The Psychology of Language
Dr. Naveen Kashyap
Department of Humanities of Social Sciences
Indian Institute of Technology-Guwahati

Module No. #01
Lecture No. #05
Speech Perception - I

Namaste, friends. Welcome back, to this fifth lecture, on the course on, Psychology of Language. Now, the highlight of today's lecture, which is the fifth lecture, and the one to follow this one, which is the sixth lecture, would be on, Speech Production. Which is, how do we produce, speech. The talking that we do, how do we, do that. What we will unfold in this lecture, is the very basic. So, please understand, that this is not a Linguistic course.

And so, we will not go into, too much detail onto the, linguistic aspect of speech production. So, we will just look at the, psychological aspects of language production, or production of speech. And so, we will be quite restrained in, our explanation. But before, we start unfolding the lecture, as I have been doing, in the past lectures. Let us look at, a little bit from, the past lectures. That is, 1, 2, 3, 4 lecture number.

And, that also helps in, getting the context of these lectures. So, each lectures, getting into the context. So, the Psychology of Language, as I said, is a course, which looks at, the psychological aspects of language. So obviously, in the first two lectures, we try to define, what is language. We started off, by explaining, the basic form of language, which is animal communication system. And, we looked at, what are the characteristics, of this animal communication system.

Since, animal's communication, is the very primitive language, or the very basic language. Now, as we said there, it is a very basic system, which can only explain, or express a few ideas. So, we looked at, that. We looked at, why do animals communicate. And, we looked at, some characteristics of this language system. Then, we moved on to, the idea of human language. We saw, how human language, is different from animal language.

So basically, the characteristic features of human language, namely, that it can produce, a multitude of ideas. And, it can have a structure, which is recursive in nature, which provides it with, the flexibility of expressing, within few limited basic structures, the whole language itself.

So, we started looking at, human language. And, we started distinguishing between, human language, and the animal language.

We then, looked at the characteristics of human language. We looked at things like, the various modes of human language, and the idea of duality of patterning. We looked at, how the human language is built, right from, the basic speech sound, to the idea of Morphemes, to the idea of word, sentences, syntax, which is how the words should be arranged, grammar, which is the rules, that a language should follow, then the idea of discourse.

So, we looked at, all those structures of, how the human language is arranged. We also looked at, certain rules, or the face structure rules, and how within the particular sentences, certain parts of the language, or certain words can be rearranged. Then, we focused on, the fact of how, language evolved. So, we saw, how the Neanderthal man, the Homo sapiens, they came across, and how they had developed this idea of language.

And, we looked at, the Continuity and the Discontinuity Theory of Language, which basically says, how language evolved. And, after that, we moved on to, and so, we looked at several evidences, of both, the Continuity and the Discontinuity Theory. And, then we looked at, the fact of how the Fossils of Language, the Living Fossils of Language, basically how languages were described, to have evolved from, the Proto language.

The idea that, the present human language, which is complex in nature, how it evolved, from a proto language, or basic language. So, there we will investigate the idea of pidgins, which are Primary form of language, which do not have, too many words in to it, too many complexities into it, but can be expressed for, expressing ideas. And, we looked at how, this language teaching, research has been done across, um various scientists.

And, how some basic language, human language, were taught to Chimpanzees. And how, Chimpanzees and Humans were difficult. So, the idea of Chomsky, which says that, humans have a language acquisition device, against the idea that, animals do not have that, kind of a thing. So, we concluded the chapter there, with this debate. Lecture number 3 and 4, was explicitly focusing on, Research Methodology into language.

So, we explained, what is the way of doing, research in language. I explained in detail, what is the Scientific method, what is theory, what is Hypothesis. How, Theory-Hypothesis observation, and patterns, they interact together, to form the Research Cycle. And, the idea of how, Induction and Reduction works, into the research cycle. And, how we build models, to explain, certain hypothesis, or certain predictions of the theory.

We looked at, the idea of experimental design, how there are, the between subject, and within subject designs. And, what are the IV'S and DV'S, and those kind of things, that we looked at. We looked at, behavioral techniques of doing language research, in which, we primarily looked at, Latency, which is the Reaction time, and Accuracy, which is the Correctness, as dependent variable measurement, or as choices of measurement variables, in most language researches.

We looked at, several examples, in terms of the, Lexical Decision task, and several other examples, there to explain in detail, what we have been covering, or what we have been talking about. And lastly, we looked at, the language in the brain, how the brain and the language are integrated together. We focused on, certain regions of the brain, certain areas, which is dedicated to language.

And, we looked at, certain equipment's, or techniques, like EEG, MRI, FMRI, and how these techniques are used, for studying language, or for completing a research on language. Now, today's lecture would be an interesting lecture, where we look at, how do humans produce speech. The way, I am speaking to you, how do I produce this speech. So, to start today's story, we will go back to the Haskins Laboratory, where the first Artificial Speaking Machines were designed, and this they called the, Pattern Playback machine.

Now, this pattern interest in language, of how language is produced, has been around for quite some time. And, just after the world war I, Haskins Laboratory, produced an artificial language machine, which was called the, Pattern Playback machine. Now of course, this machine had, no vocal cord, no mouth, in other language equipment's which human have, to produce speech. But, what this machine was able to do, was produce, some basic frequency sounds.

And, these basic frequency sounds, when they were put together in a sequence, it sounded like, the human language, or the human speech. That was the, beauty of it. But, it could produce, only very basic speech. So, even these sounds, which are produced, by this machine, when they

were arranged in complex pattern, it could not be matched exactly to, the way humans produced the phones, the basic speech sound. But, it was much closer.

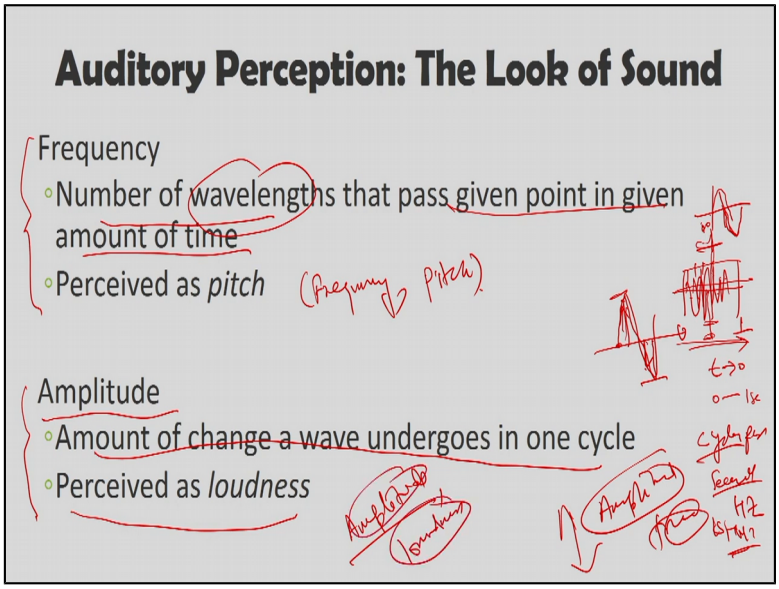
So, humans, when the playback machine were played back, to that official machine was played back to human, they could hear, or they could promise that, they hear, some basic speech sounds. That was the first attempt. From then, which was just after the World War II, to presently, we have much dedicated systems, which produce artificial language. We have Siri, we have the Google assistant, which can not only understand, but also listen to you, and then respond to you.

So, the first machines, which were producing just specific frequencies, which uttered, when uttered through these complex arrangement of sounds, sounded like human language, to the present world google assistant, which you can talk to. And, not only, it produces sound, which you can hear, and understand, it can also understand, what you are saying, and comprehend meaning out of your speech. So vast area. So, let us look at, some of the psychology behind, the speech perception.

We will start, by looking at, how language is, what is the very basic item, or element, in producing of speech. And so, the very basic element, in auditory perception is the, speech sound, the look of speech. So, how does sound, look like. Now, as you know, what is sound, the ear actually, when you produce speech, there is a pressure difference, which is created. And, this pressure difference, it moves, in terms of waves, across the space, between two people. One, the producer of the speech. And, the other, the receiver of the speech.

And, when it hits the ear, which is the perceiving mechanism, for the other person, this is decoded back, this this pressure, or this wave, which is travelling, and meaning is extracted out of it, or people understand that. The producing system, is the vocal cord, or the vocal fold, and the mouth. And, the receiving system, is the ear. So, we look at, these two systems. But, even before we do that, what is used, to transfer this vocal fold vibrations, is the wave, the sound waves.

