

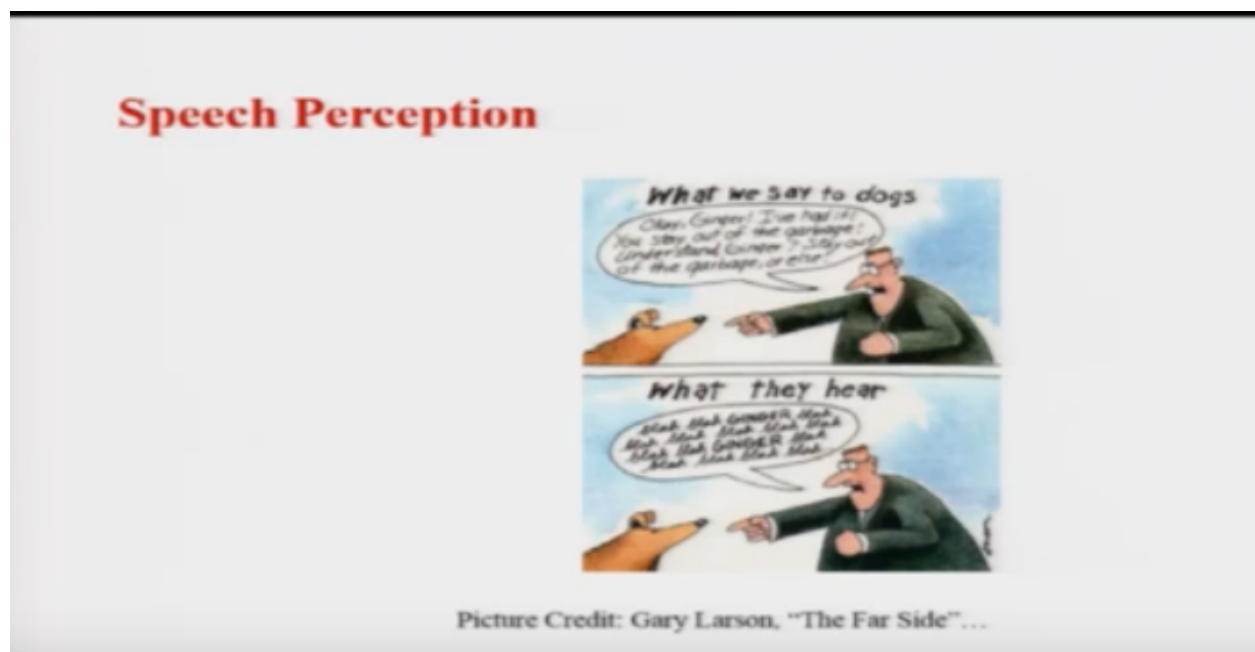
## **Lecture – 14**

### **Speech Comprehension - 1**

Hello and welcome to the course introduction to, 'The Psychology of Language'. I am Dr. Ark Verma, from IIT, Kanpur. And we are in the third week of the course. This week we have been talking about, speech production and comprehension. In the last three lectures, you saw, that I talked about different aspects of speech production. I talked about a couple of models of speech production and also some of the evidences that come to support those models. So we talked about the, Viva plus-plus model, given by level2, 1999. And we kind of saw some of the evidences, that come from, say for example, speech errors, tip-of-the-tongue phenomena or even, picture naming and picture word interference studies from the

normal individuals. These evidences, basically are a lot of experiment studies, that were done and some of the errors, that were observed and analyzed, to basically tell us, what part, of what step, of the speech production process, could be damaged or say for example, could be affected, to lead to those kind of errors. In the last lecture, I discussed a little bit about, dense spreading activation model of speech production. And we also saw some of the phenomena that, some of the basic assumptions that, Del's model had, which were different from, Viva plus-plus model. Both were kind of, slightly different models, but both of them kind of, in some sense shocked out the process in very similar ways. As I said, today's lecture, is going to be about speech comprehension. We will discuss a couple of theories, about how speech comprehension really happens. And basically what is it that goes for a particular listener to understand and comprehend, whatever speech signal that has been created, you know, how to decipher that. If you remember in the starting of this unit, I talked to you about the fact, that the process is, sort of cyclical. So the process is that you know, going from an idea, to really you know, formulating the idea into particular linguistic words and then going to the articulation part. You actually go to something that is actually physical. So you create a visible sound, which basically, is what that reaches the listener. So the listener has to basically do what? The listeners task starts, from hearing the sound and then from deciphering, what the sound really contains, go to the concepts that most probably, the speaker of these sounds, would have had. So it's sort of a cyclical process in, in the end. And the, first part of that cycling we have done in the speech production part, in the first three lectures, the second part of their cycling, cycling process, comes and the speech comprehension and this is what I will discuss in today's and yeah and the end tomorrow's lecture.

