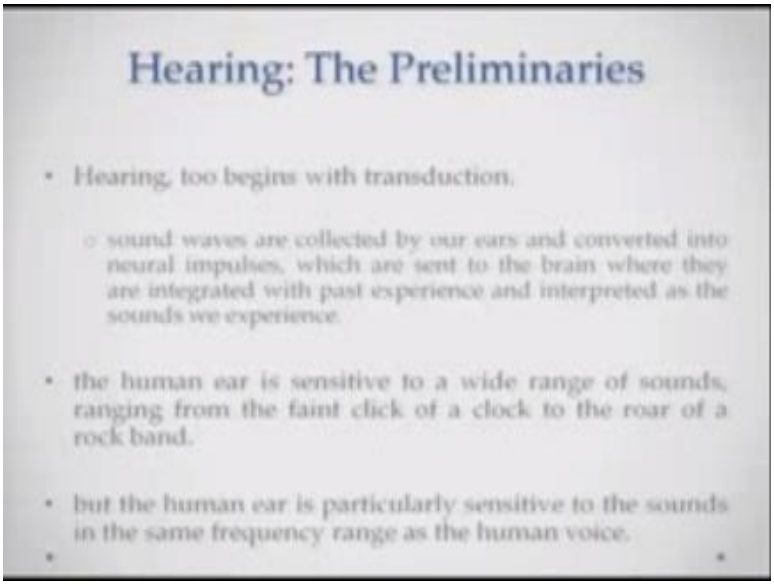
(Refer Slide Time: 00:21)



You have seen that we have talked about visual perception till now, we have talked about the different aspects of visual perception you talked about how visual perception can be used to recognize objects how it can be used to interact with the environment and we have seen the different modalities and theories about visual perception itself. Today I will begin talking about auditory perception.

Now auditory perception is basically hearing okay, and hearing also begins with transduction as we saw in the case of visual perception any sensory input that is coming via the senses to the brain that needs to be converted into a representation which the brain can handle and that process is called transduction as far as auditory perception or hearing is concerned sound waves are collected by our ears and they are converted into neural impulses which are then sent to the brain.

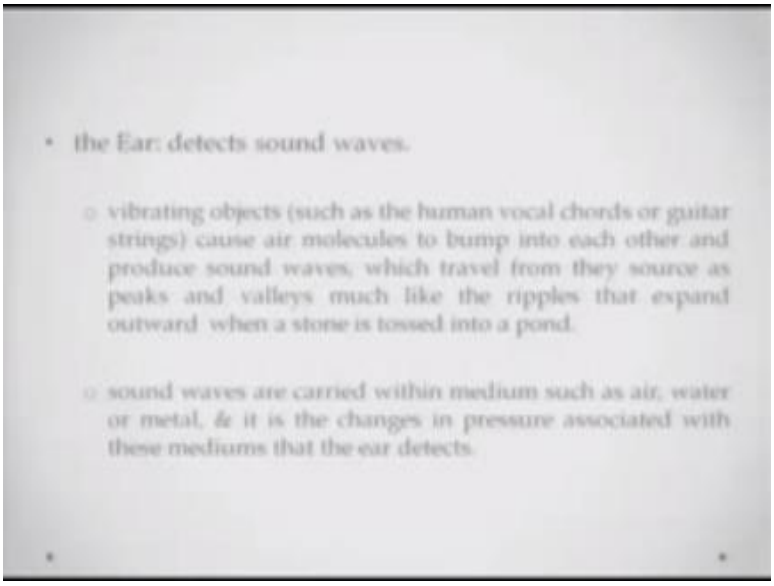
(Refer Slide Time: 01:12)



Where they are integrated with the past experiences, whatever knowledge about those kind of sounds you might already have and then that information is used to recognize or understand whatever that sound is supposed to convey. Now the human here is the sensitive to a wide range of sounds ranging from a faint, click of a clock to the roar of a rock band but the human ear is most especially sensitive to the range of frequencies which coincides with the range of frequencies of human voice and that you can understand is the evolutionary advantage.

Because the most important stimulus will be hearing in your life things is people talking and in that sense you will use that to understand most of what their environmental input will be about.

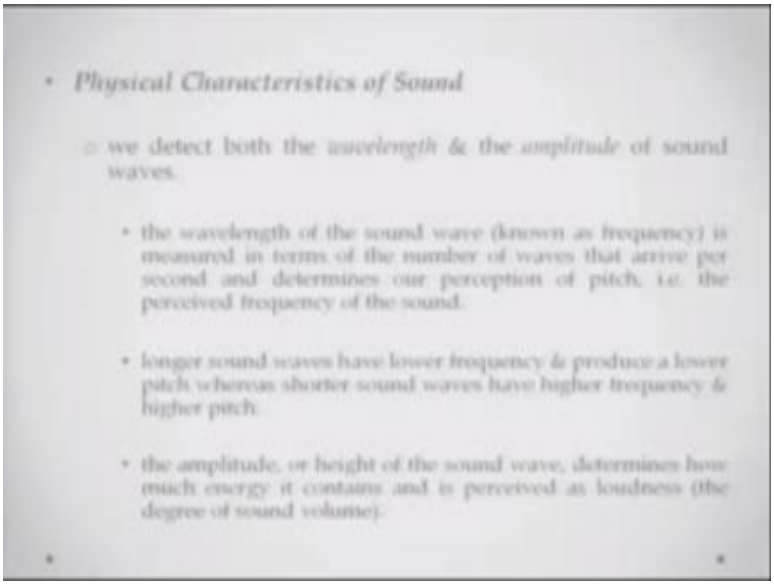
(Refer Slide Time: 01:54)



Now the ear it is basically an organ that it takes sound waves much as the eyes detect light energy, now sound are generated by vibrating objects say for example if somebody is talking their vocal cords are vibrating or say for example if you pluck a string of the guitar produces the sound that you hear coming out of the guitar. Now this vibration basically causes these air molecules to bump into each other and produce what are called sound waves and these sound waves travel from the source of the vibration in peaks and valleys like ripples and they expand outward.

You can actually make this look like, say for example if you are putting a stone in a pond of water the stone where the stone has felt will create those kind of ripples and it will be expanding outwards that is much similar to how sound waves really propagate. Now these sound waves need to be carried within a medium such as an air or water or a metal and it is the changes in the pressure associated with these mediums that your ear is able to detect.