(Refer Slide Time: 12:26)



So, how do you explain, this wave. If you sever sat, near a pool of water, taking a rock, and thrown it, into the water, you will see patterns of, high pressure area, and low pressure area, or certain patterns moving away from, the point of impact of the stone. And, this is what a, wave is all about. And, this is how, the sound wave also, looks like. So, any wave, a wave is basically, the highs and lows, of pressures.

So, when wee when we speak, what we are actually doing is that, there are air molecules between, our mouth, and the other person's ear. And so, these air molecules are vibrating, in a particular frequency. And so, when we speak, what happens is, the vibration changes in a certain way. And, so it travels in the, form of a wave. Now, any wave, so basically, sound is a wave.

So, any wave has, two fundamental things, or there are two fundamental properties, of any wave. The first is called the, frequency. The frequency of the wave. And so, what is the frequency. It is the, number of wavelengths, that pass, by a given point, in a given amount of time. So, if this is my measurement axis, or this is where, I am doing the detection. And, let us say, this is my time axis. So, let us say, from 0 to 1 second, the number of times, this is 0 second, and this is 1 second.

The number of times, this vibration is there, is called the frequency. So, number of wavelengths, this is one wavelength, this is two-wavelength, this is three-wavelength, four-wavelength, and so on, and so forth. So, the number, which is there, that gives the amount. And, this is measured

in, cycles per second, or hertz. How many cycles. So, this is one cycle, this is two cycle. This is sine wave. Right.

A Sine wave starts at the 0, 0 at start, goes to a maxima, a local maxima, comes back to the baseline, goes back to a negative minima, and then comes back. So, this is what, a Sine wave looks like. I am pretty sure, you must know, what a Sine wave is. And so, most sound waves are, considered to be, sine waves. So basically, a wave is explained then, in terms of, how many waves, or how many wavelengths, pass a given point.

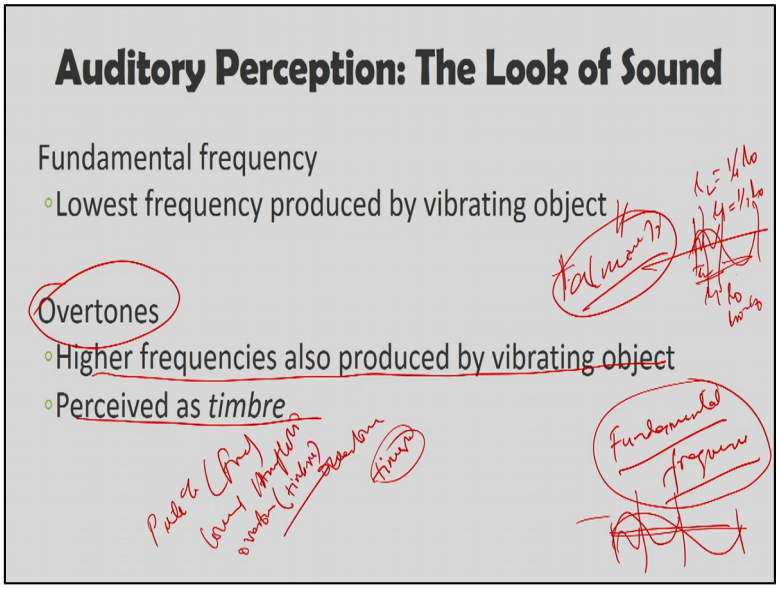
This is my point x, detection. And, in 1 second, how many of them are PA passing. Let us say, there are 15. So, it is 15 cycle per second, or 15 hertz. The psychological perception of sound, wave frequency, is called a, pitch. So, what we call, frequency, in the physical domain, is called a pitch, on the psychological domain. So, when we talk about pitch, it is, how many, wavelengths are passing, a particular point of detection, in a particular period of time.

So, this is the first thing, you should know, about a wave. And, I am pretty sure, you know these things. So, we are just, repeating these things, or we are just, adding on to it. The next thing, that you should know, what a wave, is the amplitude. This is the height, of a wave, from the baseline, the deviation from the baseline. So, amplitude is the amount of change, that a wave undergoes, during one cycle. So, wave starts here, goes to a maxima, goes to a minima, and comes back to the baseline.

So, the maximum change, that it is facing, this point to this point, is called the amplitude. This is also, in terms of sound waves, the difference between, the highest pressure, and the lowest pressure, or the sound wave, is called the amplitude of that wave. So, the amount of change, that a wave undergoes, in one cycle, is the amplitude. These two things, are very necessary, the frequency, and amplitude. And, most measurements of wave, are done, on this basis only.

So, on the physical domain, this is called the amplitude, in the psychological domain, this is called loudness. So, what you call amplitude, in physical domain, is loudness, in the psychological domain. Now, these two properties, the amplitude, and the frequency, it describes the sine wave, so called. A Sine wave is so called, because, it is a wave, that can be described, by a Trigonometric function, which is called the, Sine function.

(Refer Slide Time: 17:21)



Sine waves, why do we call as Sine wave because, it is basically, it can be explained, through a trigonometric function of sine. Now, some other things, that we need to know. We need to know, what is the fundamental frequency, of any wave. If you take a rubber band, and put it across your two fingers, and pluck it, bring it near the ears, it will produce a sound. You will see the, rubber band vibrating. The vibrations, which happen, across the entire length, of the rubber band, is called the fundamental frequency.

So, vibrations along, the entire length of the string, or rubber band, generates the fundamental frequency. And, which is also the lowest frequency, produced by the vibrating object. Now, since it is for the entire length, obviously the amplitude has to be less. So, this is the low frequency. Since, across the thing, so the frequency will generally be 1. And so, the low frequency produced by vibrating object, is called the fundamental frequency.

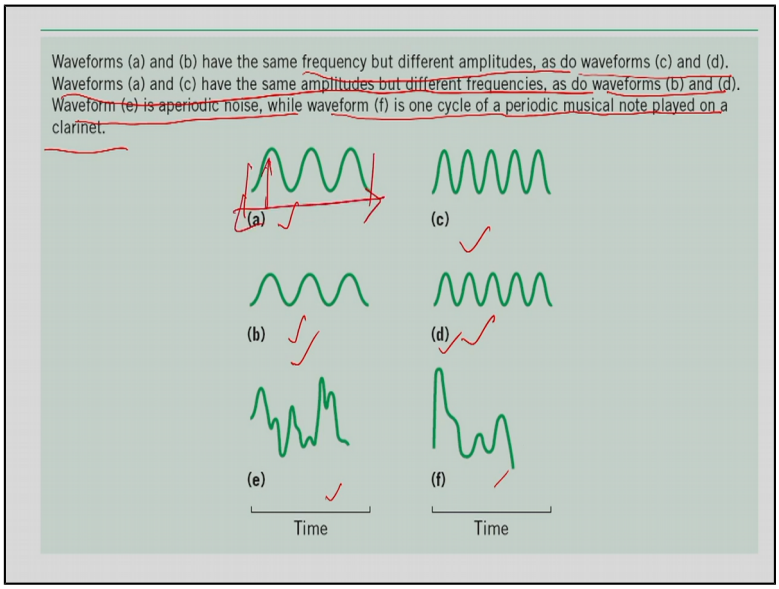
But when, you look at a vibrating object, this object does not only produce, waves, across the entire length. Right. Vibrations also happen, at half the length, and then, one-fourth the length. So, when you have a vibrating object, or a tuning fork, this is my fundamental frequency, this is another frequency, and this is another frequency. So, this is the full length L0, this is L1, and this is L2. So, L2 is, 1/4 of L0, L1 is, 1/2 of L0, and L0 is of course L0.

Now, this type of frequencies, vibrations at half length, third length, and so on, generate, what are called the, Overtones. So, if you have, a guitar, and a violin, they sound different. Because, although the fundamental frequency, if they are playing the same string, say, A-minor, or something like that, although they are playing the same sound, they sound differently. Because,

they have different overtones, the lengths are different. And so, they produce different overtones. So, higher frequencies are also produced, by the vibrating object.

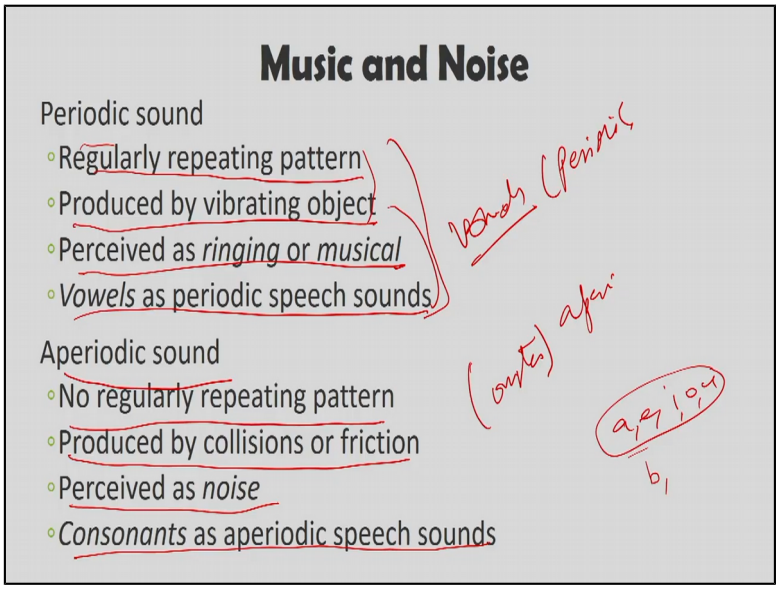
And, these are perceived as, timbre. Frequencies, higher than the fundamental frequency, is called the overtones. Because, they are at 1/2 length, and 1/4 of length. And these, in the psychological domain, is known as timbre. So, overtones, are called timbre. Now, if you remember, these are the three things, that we have in sound. We have the pitch, which is the frequency. We have the loudness, which is the amplitude. And then, we also have the overtones, which is partial frequencies, half frequencies, the frequencies higher than the fundamental frequency, and these are called the, timbre.

(Refer Slide Time: 20:35)



Waveforms A and B, have the same frequency, but different amplitudes, as do waveforms C and D. Look at, A, B, C, D. Waveform A and C, have the same amplitude, but different frequencies, as do wave length B and D, B and D. Waveform E, is aperiodic noise. While, waveform F, is one cycle of Aperiodic musical keynote, played on a clarinet. So, this is Periodic. This is, aperiodic. And so, we will discuss that, aperiodic, Periodic. So, this is my, from this tone, to this tone, this is my frequency, and this is my amplitude.

(Refer Slide Time: 21:11)



So, vibrating object, produces a sound, with a regularly repeating pattern. So, when something is vibrating, an object like a tuning fork is vibrating, it produces sound, with repeated, regular repetition. This is known as, the Periodic sound, to objects. But, that is not the only sound, which is there. So, most vibrating objects, they produce Periodic sound. But, some of them also produces, sometimes, they also produce, Aperiodic sound.

So, two objects, rub against each other, produces sound, with no regular repetition pattern. And, these are called the, aperiodic noise. For example, you rub your hand, like this. And, bring it near, your ears. The sound, that you hear, is not Periodic. And, this is called, aperiodic noise, aperiodic noise. So, Periodic sound, these are regularly repeating pattern, produced by vibrating objects, perceived as ringing or musical in nature, and vowels as Periodic speech sounds.

So, this these two, we have not explained. But, these two, we have explained. And so, most vowels, are Periodic. And, most consonants, are aperiodic. So, A E I O U, you can sing this. This are produced by the vibrating quad, without any restriction, to the air which is, sent out by the mouth. So, A E I, no restriction, and you can sing it. But, with other things, the consonants. So, anything, except the A E I O U in English language, is a consonant.

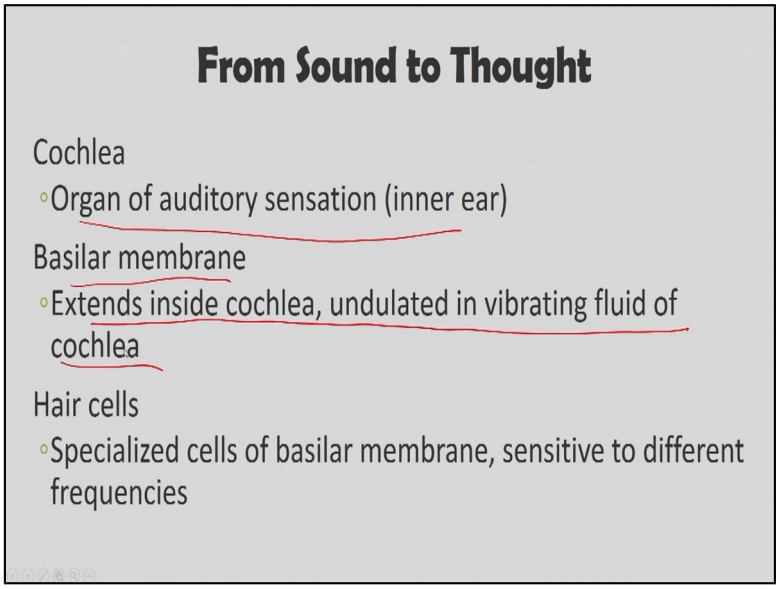
So, example, B. B, as there is a restriction, in the sound, which is coming out of the mouth. And, this is called a, consonant. So, Aperiodic sound, these are generally produced by consonants. No, regular repetition pattern. Produced by, collisions or friction. Perceived as,

noise. And, consonants are, Aperiodic speech sounds. Speech sounds, can be broadly categorized into, vowels, and consonants.

Vowels are Periodic, which is, they are ringing, their musical character, to them. For example, AAAAAAAA, I can, keep on singing, like that. But then, consonants are Aperiodic, and they are noisy. I cannot say, TEEEE. On saying, TEEEE, and not lose my breath. But, I can. After a period of time, I lose my breath. There are however, some consonants, which can be produced without friction.

For example, the SH, ssh. You can do that. That, such noise, produced with, without friction, this is also consonant. And, you can do it, for a longer breath, across your breath. But then, there are other consonants like P, you cannot sing it, for a longer period of time. So, these are the basics of the sound, or the basic of the wave from which, the speech is produced.

(Refer Slide Time: 24:09)



How the speech is produced? Next, we look into, how this speech, which is produced by the vocal tract, how this is heard by, the ear. So, if I ask you this question, how do you hear sound. The answer to this is, through my ear. And, maybe, you will point to this this this cartilage, just coming out from the head. Now obviously, this has, nothing to do with. This is, just a directional mechanism. And, this part, even if we cut out like, when god did, you will still be able to hear sounds.

This has, nothing to do with, the ear. The ear, perception of sound in the ear, actually starts, in the inner ear, at the inside of an ear. So, basically, the organ of auditory sensation, is the Cochlea, that is the, a snail-like organ, inside the inner ear, which has hair like structure, which can see, or which can perceive, changes in sound, and then convert these changes, into electrical impulses, which are then perceived by the, Primary Auditory Cortex. The Basilar membrane, extends inside the length of the Cochlea, and its hair cells, are sensitive to particular frequencies.

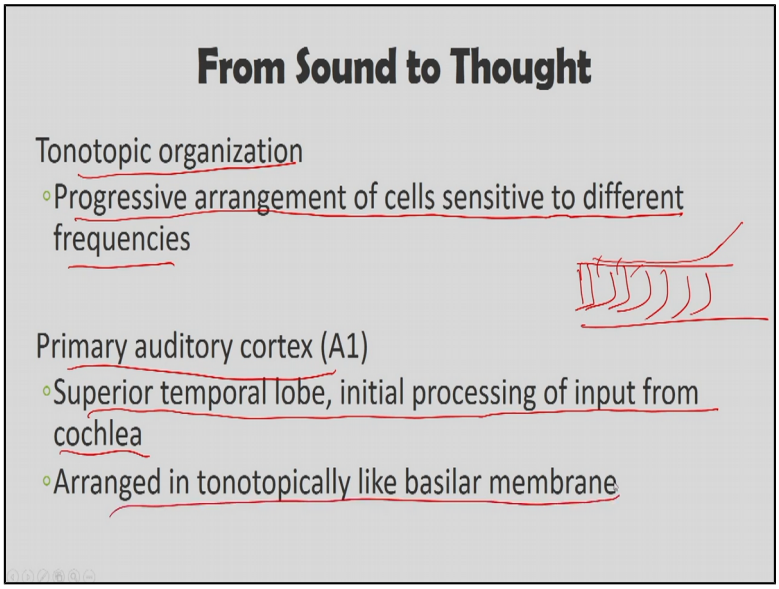
So, the Cochlea has a, Basilar membrane. It is, a two-part system. And so, the Basilar membrane is lined by hairs. It is filled with fluid. The sound, that we are producing, the wave pressure, the waves that we are producing, at different frequencies, they are perceived by the Cochlear hairs, and they are transposed, or they are changed into electrical signals, which are perceived by the, or picked up by the Primary auditory cortex, through a mechanism.