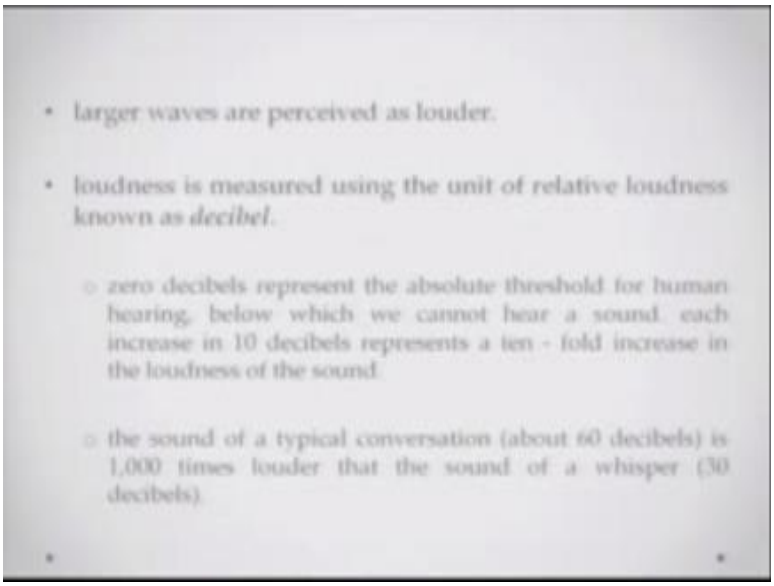
(Refer Slide Time: 02:57)



Now let us talk a little bit about the physical characteristics of sound there we detect both the wavelengths and the amplitude of these sound waves, now the wavelength of a sound wave also known as frequency is basically measured in the number of ways that are generated per second this wavelength also determines what is called the pitch of the sound or the chillness of the sound that is the perceived frequency of the sound.

The longest audience will have lower frequency and will produce a lower pitch there a shorter sound waves will produce higher frequency and have a higher pitch. Now the amplitude or the height of this sound wave will determine how much energy it contains and in that sense it is perceived by you as the loudness of that particular sound.

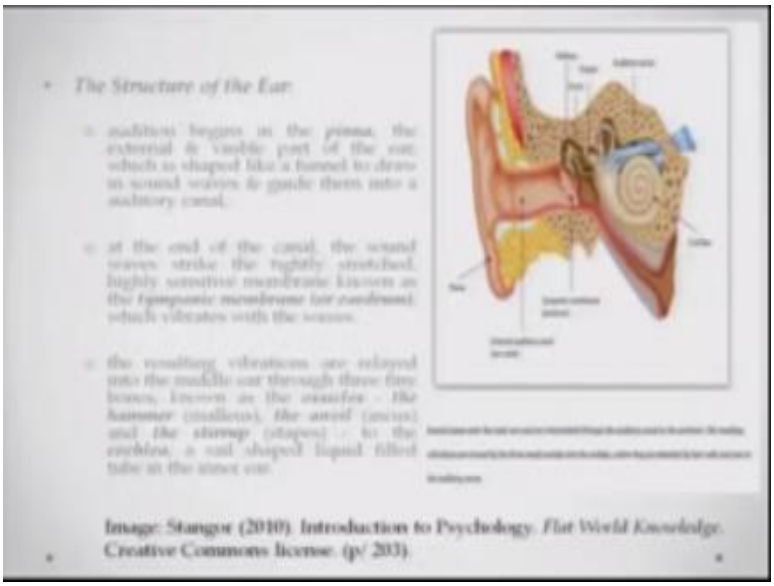
(Refer Slide Time: 03:39)



In that sense larger waves are perceived as louder waves, now loudness is your experience of whatever sound you are hearing and this experience is quantified in a unit known as decibel, 0 decibel basically is about the absolute threshold of human hearing so anything below 0 decibel you cannot hear anything above 0 is when you start hearing. Say for example, a typical conversation might be around 60 decibels the sound of your breeding is around 10 decibels and so on and so forth.

It basically increases in10 volts you have to multiply the quantity with 10 every time you are kind of moving up by 10 decibels.

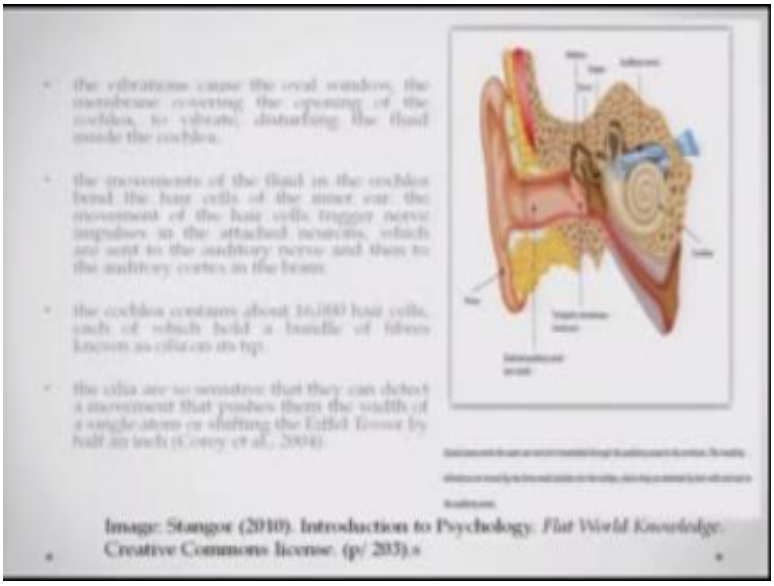
(Refer Slide Time: 04:22)



Now let us concentrate a little bit about the structure of the ear so audition or here begins in the external ear or the pinner which is the visible part of the ear. Now you can see that this external ear is shaped like a funnel and this funnel is used to draw the sound waves from the outside into what is called the auditory canal. Now this auditory canal at the end of this is a sensitive membrane called the tympanic membrane or the eardrum which receives these vibrations from the external environment.

Now these vibrations when received at the tympanic membrane are conveyed through the middle ear through three tiny bones which are called ossicles and these bones are namely the hammer the anvil and the stirrup these convey these vibrations to the snail type structure you can see which is the cochlea which is the innermost hear.

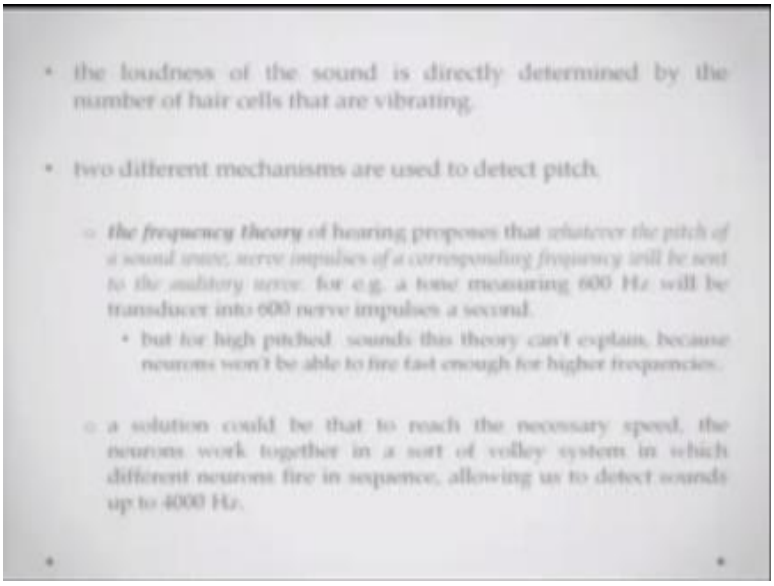
(Refer Slide Time: 05:12)



The vibrations cause this oval window of the membrane of the cochlea to vibrate and which deserves the fluid collection inside the cochlea. The movements of the fluid inside the cochlea bends what are called hair cells or the inner ear, the movement of these hair cells trigger neural impulses so here you can see how the vibration is being converted into neural impulses and then they will be connected conveyed to the auditory nerve which is connected to the brain which will do the further processing of whatever sound that is received.

This cochlea contains around 16,000 hair cells each of which holds a bundle of fibers called cilia on its tip and the cilia are so sensitive that they can detect a movement that even pushes them as slightly as by the width of an atoms even if you push these cilia a little bit as much as the width of an atom they will be able to detect it and respond to that. Now the loudness of the sound basically is directly determined.

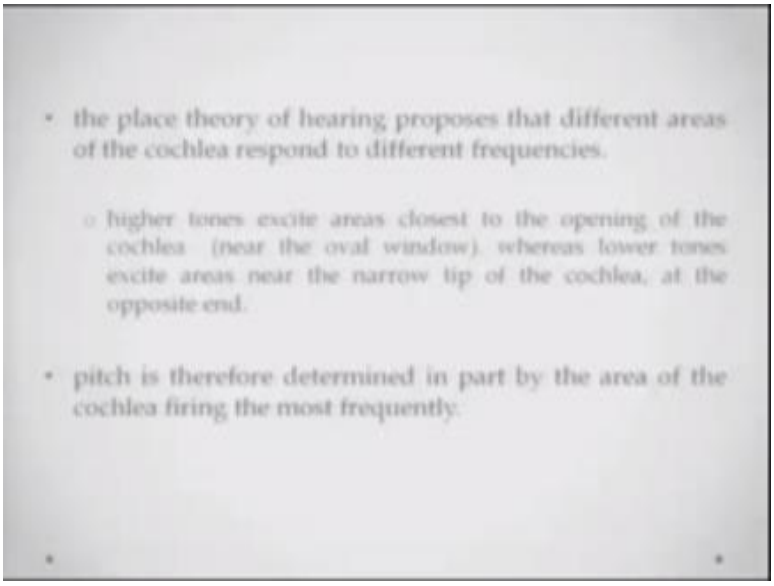
(Refer Slide Time: 06:13)



By the number of hair cells or number of cilia that will be vibrating and these there are two different mechanisms that have been proposed to detect page. One theory or the frequency theory proposes that whatever the pitch of the sound wave a proportional number of you know a proportional corresponding frequency will be sent to the auditory, say for example if a tone measuring 600 Hz will be transduced it will be translated into 600 nerve impulses per second.

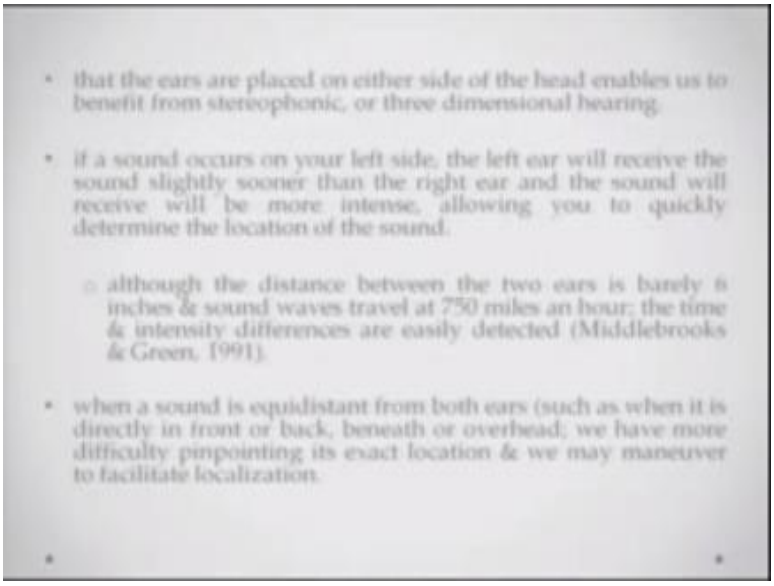
But this is slightly impractical say for example if you have a sound which is of much higher pitch then the celia will not be able to convert those kind of neural impulses. So alternative solution is proposed that maybe to reach the necessary speed then you also work together in some sort of volley system in which different neurons fire in sequence allowing us then to detect sounds up to 4,000 Hz. So there might be you know sections of neurons vibrating at one point and in the other section and so on and so forth.

(Refer Slide Time: 07:09)



The place theory of hearing the one which I was referring to proposes that different areas of the cochlea respond to different kinds of frequency this might be informative as higher tones will excite areas closest to the opening of the cochlea which is near the oval window there is lower tones will excite the areas near the narrow tip of the cochlea which is at the opposite end which can therefore be determined in part by the area of the cochlea which is firing most frequently. Another important fact about the placement of the ear around our head is also important to look at.

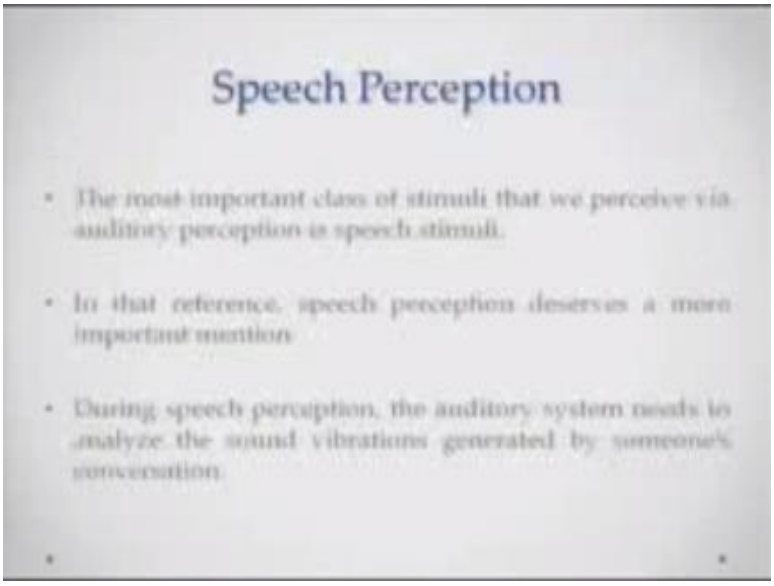
(Refer Slide Time: 07:42)



So you know that the ears are placed on either side of the head and they kind of give us the benefit of stereophonic hearing which is that you can locate the sound in three dimensional space, if a sound occurs on your left side the left ear receives the stimulation earlier and you will respond by the left ear if the sound happens on the right side you will be able to detect this from the right ear and you respond accordingly.