So, Cochlea, it is an organ of auditory sensation, in the inner ear. Basilar membrane, extends inside the Cochlea, undulated vibrations fluid of the Cochlea. As I said, the Cochlea has hairs. The Basilar membrane in the Cochlea has hairs. And, it is filled with fluid. And, in this fluid, the sound waves travel, and they are picked up by the, hair cells of the Basilar membrane.

Now, the Basilar membrane, which you have in the inner ear, which actually picks up sound, has something called a, Tonotopic organization. Now, what is Tonotopic organization? It means that, if you have ever seen a piano, you have high sounds, to low sounds, if you play it, this way. Different frequencies are range. So, lower frequencies to higher frequency.

So basically, what Tonotopic organization means, that it has sensitivity to different frequencies, at one end, and lower frequencies at the other end. So, you have, the hair cells are arranged in such a way that, higher frequencies are perceived at one end, and lower frequencies are perceived at the other end. So, that is how the, hairs are arranged. The Primary auditory cortex, is located, in the superior temporal lobe. And, it has the same Tonotopic organization, as the Basilar membrane.

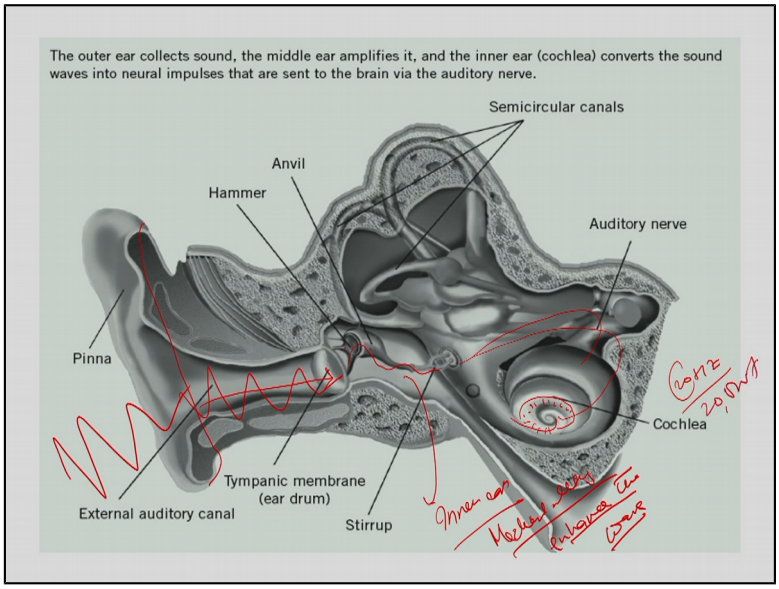
(Refer Slide Time: 27:29)



So, the Primary auditory cortex, which picks up the signal, from the Cochlea, it is arranged in the same way, that the Basilar membrane, inside the Cochlea is, which is Tonotopic organization. So, high frequencies. And, if you ever seen a piano, you have, things like this. Right. These are the keys of a piano. And, if you keep on pressing, you will see the frequencies, from low to high, it will keep on going.

So, what is Tonotopic organization. It is a progressive arrangement of cells, sensitivity to different frequencies. The Primary auditory cortex. This is the region, which gets input, from the hair cells of the Basilar membrane, inside the Cochlea. So, superior temporal lobe, initial processing of input from the Cochlea, arranged in tonotopically, like the Basilar membrane.

(Refer Slide Time: 28:18)

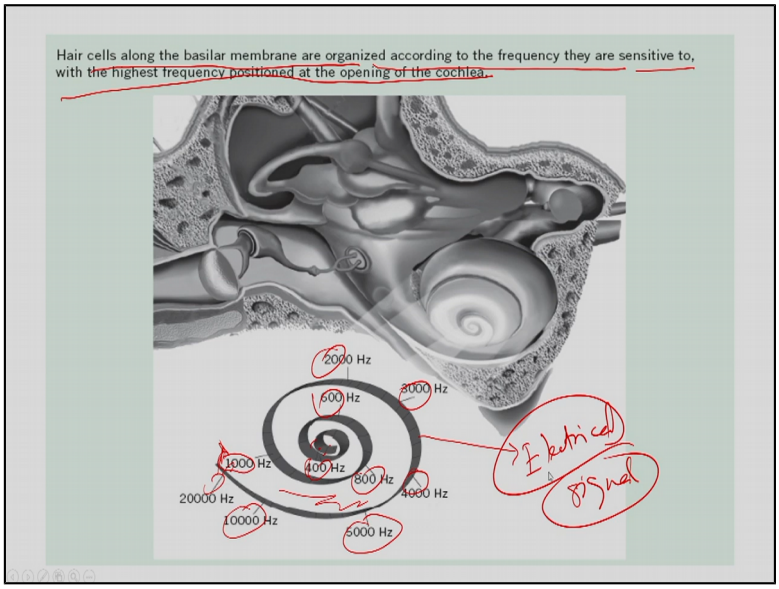


This is, how the ear is. So, when you actually hear something, comes here. So, even if you cut this, Pinna, External auditory canal, nothing happens here. The perception starts, or the actual hearing starts, here. This is my eardrum. Now, this is a membrane on which, the waves will come and bounce. And, this is attached to a three-part system, which is called, the Hammer, the Anvil, and the Stirrup.

And so, these wheels, this is the part of the inner ear. And, what they do is, they mechanically, they enhance the wave. And then, from here, it moves to the inner ear, which is the, this is where the Cochlea is. Inside this, you will have the Basilar membrane. Now, in the Basilar membrane, there are hair like structures, which are arranged. And then, it is filled with the sound waves. They move into this, Cochlea, or move through the, liquid of the Cochlea.

And, these are perceived, or these are picked up, by the hair cells, which are here. These are arranged from, 20 hertz to different hertz. Right. And so, different frequencies are there, right from 20 hertz. So, this is the beginning. 220,000 hertz, which is the limit. And, it can be picked, in that way.

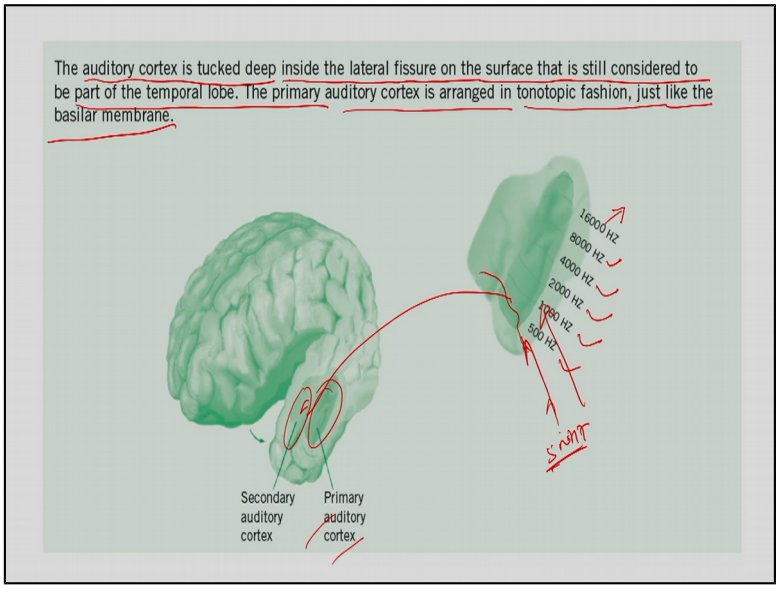
(Refer Slide Time: 29:59)



Hair cells are organized, according to the frequency, they are sensitive to, with the highest frequency positioned, at the opening of the Cochlea. As, you can see, this is 1,000 hertz, 20,000 hertz, 1,000 hertz, sorry 10,000 hertz, 5,000 hertz, 4,000 hertz, 3,000, 2,000, 6,000, 8,000, 6,000, 800, 600, 400, and so on, and so forth.

Since, hair cells are there inside this, this is filled with fluid, moves here. If it is at, 20,000 hertz, pitch that you are producing, it was picked up from here, we put it to the Primary auditory cortex, which actually takes this signal, and converts into electrical signal. Because, inside the brain, everything is electrical in nature. Sound waves are converted into, electrical signals, inside the ear.