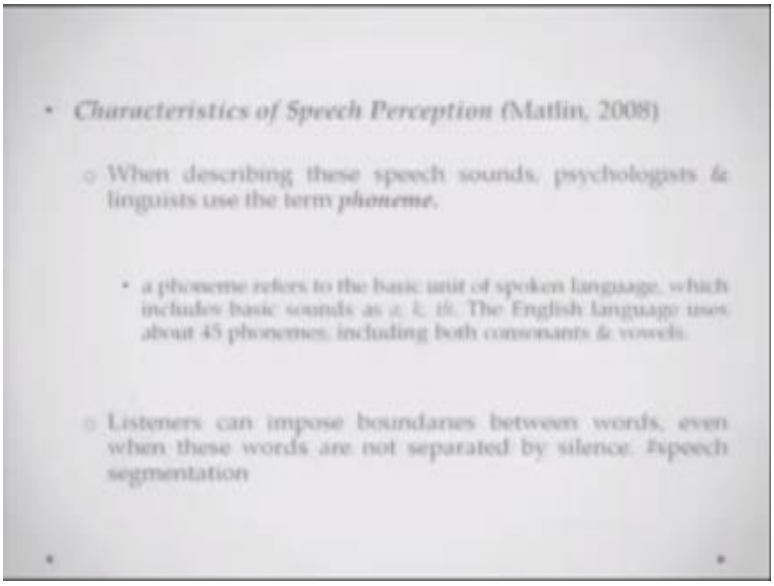
Now although the distance between the two ears is around six inches and sound travels you know around at 750 miles per hour these time and intensity differences of various kinds of songs are easily detected. When a sound is equidistant from both here say for example if it is in the front or in the back and the distance is equal from both ears then we have slightly more difficulty in pinpointing the exact location of the sound and then we might maneuver our ear position in order to get the best hint of whatever the sound is. You might have seen there are certain animals that do that.

(Refer Slide Time: 08:40)



Now the most important part of auditory perception as I said earlier also is listening to speech, so we will spend some time here talking about speech perception as a special case of auditory perception. Now what happens during speech perception is that the auditory system needs to analyze the sound vibrations which are generated by somebody's conversation okay.

(Refer Slide Time: 09:06)



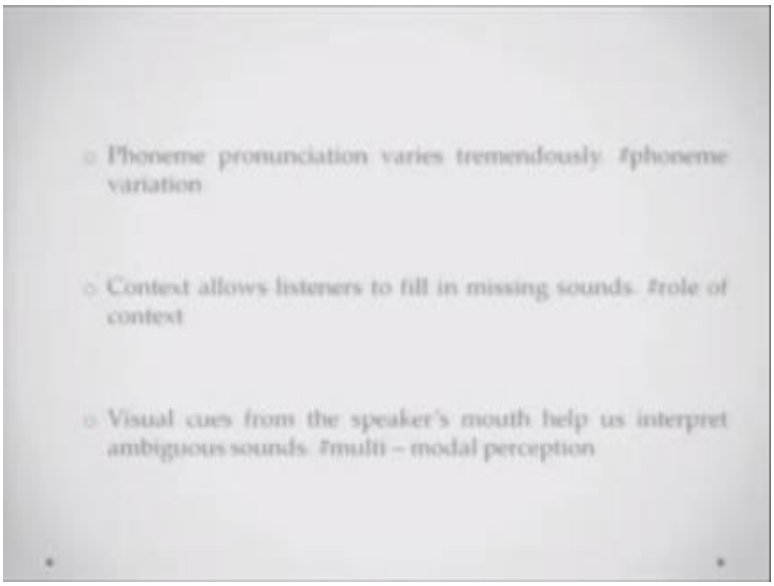
So there are a few characteristics of speech perception that have been proposed by Macklin say for example when you start talking about speech sounds the fundamental unit of speech sounds proposed by linguists and psycholinguists is the term called phoneme and the phoneme refers to any basic sound so say for example if you are listening to a language you can reduce the entire language into a set of very basic sounds called phonemes.

Say for example if I am talking about Hindi any sounds like Burma those kind of things if I am talking about English then the same kind of sounds like ABC those things can be referred to as phonemes. Now another important characteristic is when you listening to a speech when you are listening somebody conversing listeners will need to impose word boundaries. Because otherwise everything will be jumbled up and you are not really able to make sense of what message is being conveyed.

It becomes much easier when you are reading something on a text and then there are spaces which tell you that the word boundary ends here and the other word begins somewhere there but with the spoken signal but with the speech you have to do it slightly you know by yourself and that is something which is an important feature that is achieved by humans. If you kind of want

to appreciate how difficult or easy it is you might want to listen to your movie or a song in a language that you do not know at all and then you see how difficult it is to really generate word boundaries.

(Refer Slide Time: 10:34)

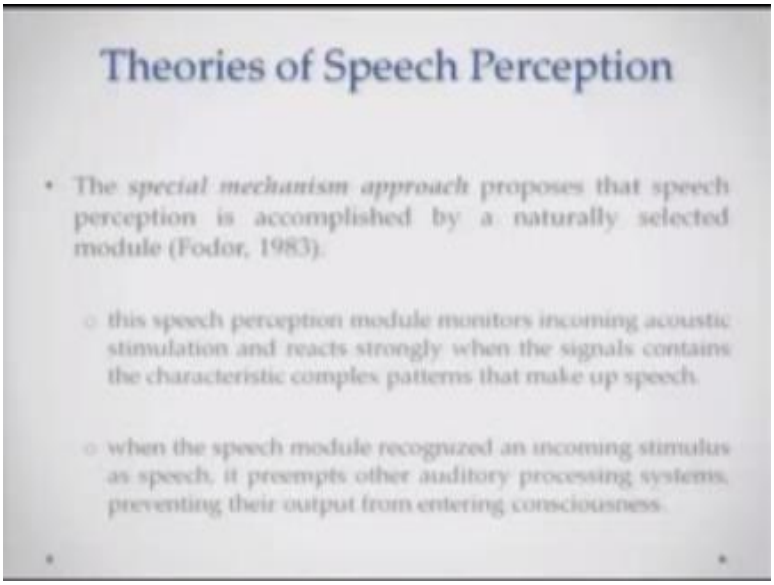


Another thing is that the pronunciation of these phonemes or these basic sounds in speech is a varies extremity from speaker to speaker and also within speaker from situation to situation. Now an important task of the auditory perception system is actually to you know understand this variation and still not produce too many different results for the same set of sounds. We will talk about this in more detail further, context also allows listeners to fill in missing sound.

Say for example if you are talking to somebody on a disturbed phone line a lot of times you will see that part of the signal will be missing but you will still be able to understand what is being said by the other speaker this helps you this means you are being able to do this because you have the context information of what the topic is who is talking and all those kinds of information which basically form the context and you use that context to fill up the missing bits of information and still understand the conversation very well.

The final thing that might be very important from a listeners point of view when talking about speech perception is also the use is the use of visual cues you are not only listening to somebody you know from the audition perspective you are also looking at somebody most of the time when they are talking you are also you are you know and you have the input from the lip movement from the gestures and those kinds of things and these kind of information these multi-sensory multimodal information also helps you solve this puzzle of speech perception.
We talk about speech perception in more detail now.