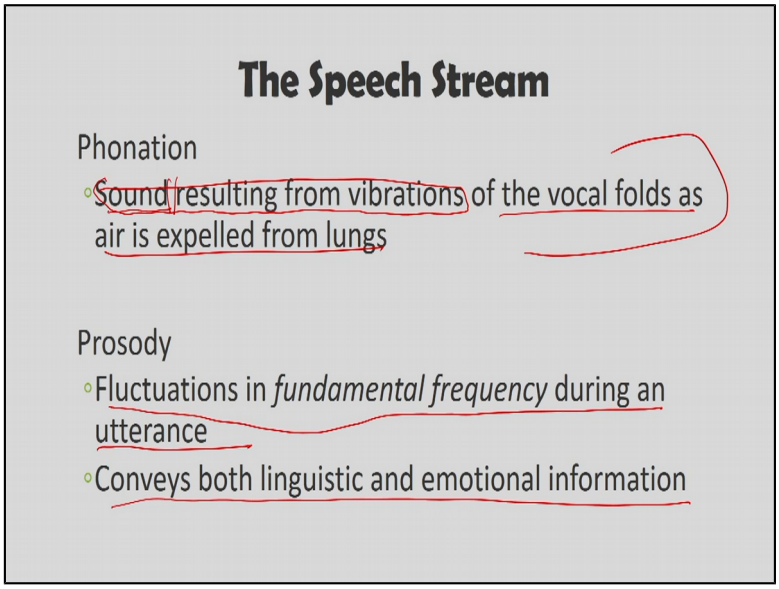
(Refer Slide Time: 30:52)



Now, the auditory cortex, is tucked between, inside the lateral fissure of the surface, that is still considered to be part of the temporal lobe. The Primary auditory cortex, is arranged in tonotopical fashion, just like the Basilar membrane. This is my, Secondary auditory cortex. This is my, Primary auditory cortex. It is arranged in the same way, in which the Basilar membrane needs. And so, as you can see, 16,000 hertz, 8,000 hertz, 4000 hertz, 2000 hertz, 1000 hertz, 500 hertz, and so on, and so forth.

And so, the Basilar membrane hair cells, connect to this, different, different. So, this is the 500 hertz Basilar membrane air cells, gives a signal here. Electrical signal from here, it passes. Primary auditory cortex passes, the secondary auditory cortex. And, secondary auditory cortex then, makes a perception of it. Basic idea is that, they are any in the same way, like a piano.

(Refer Slide Time: 31:45)



So, the Primary auditory cortex, is located in the superior temporal lobe. Temporal lobe. And, it has the same Tonotopic arrangement, as the Basilar membrane. Neighbouring cortical regions, including the Wernicke area, in the left hemisphere, do higher-level processing of the auditory input. So, Wernicke area then, takes him from the secondary processing area, and it does, high-level processing of the sound. Now, we cannot recognize objects and events, just by the way, they sound.

The idea is that, sound is enough for us, to recognize, objects. And so, we can recognize object just by, the sound of it. For example, falling glass, the whistling of the tree. We can just recognize, objects, or object moments, through sound. Speech sounds, are likewise auditory events, that are extremely short duration. Yet, our ability to perceive them, accurately, is remarkable. Speech sounds, although, most speech sounds are time-based. So basically, speech sounds, or auditory perception, is a time-based event. It moves across time.

So, single channel event. And so, what happens is that, they are very short duration, but we can perceive them. Think of, the last office meeting, you have. There is, so much sound. Everywhere, it was heard. All around, there was so much sound. But, you can very easily, in that in in in that humdrum of so much sounds, you can still perceive, different speech sounds, from different objects, and can locate, or can think about, what object is making that sound. That is the beauty, of the speech system.

So, auditory perception unfolds along a time dimension. It happens, across the time. When the time goes away, or it moves in a time dimension. So, you have to perceive it, across time, not

across space. As in visual perception, happens across space, but auditory perception, generally happens across time. So then, having said that, having looked at the wave, which is the Primary source of auditory perception, and the apparatus that makes sound, or that perceive sound, let us look at the speech, stream in itself.

When we speak, what really happens? Now, written text, what does the speech stream consist of. So, if you look at, written text. Let us look at, this text. Sound results from, vibrations. Let us look at, this text. Now, if you look at it, there are certain words, which are close together, and there are certain spaces, which are there, and so we can distinguish between words. So, written text, they have discrete letters, with each word, separating through spaces, from immediate neighbours. You can, know of that.

Speech stream, they do not have, discrete phonemes, and clear word boundaries. Speech is a continuous. So, when you speak, there is no space between it, there is no clear boundaries. Speech is continuous in nature. And so, speech is a continuous flow, of ever-changing frequencies, and amplitudes. When we speak, we speak on. And so, there is always, changing frequencies and amplitudes. Which are moving. There is no gap, between that.

Speech perception system, they infer, intended phonemes, and word boundaries. So, it is the beauty of the speech perception system, which generates boundaries, of where, there should be gap in the, spoken speech. If you ever look at a spectrogram, which is a machine, which converts, the spoken speech, into its fundamental frequencies, you will see certain regions, which are not populated, in speech output.

These regions, which are not populated, are actually not, periods of silence. There are certain consonants, which are called the stop consonants, and is basically the stop consonant, which is putting that, white space in the spectrogram. So, the spectrogram. So the spectrogram allows us to visualize, the structure of speech stream. Spectrogram is a machine, which takes in the speech stream, or the sound, which is coming out of your mouth, and it projects it. And, we can study the speech stream, through that.

Now, what is a spectrogram. The spectrogram or speech sample, consists of alterations of Aperiodic and Periodic sounds, roughly corresponding to consonants, all over the language. So, when we look at a spectrogram, it actually consists of Aperiodic and Periodic sounds, which

basically is that, it is consisting of consonants and vowels, how they are produced, by the speech stream. And, that is what is the display, actually showing.

And, as I said, the periods of silence, are actually consonants, which are called the stop consonants. Certain consonants in English language, which are called the, soft consonants. Now, at the bottom of any spectrogram, the fundamental frequency of the speech is displayed. This is the fundamental frequency, or the sound, resulting from the vibration of the vocal fold, as air is expelled from lung, and which is called the, phonation.

So what is phonation? Phonation is, the fundamental frequency of the sound, which is resulting from the vibration of the vocal fold, as air is expelled from the lungs. Sound resulting from vibration of the vocal fold, and air is expelled from the lung, is called, is what is called, phonation. So, the Primary frequency, that we have, the primary vibration, that the vocal tract is doing, is called the phonation.

And then, a related term is called, the Prosody. What is Prosody? Now, this fundamental frequency, the rate of vibration of the vocal fold, results from fluctuation, in the fundamental frequency, during an utterance. Now, this fundamental frequency, is never constant. The fundamental frequency is the, vocal fold vibration. It is different for, different people. Now, these fundamental frequency, is also not constant. It varies, across time, and so, across an utterance, across a speech.

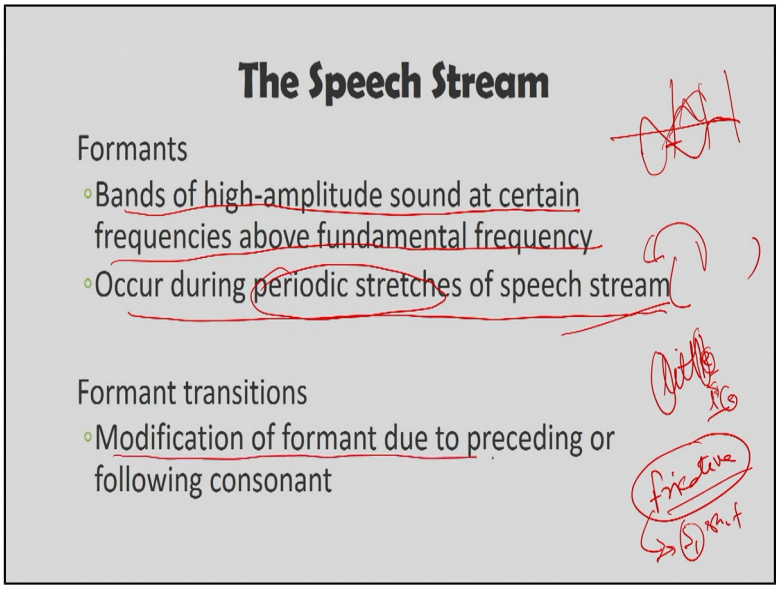
And, these vibrations in speech, in frequency, is called the Prosody. So, what is Prosody. The way you speak, when you speak, there will be variations in the, vibrations of the vocal tract, the frequency of vibration of the vocal tract. Of course, everybody has a fundamental frequency. But, this fundamental frequency, keeps on changing. And, that is what is, Prosody. So, what is Prosody? These are fluctuations in fundamental frequency, during an utterance, conveys both linguistic and emotional information, that is what a Prosody is.

And, why is Prosody important? Because, it conveys linguistic, and both emotional information. That is why, it is imp important. So, Prosody serves, a number of linguistic functions, as well as, conveying information, about the speaker emotional stream. How are you, if I say that? How are you, if I say that? Oh, how are you, if I say that. And, so the same way, the way we speak, this, how are you, that is basically, Prosody.

The change that, I am doing to the fundamental frequency. Everybody has a fundamental frequency. But, the change in the fundamental frequency, which I am doing, or the change in the pitch, that is what is called the, Prosody. The Prosody stretches, or the speech stream reveals, bands of high amplitude sound, at certain frequency, above fundamental frequency.

So, if you look at this, prosodic stretch, or speech stream, the change in the fundamental frequency, in an utterance, if you look at that, it reveals, bands of high amplitude sound, at certain frequencies, and above the fundamental frequency. So, there will be some part, this, this Prosody, will be high, at some part, about the fundamental frequency, and low, at some part. These BA bands are called, Formants. And, result from the fact that, the shape of the vocal tract, dampens certain harmonics, and enhances others.

(Refer Slide Time: 39:52)



So, Formants are what, these are the bands of high amplitude sound, at certain frequencies, above the fundamental frequency. When you produce sound, it is not always fundamental frequency. Sometimes, it is high above the, fundamental frequency. Sometimes, it is below the, fundamental frequency. The bands of high amplitude sound, at certain frequencies, above the fundamental frequency, are called the Formants.

And, why do they come? They happen because, these Formants happen because, the vocal tract, the one which is producing the sound, they dampen certain harmonics, and they enhance other harmonics. And, that is why, you get a Formant. They occur during, Periodic stretches of speech

stream, which is basically, the vowel. Periodic stretches is, yeah, not the vowel, basically. Now, so these are the, of formants.

So, this this change, that you are seeing, up in an utterance, about the fundamental frequency, and below the fundamental frequency, are basically, Formant. The relative distance, between the first and the second Formant, in comparison to the fundamental frequency, is used to distinguish, vowels. So, how do we distinguish, vowels. The change, that we see, between two alternate formants, and by comparing this, to the fundamental frequency, is how, we distinguish, different vowels.

Now, Periodic stretch, in the speech stream, may also signal, something called a Sonorant. Now, this Periodic stretch, may also signal something, called a Sonorant. So what is this? This is a speech sound, that usually serves as a consonant, but sometimes as a vowel. So, what is Sonorant. A Sonorant is basically, just like it is, we use a Formant, to distinguish a vowel, we use a Sonorant, to distinguish a consonant.

And so, what are Sonorant? These are speech sounds, that are usually serves as a consonant, but sometimes, they also act like, vowels. For example, the L, the R, the N, and M, in English language. Example, the first L, if I if I write Little, the first L, acts as a consonant, where the second L, acts as a vowel, because E is silent here. So, I said, Little, Little. I do not pronounce this E, and so this second L, acts as a vowel, or seems like a vowel.

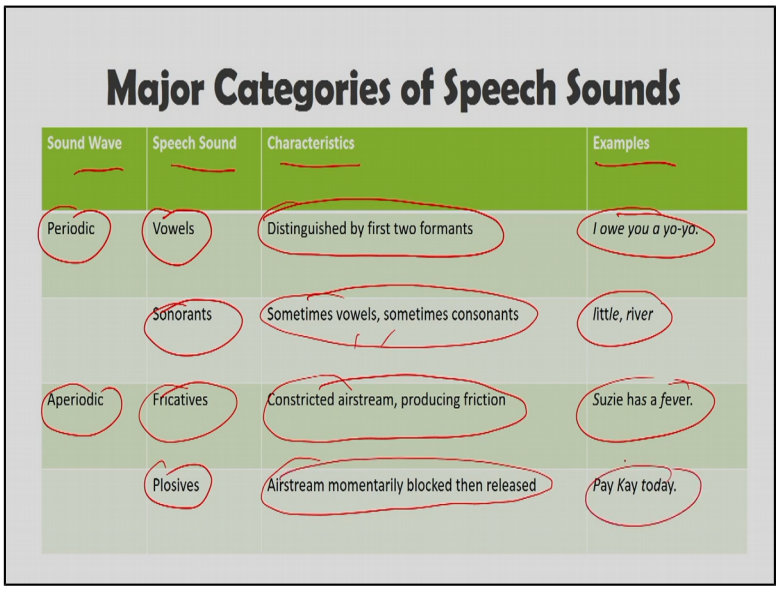
And, that is how, it looks like. Now, Aperiodic portion, of the speech stream, clearly indicates, a consonant. So, Periodic is vowel, Aperiodic is consonant. The first type, there are two types of, the first type of, Aperiodic stretch, Aperiodic portion, in the speech stream, which displays by spectrogram, is the Fricative. Two types of Aperiodic stretches. The first type is the, called the, Fricative.

What is it. It is a consonant, that is produced by, constricting the airstream, to create friction. Fricatives is, basically a consonant, which is produced, by blocking the stream of air, which is coming from the vocal cord, so that a friction happens. Try saying, S, SH, and F. So, SSSSS, the way you do it, the hissing sound, that is produced here, is basically the Fricative. And, this is the consonant, that I am producing, one type of consonant.

The second type of consonant, that is produced, or that can be seen, in a speech stream, in our Periodic speech stream, is called the Plosive. And, what is the Plosive. It is a consonant, that is produced, by momentarily blocking, and then releasing the airstream. So, in Fricative we, constrict the airstream, to create a friction, and out of that, the vowel comes in.

The second type is, we momentarily block the airstream, which is coming from the vocal cord, and then release the airstream. And, the vowel produced out of that, is basically, the Plosives. For example, the P, the D, the T, these are also called the, stop. So, basically then, there is also something called, the Formant transitions. And, these are the modifications, the Formant, due to the preceding or, following consonant. These are, transition in the Formant.

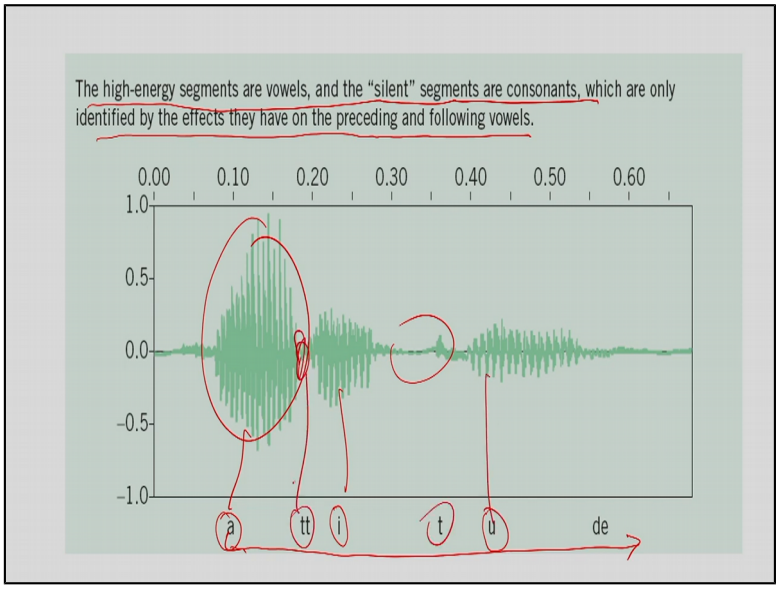
(Refer Slide Time: 44:36)



Now, major categories of speech sounds. Sound Wave, Speech Sound, Characteristics, and Examples. And, for example, Periodic, vowels, distinguished by the first two formants, I owe you a yo-yo. Similarly, the Sonorant, as I said, some of the some of the Sonorant, can act like a vowel. For example, sometimes vowels, acts as consonants.