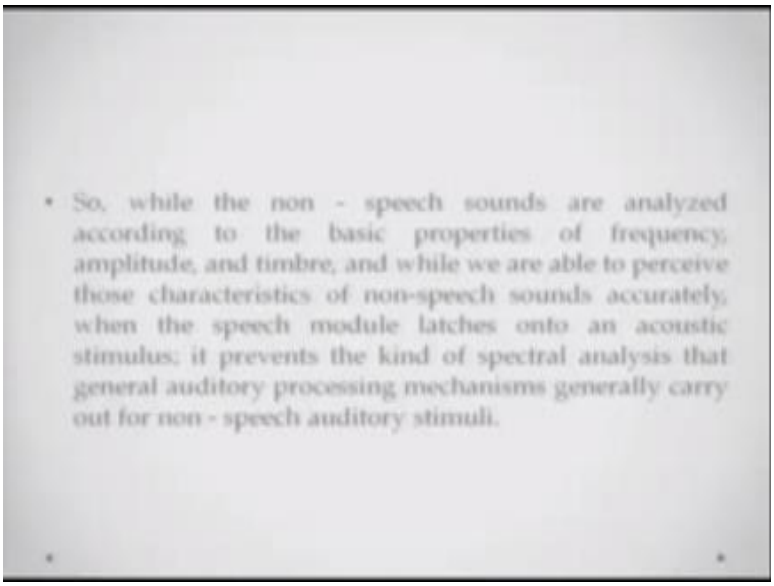
(Refer Slide Time: 12:04)



Because we are going to talk about what is called theories of speech perception. Now there are two classes of theories in a speech perception of three classes of theories in speech perception the more prominent of them is called the special mechanism approach or also the motor theory of a switch perception as it is referred to and it says that speech perception is accomplished by what is called a naturally selected module they say that speech perception that this special speech perception module monitors whatever incoming acrostic stimulation that you are receiving and then it reacts strongly whenever the incoming signal contains characteristic patterns related to speech.

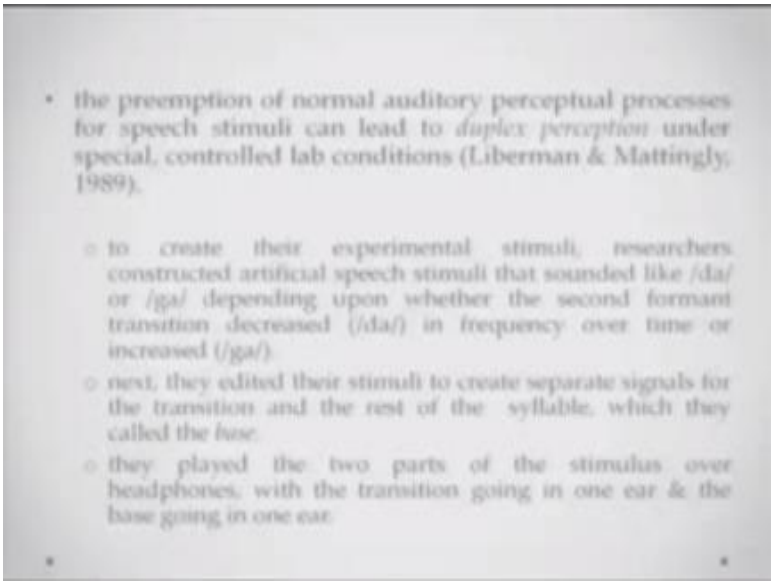
And when the speech module recognizes that an incoming stimulus contains speech contains a speech signal then what it does is it prints or disallows the other auditory processing systems to act on this particular speech signal, because the processing that is required to understand speech is slightly different from the processing that is required to understand other kind of non speech sounds.

(Refer Slide Time: 13:06)



Now and this is what it is so the null space sounds basically are analyzed according to basic properties of frequency, amplitude, timbre and why we are able to perceive the characteristics of non speech sounds accurately when the speech module latches onto an acrostic stimulus it will prevent this kind of analysis and it will basically try and understand speech in terms of whatever prior knowledge in terms of linguistic variables that you have and that is more useful in understanding speech information.

(Refer Slide Time: 13:33)



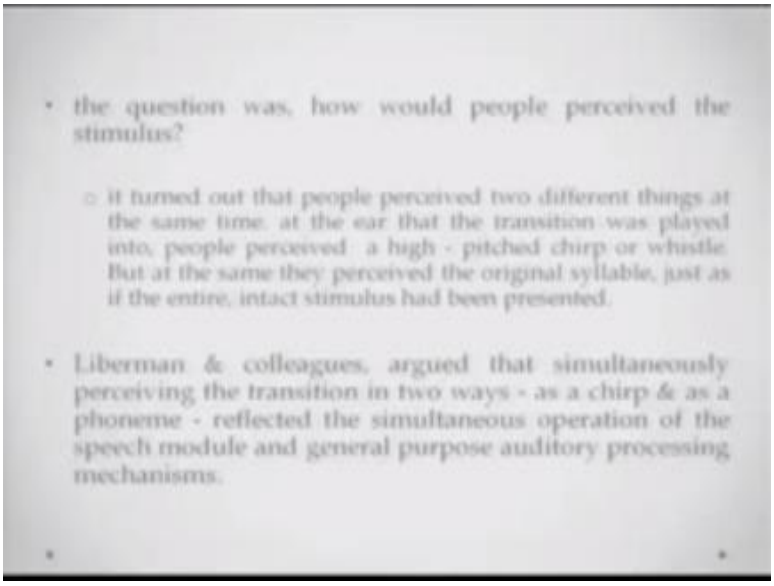
This preemption or this stopping of a general auditory processing on speech signal sometimes can lead to what is called a duplex perception and the particular experimental scenarios. I want to describe a very simple very interesting experiment to you guys which was done by Lieberman and Mattingly in 1989 so what they did was they created artificial speech signals and the three signals represented phonemes or sounds like da and gah and basically they differ from each other in one sense that there is one part of these SP signals which is flat which is not changing.

And the other part of the speech signal and referred to as the second formant might either be increasing or decreasing. So these P songs have two parts and I can just show you right here, so these P speech sounds could be something like.

(Refer Slide Time: 14:25)

A dark would be represented by a flat line and an increasing curve and girl could be represented by a flat line and a decreasing curve. The flat line was referred to as base and the increasing or decreasing curves could generally be called transitions. Now what these people did was they bifurcated the base and the transition and in by use of head phones they presented the base and transitions separately to two different ears, so what is happening is that in your ear you are still getting the whole signal but you are getting this in the left here and this one in the right here. Now this creates an interesting scenario and the question that one can ask is.