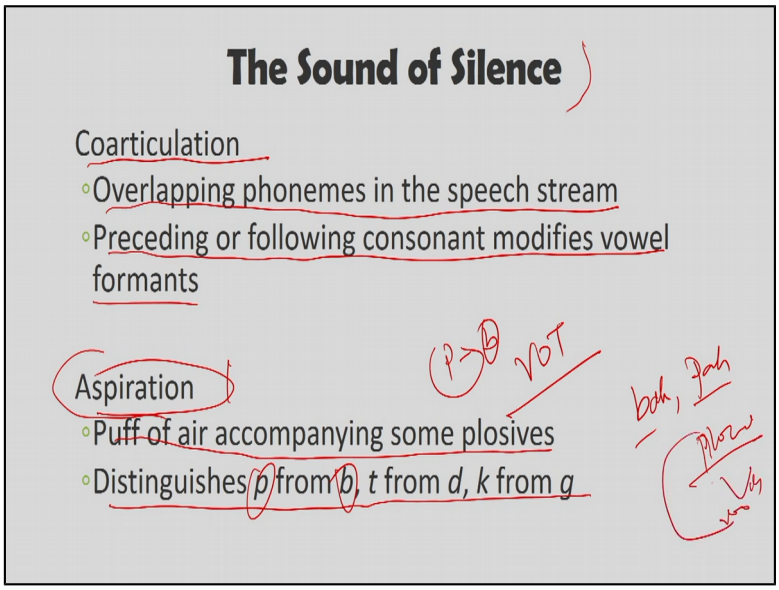
And, these are sometimes vowels, and sometimes consonants. For example, Little, River. Aperiodic, if it is, then I have, the Fricatives, which are constricted air stream, producing friction, Suzie has a fever. And similarly, if it is Plosives, then constructing air stream, and then releasing it. Air stream momentarily blocked, then released. For example, Pay Kay today. Basic speech sounds.

(Refer Slide Time: 45:29)



The high energy segments, are vowels. And, the silent segments are, consonants. This is my, vowels. This is my, consonant. Which only, identified by the, effects that they have a preceding for, vowel and consonant. And, this is the pronunciation, or the spectrogram, for the word, attitude, the way, ear says. This is my vowel. I, vowel, as you see, U, consonant, this region. Ok. There are, other interesting thing, in the speech sound, that is something about the Voice onset time, but will not take that now.

(Refer Slide Time: 46:12)



Now, we will move on to, how Coarticulation, and Aspiration, really works. What is Coarticulation? Overlapping phonemes, in the speech stream produces, Coarticulation. Preceding or following a consonant, modifies the vowel Formants. So, a consonant, modifies a vowel Formant, by proceeding, or succeeding it. So, although, they are not more than the brief bits of silence, Plosives are common in English language, and in languages around the world.

So, most English languages, and languages around the world, have something called, Plosives. And then, what is Aspiration. It is a puff of air, which is accompanying some Plosives, distinguishing B from P, and T from D. So, let us look at, what is Aspiration, then. Let us consider, consonant pairs like, B and P. Aspiration is the, sound of air, which is, or the puff of air, which is coming out, from the vocal tract. So, how do we din distinguish, B and P, from each other, the consonant B and P.

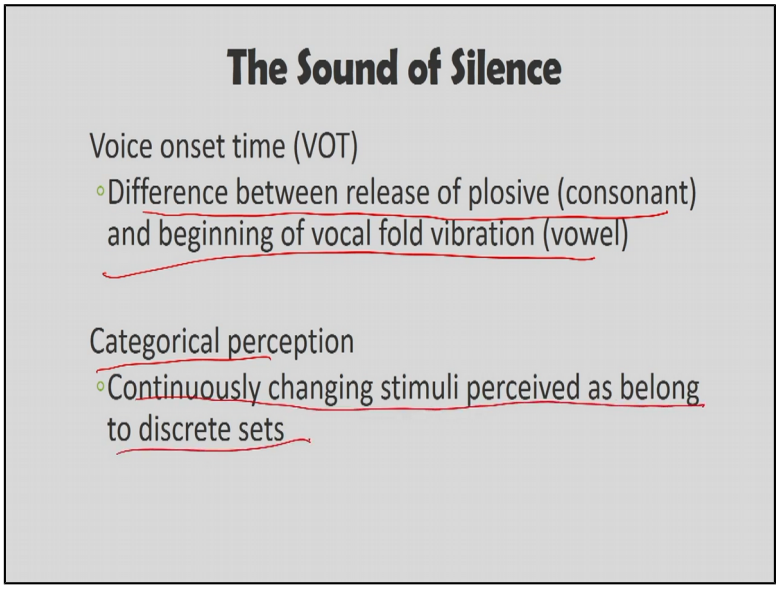
And, that is based on the, the Aspiration. So, vowels are easy, but consonants are little bit difficult. Now, imagine this. Put your left hand, on the throat, and the right hand, in front of your mouth, and twice saying, BA, and PA. Do that. If you do that, you will notice two things. There is more phonation, or vocal cord vibration, in BA, than PA. So, in terms of the left hand, it when you say BA and PA, in BA, there is more phonation, there is more vibration, in the vocal cord, but in PA, it is not.

And, how do you distinguish between, B and P, that kind of a Plosive. Also, if you look at the right hand, there is a greater puff of air, with BA, than PA. So, more puff of air, and more vibration, on the vocal track. Now, the puff of air, accompanying the release, of some Plosive is called, Aspiration. For example, if we look at word pairs like, DA and TA, and BA and PA, both of them, are accompanied by a puff of air, when they are produced.

And, this air production, when producing certain Plosives, which are certain type of consonants, are called, Aspiration. The two characteristics, of Phonation, and Aspirations, are generally explained, in terms of something called the, Voice onset time. And, what is Voice onset time? The voice onset time, so how are we measuring this. Between, P and B perception, it is in terms of the Voice on set time. And, what is the Voice onset time.

The Voice onset time, is the difference in time, between the release of a Plosive consonant, and the beginning of a vocal fold vibration, or the release of the consonant, the Plosive consonant, and the vibration of the vocal cord. The difference between that, it what is called the, Voice onset time.

(Refer Slide Time: 49:26)



It is the difference between, the release of a Plosive, a consonant, and the beginning of a vocal fold vibration, the vowel. Now, speech sounds, they do not occupy, discrete sections of speech stream, but rather, they overlap each other, in a process which is known as, Coarticulation. So, what happens, or what is Coarticulation is that, when we speak, the speech that we are producing, is basically overlapped, over one another. And, that is what is called, Coarticulation.

Speech sounds, are perceived, categorically, even though, they are produced differently, depending in the context. So, when we perceive speech, we do not look at the specific sections, of the speech, the specific Plosives, and or Fricatives, or vowels, we look at speech as a whole, and we perceive them in, certain categories. And, that is what is called, Categorical perception. So, speech sounds are perceived, categorically, even though, they are produced differently, depending on the context.

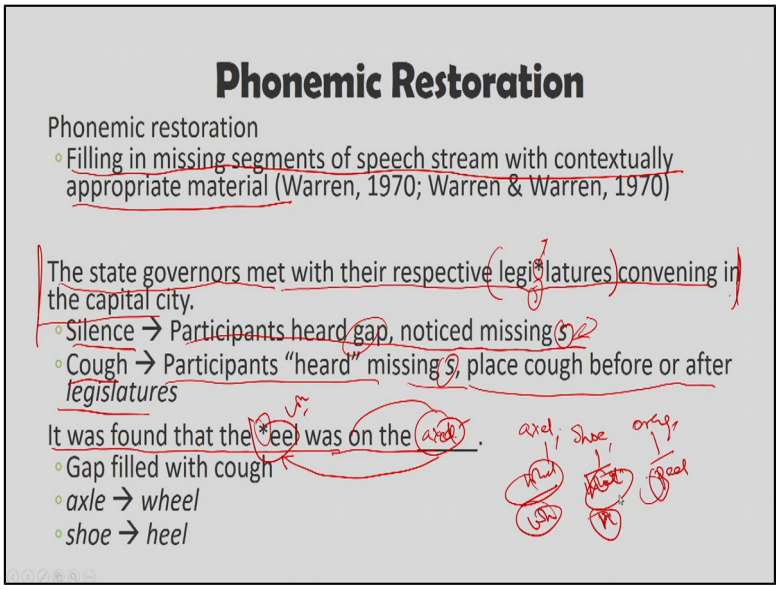
So, we categorize, certain speech sounds, and then we perceive them, together. Originally, categorical perception, was believed to be a unique characteristic, of speech perception. Now, it was soon understood, that as a general as a general. So, categorical perception is a, general cognitive principle. Simply put, categorical perception, is one way, our brains deal with the message, of the real world.