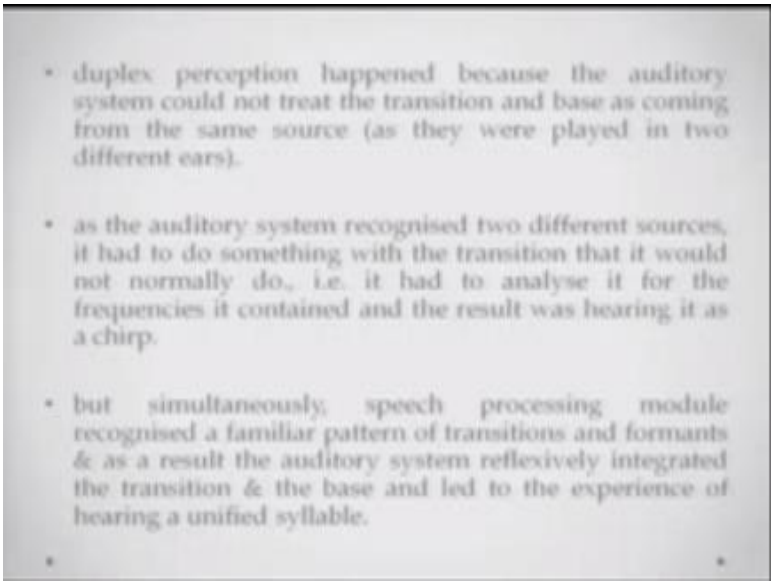
(Refer Slide Time: 15:09)



How would people perceive this kind of image they are still getting the whole stimulus but they are getting it bifurcated into two ears, now they found that people were perceiving two different things at the same time at the year that the transition was played into let us say the right ear here they perceived a high-pitched whistle or a chirp sound but at the same time they also perceived the original syllable so they are perceiving this chirp sound as well and this whole syllable as well.

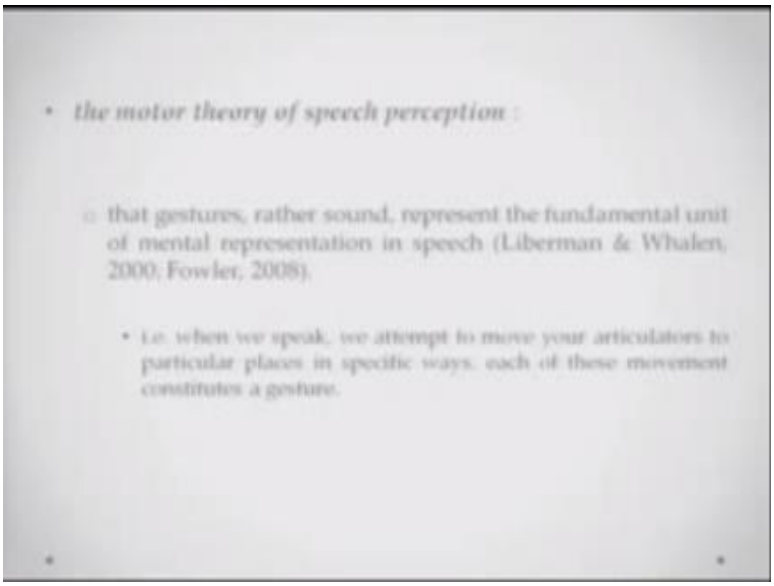
Now this was interesting and people wanted to understand why this is happening, so Lieberman and colleagues they argued that the simultaneously perceiving of transition in two ways as a chirp and as a phoneme it reflects the simultaneous operation of both the speech module and the general purpose auditory processing mechanism which will analyze speech as a non you know it will analyze the incoming about sixty minutes as a non speech their stimuli.

(Refer Slide Time: 16:06)



So what is happening here is this duplex perception is happening because the auditory system could not preach the transition as base as coming from the same source, so what happens is it starts to analyze the transition separately using the non you know non speech analysis and it later at the same time recognizes that no this is part of a larger signal which is the base as well, so what it happens is it generates two kinds of results one is that this is listened to as a whistle and the other is that it is listened to as the same phoneme now this was interesting.

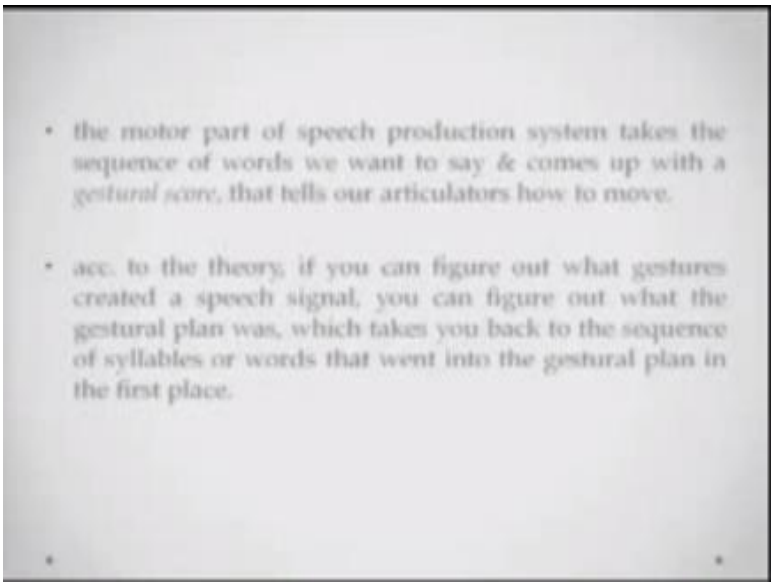
(Refer Slide Time: 16:40)



Now another important effect about this special mechanism theory of perception is also as I said that it is also the motor theory of perception. Now it says basically that gestures we were talking about that phoneme is the basic unit of speech perception but they proposed differently they say that the gestures that are involved in producing the speech sound should be treated as fundamental units of representation.

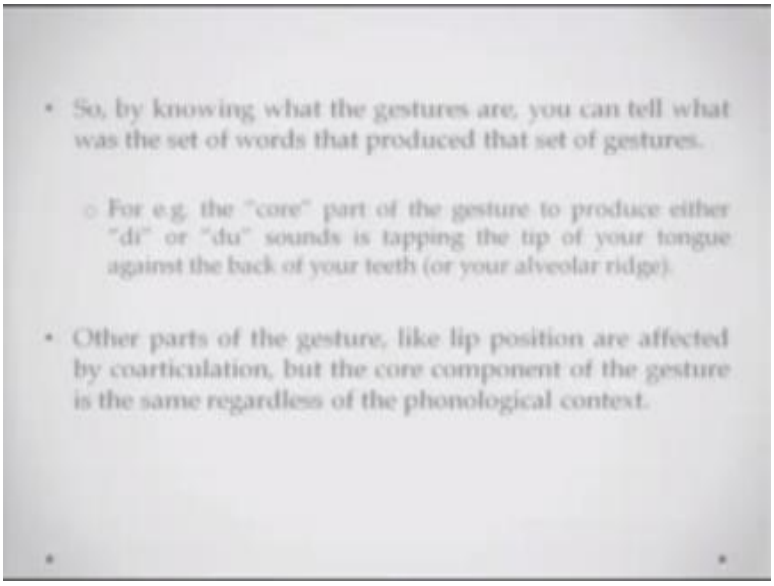
So they are saying that if you want to understand a speech sound you want to understand which vocal gestures produce that sound, now what are these vocal gestures so you can see that when you are trying to speak you are you know moving your lips, you are moving your tongue, you are moving your mouth and vocal cords in special ways all these organs are called auriculars and specific movement patterns in them lead to generation of sounds this is what is called a gestural score.