If you look at the real world, there is so much sound, which is there, and so much things, which are there, people are talking, things are falling, so many sounds are there. So, the best way for speech perception is, by putting category. So, this is human speech. This is, what the brain does

is, it separates from a different number of frequencies around you. It does, what it does is, it separates certain speeches.

And, it says that, this is human speech, this is the object speech, this is the speech of an animal, this is the speech of something. And, that is how, it perceives. So, it puts them into, certain characteristic, and that is how, it basically goes through, the everyday messiness. The speech perception system, relies on context, to fill the missing information. So, what is categorical perception? Continuously changing stimuli, perceived as, belonging to discrete sets.

(Refer Slide Time: 51:38)



We take, number of stimuli, and perceive them together. So, speech perception system, they rely on context, to fill in missing information, from a speech stream, that has been marked, by ambient noise, in the process, is known as what, fa Phonemic Restoration. When we speak, just to prove that, this categorical perception happens, if we are speaking, and certain segments of the speech are missing, the speech perception system is able to manufacture, or fill this thing. And, this is what is called the, Phonemic Restoration.

So, filling in, missing segments of speech stream, with contextually appropriate material, is what is called, Phonemic Restoration, so, if I have a context, if I have certain words missing, certain letters or certain parts of letters missing, from a speech stream, from a sentence, I can restore that. And, that is what it is called, Phonemic Restoration. For example, Warren & Warren, they tested this Phonemic Restoration effect.

What they did was, they produce sentences, like this, the state government met with their respective legislatures, and they put a star here, in the S, which is missing, convening in the capital city. So, this is the sentence, they read to people. This is the sentence, that is read to people. And, in this particular word, this missing S, was either, replaced by a cough, or by a silence. When a silence is produced, legi lators, this way, people are not able to restore the, S.

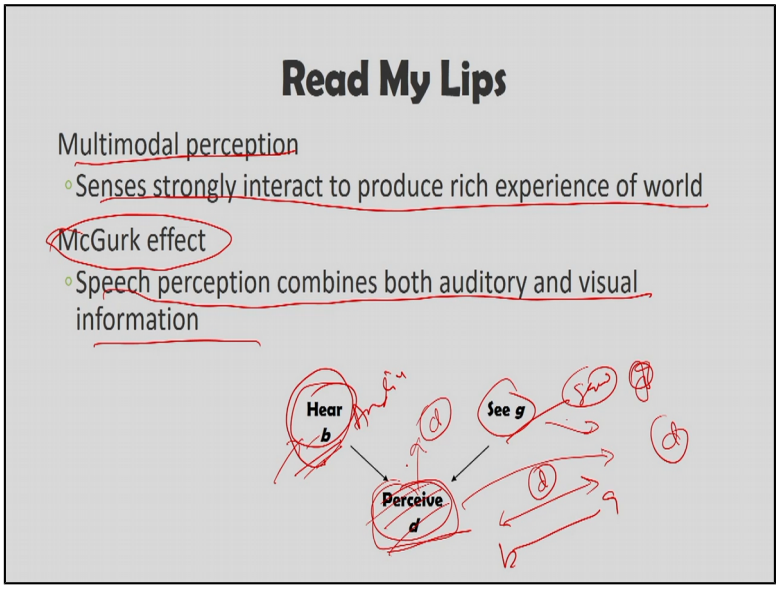
But, when people cough in between, for example, legi huhumm humm lature, legis huhumm latures, in this case, people are able to restore the, S. So, participants heard gap, noticed who had gap, they noticed the missing S. But, participants, heard missing S, place cough before, or after legislature. So, if you cover it, with a cough, then people restore this S. And, this is what it is called the, Phonemic restoration. People are able to, create the phone back, restore the phone back.

So, Warren & Warren in 1970, they modified the following sentence, by splicing out, the WH, from this word. So, it was found that, the WH, from this was, splicing out the WH from the wheel, it was founded, the *EEL here, people were able to, replace that. Right. It was also found out that, if I change this thing, here, the last word, the context will design, how what people are, pa filling in. So, what letters, what phoneme is being filled here, is dependent on the context, to prove that.

What Warren & Warren did was, they took a sentence, like this. It was found out that, the *EEL was on the. And, when they filled up, this last word with excel, people heard it was, wheel. But, when the last word was Shoe, people heard that, or people re reported that, they heard, Heel. So, in excel, they represented wheel, but in Shoe, they represented Heel.

And. That is the restoration, that was happening. So, it was context dependent. The restoration was, context dependent. So, this was replaced by, a cough, also excel was replaced by, Shoe. Similarly, Shoe. And, if I replace the Shoe, with an Orange, in here, people would say, found that the, Peel. So, P here, H here, and WH here, people were able to replace these, phonemes.

(Refer Slide Time: 55:21)



Now, Psycholinguist, used to assume, that the senses operate, independent of each other. It was believed, by Psycholinguist that, auditory perception happens on its own, visual perception happens on its own. And so, there is no link, between that. But, it was soon found out, that there is a relation between, auditory perception and visual perception. And, perceptions of auditory nature, are multimodal in nature.

In recent years, researchers agree, the idea, that the senses strongly interact with each other, to produce, our rich experience of the world. And, this is called the, Multimodal perception. So, Multimodal perception says that, senses strongly interact, produce rich experience. So, when we are producing auditory perception, the visual information is also included in it, in the in the auditory perception, and they integrate together, to give up the whole idea of the sentence meaning, that is being generated.

And, to test this, McGurk, what they did was, they produce an illusion. And, in this illusion, they made people, hear a particular word, they made people, see somebody speaking, a particular word, but there was no voice. So, in one case, people spoke, and so, you can auditorily hear, what the person is speaking. In the other case, you can see, read the lip of the person, which is producing, he is producing, a particular letter, or a word.

And, what they did was, they made sure, that people hear a D, and people see G, being produced. So, the speaker, in in in a visual form is producing G, in the auditory form, or the audio that somebody is hearing, he is hearing B. When he is seeing somebody speaking it, what

the experimenter, or the volunteer is seeing, G. And, well, it was a letter ask, what do you see, or what letter was being produced.

It was founded that, it was in between, B and G. What people perceived, or experimenters perceive was, D. So, somebody, who saw somebody producing, somebody who heard, somebody producing B, or volunteers, when they heard somebody producing B, in auditory domain, and saw him, or read his lip. And, when they read his lip, the person who was acting, was producing G. What later on they revealed, what this two people were producing was D, which was in between B and G.

So, speech perception combines, both auditory and visual information. This is the first effect, or the first time, it was found out that, speech perception, happens from, both the streams. It happens from the, visual stream, as well as the, auditory stream. So, the McGurk effect, is an artificially induced illusion, in which the auditory information from one speech sound, is combined with the visual information from another speech sound, to produce the perception of a third speech sound.

As you said, they hear, they see somebody speak speaking B, and they see somebody producing G. When later on, when they were asked, what is being produced, they report hearing D or C, D in between. So, that brings us, to the end of this session, of lecture number 5, where we looked at, how speech is produced for the first time. So, what are the characteristics of speech. So, we looked at, the characteristics of a wave.

And, we also looked at, the organization, of the human ear, which does perception of speech. Then, we moved on to looking at, how this speech perception, is represented, in terms of, the spectrogram. We looked at, how Aperiodic and Periodic sounds, which is vowels and consonants are represented. And, how these vowels and consonants are perceived, and then it is perceived back, by the speech systems.

We also looked at, speech perception in the multimodal form, and what are Plosives and Fricatives, and other kind of interesting things. How the speech is translated into thought? And then, how this, the speech stream, is actually perceived, or how it progresses. And then, we looked at, how the phoneme restoration effect, which is a, certain words, or certain letters, are not spoken in a speech stream, how they are covered up by the brain.

So, that is, that is what, we did in, today's lecture. When we continue on with this lecture, in the next session, what we will do is, we look at, how this system actually evolved. And, we look at, especially children, and try and see, how this system of speech perception, or speech production rather, they evolve, and several other interesting facts, related to, the production of speech. So, till we do that, when we meet next, and till that point of time, it is Goodbye, and Namaste, from the Studios. Thank you.