(Refer Slide Time: 17:34)



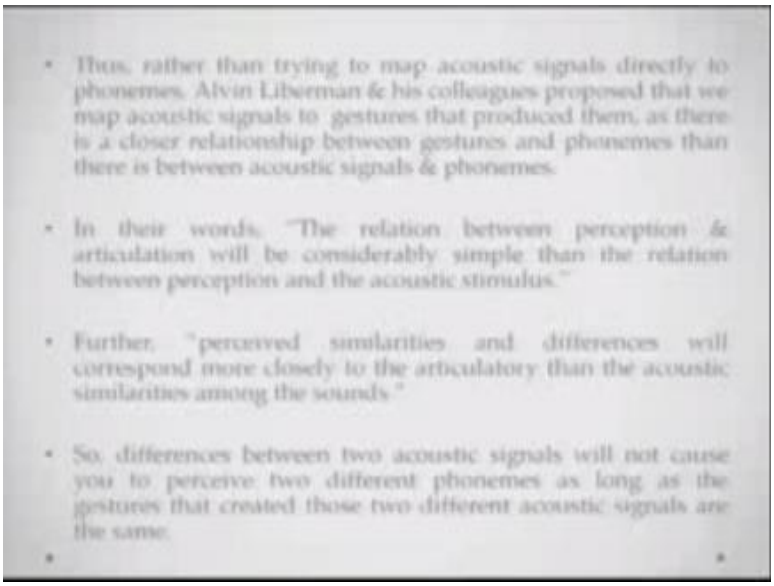
So this theory is saying is that you know the motor system when you know when it has to produce any kind of sound it basically creates a particular plan of movement of these things that this particular part a will vibrate now and part B will move so this kind of plan is called a gesture score and the motivated speech perception says if you can figure out this gestural score you can understand what speech is you know whatever speech is being said, you do not really need to go to the sounds to understand this you just make out what movements are made and you know that which movement produces what sound and that will help you understand the sound much better than say for example if you go by the phoneme level of analysis.

(Refer Slide Time: 18:15)



Now this is what they are saying, say for example I can give you an example the core part of the gesture to produce either you know de or do if there are two sounds is the same the core sound is the der sound which is produced when you put your tongue at the back of your teeth the other part however your lips move when you are saying de or do is different, so they think if you can get this core part alright you will be able to understand which sound has been produced I will not be confused by coarticulation effects of the you know the lip movements and those kind of things.

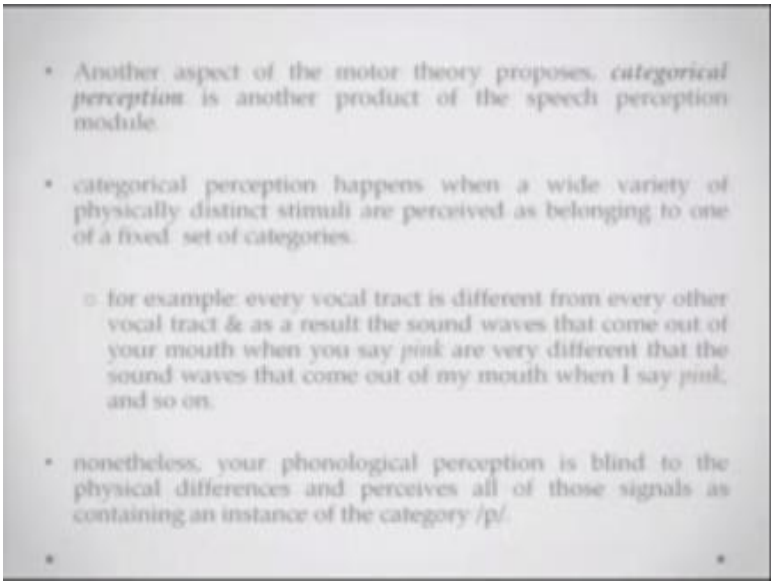
(Refer Slide Time: 18:52)



Rather than trying to map acoustic signals directly to phonemes Alvin Lieberman and his colleagues proposed that we must map acoustic signals to directly to gestures that produce them as there is a closer relationship between gestures and phonemes. They say the relation between position and articulation will be considerably simpler than the relationship between perception and sound, they further say perceived similarities and differences will correspond more closely to the articulatory rather than the acoustic similarities.

Say for example the difference is between two acoustic signals will not cause you to perceive two different phonemes as long as the core you know articulatory gesture is the same.

(Refer Slide Time: 19:31)



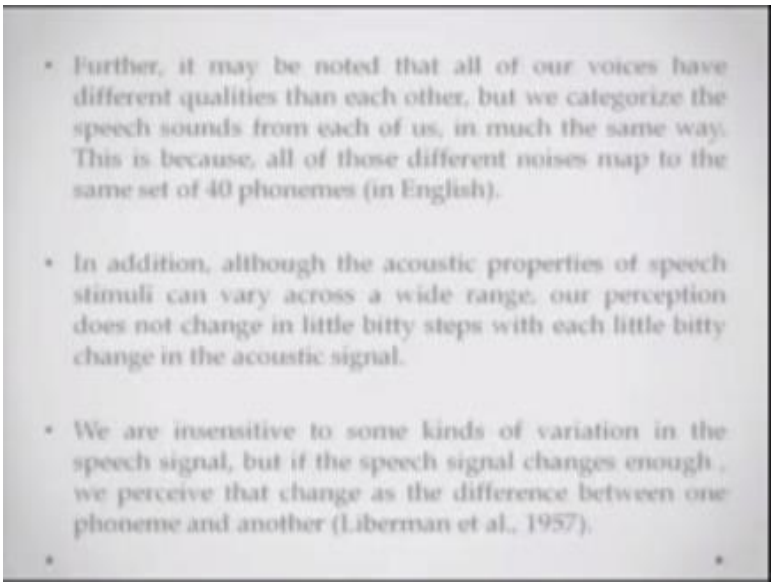
Another aspect of this theory of perception is the concept of categorical perception, now what does categorical perception categorical perception is that when you are hearing a wide variety of physically distinct acoustic stimuli you will not treat all of them differently what you will do is you will say these set of stimuli belong to one category and these other set of stimuli belong to other category.

For example, you might notice that every vocal tract mine and yours are different from each other and they produce different kinds of sound waves so the way I have to say a particular word let us say pink or blue or black and the way you will say pink or blue or black or the way somebody else will say pink or blue or black might be very different, but you must notice you must have noticed that we treat all of these sounds as very similar and we kind of did I the same meaning sort of these sounds even though the physical or the acoustic signatures might be drastically different from each other.

Even say for example within the individual if I am say for example sleepy or if I am very tired or am panting and exhausted the kind of speech signal physically that I will produce will be very different. But what I do is out of what are phonological system does is it is kind of blind to these

kind of physical differences and you will perceive all of these different kinds of signals as an instance of one category let us say if I am talking about a sound pink it considers everything else as you know an instance of the sound P. Now this is interesting and important in the sense that it helps us understand a wide variety of sounds as one even though the physical signatures might be slightly different.

(Refer Slide Time: 21:12)

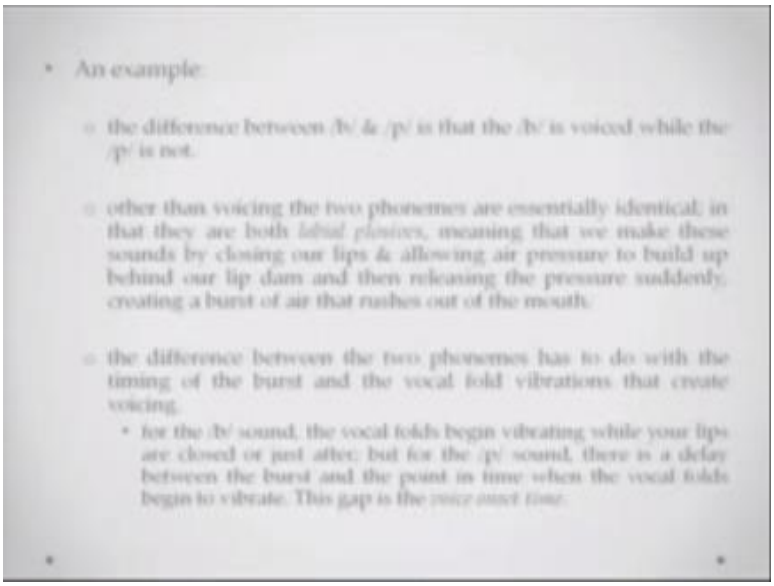


Further it may be noticed as I already said that all of our voices are of different qualities than each other but we again you know categorize here the speech sound coming from each of us as in one category, one of the reasons why we are able to do this so easily is that anything that you are saying say for example in a language called English can be you know broken down into around 40 to 45 phonemes that this language has.

So anything that is said in any way by anyone can again be mapped down to these 40 or 45 phonemes. In addition although the acoustic properties of speech similarly can vary across a wide range our perception does not change in these little, little steps okay, so we are said insensitive to these kinds of variation in the speech signal but if the speech signal now starts

changing too much then we will kind of you know create different phonemes for them and then you start hearing them slightly differently. Let me take an example of something like this.

(Refer Slide Time: 22:08)

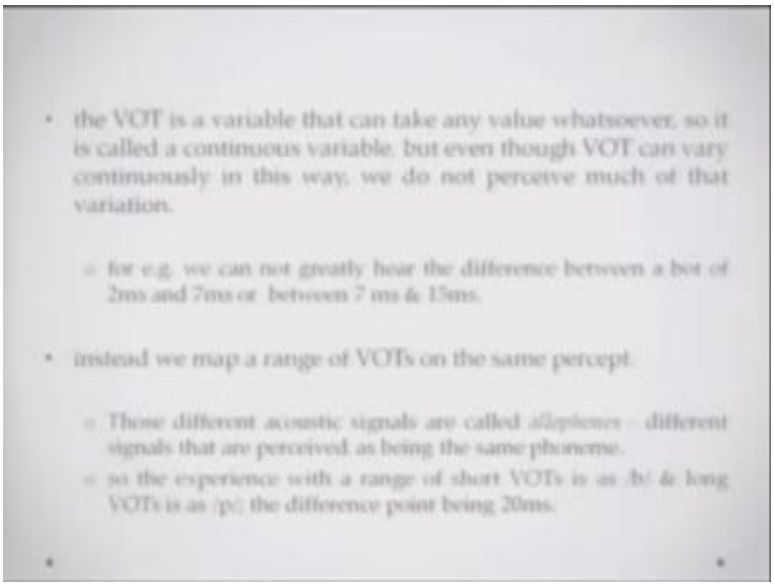


Now the only difference between the sounds B and the sound P in English is that B is a sound whereas P is not a voice sound, now the details are not really important here you just have to get the point voice or not voice basically means that even though both of these sounds are technically called labial process, labial process basically means that you use your both the lips to you know converge and then you block the flow affair and then suddenly there is a burst of air.

After this verse of it there is either a gap and then your vocal cord vibrates or not a gap and then your vocal cord variables when the vocal cord vibrates after a gap this time delay between the vocal cord vibrating and the air burst happening is called voice onset time. Now as far as B and P are concerned for the B sound the vocal cords start vibrating while your lips are still closed for immediately after.

But for the P sound the vocal cord starts vibrating after a slight delay after a particular voice onset time, now this is the only difference this voice onset time is the only difference between the sounds B and P.

(Refer Slide Time: 23:16)



This voice onset time is a variable that can take any value whatsoever so it can be called a continuous variable. Now imagine there is a set of sounds on this variable where in their different voice onset times are possible so how we will perceive these different you know sounds BP and others. If the variables if the sounds B and P differ from each other by at least 20 milliseconds we will listen to them as B and P separately but if the difference is less than 20 milliseconds that is somewhere around 7 milliseconds, 15 milliseconds or something like that then we perceive them as the same sound.

When we perceive them as the same sound is different or having different acoustic signals they are called as allophones very similar several happiness will experience the 16 is the sounds with a range of short beauties as B anything smaller you know then let us say 15 milliseconds or anything longer we can start perceiving this as P. Now here you could see that you know the sound system or the sound perception system or the auditory perception system is sensitive to

differences but it kind of washes over all the places that are less than meaningful that are not really you know that are just variations of different speakers speaking that and those kind of things.

It is an important thing because it helps us understand this speech which is produced by a variety of speakers under a variety of circumstances. I would like to close here now and we will start talking about speech perception or other aspects of the motor carrier speech perception in the next lecture thank you.

Acknowledgement

Ministry of Human Resources & Development

Prof. Satyaki Roy
Co – ordinator, NPTEL IIT Kanpur

NPTEL Team
Sanjay Pal
Ashish Singh
Badal Pradhan
Tapobrata Das
Ram Chandra
Dilip Tripathi
Manoj Shrivastava
Padam Shukla
Sanjay Mishra
Shubham Rawat
Shikha Gupta
K.K Mishra
Aradhana Singh
Sweta

Ashutosh Gairola

Dilip Katiyar

Sharwan

Hari Ram

Bhadra Rao

Puneet Kumar Bajpai

Lalty Dutta

Ajay Kanaujia

Shivendra Kumar Tiwari

an IIT Kanpur Production