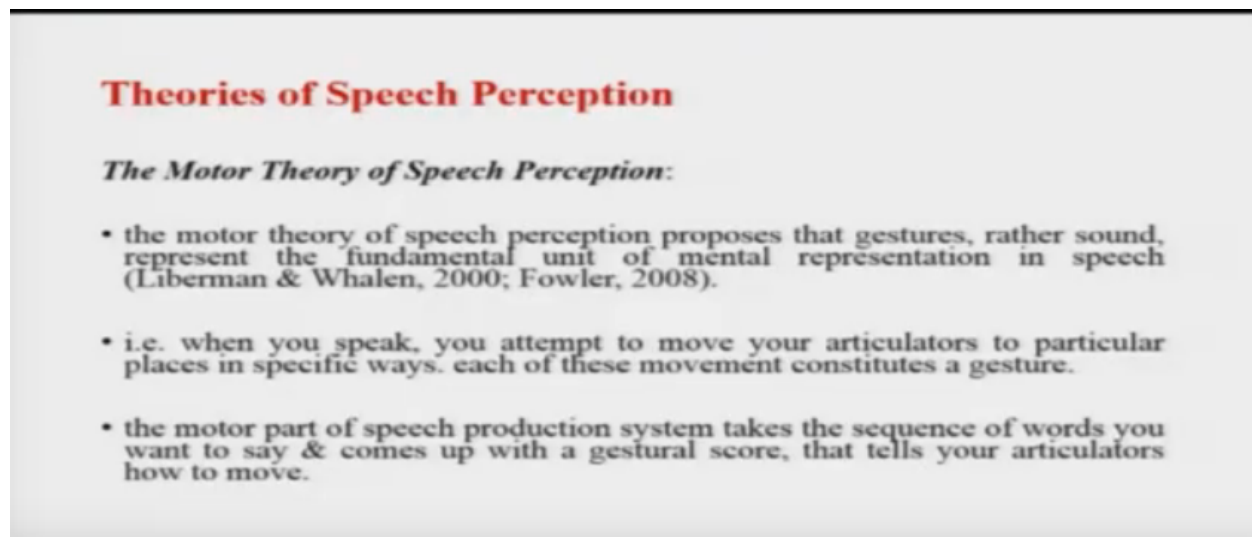
Refer Slide Time :( 2:55)



So how does speech perception begin? There is this cartoon, basically made by, Gary Larson and at the far side gallery, I kind of borrowed it, to illustrate a particular point. There is this gentleman who says, you know, 'What do we say to dogs? And it could be dogs or cats or any other person, whatsoever and you see that a person is saying, a lot of things. So he's saying, say for example, 'Okay? Ginger, how did

you stay out of the garbage, understand? Ginger stay out of the garbage, or else'. Then you know, they'd, it's almost a thread or something, something, something. And you see, is the, the next part is, pretty interesting. Because ginger is actually not hearing anything else, only ginger is, listening to his name. So, this is something which is very interesting and it's interesting to me, because, in essence of whatever you say, it should be intelligible, it should be understandable, to the listener that is what the essence of speech comprehension, has to be about. So we are, kind of going to look into some of the processes, that basically, go into making gingers understand, how you know, the speaker is speaking, from where he's coming from and ginger has to kind of achieve this, in order for conversation or communication, to work effectively. Because if only one person speaks, and the other person does not really understand, it's not really, you know, a great loop, that is being completed and that's communication, which is sort of a loop. Okay? So we'll try and understand, some of those things related to understanding, what the other speakers say.

Refer Slide Time :( 4:27)



**Theories of Speech Perception**

*The Motor Theory of Speech Perception:*

- the motor theory of speech perception proposes that gestures, rather sound, represent the fundamental unit of mental representation in speech (Lieberman & Whalen, 2000; Fowler, 2008).
- i.e. when you speak, you attempt to move your articulators to particular places in specific ways. each of these movement constitutes a gesture.
- the motor part of speech production system takes the sequence of words you want to say & comes up with a gestural score, that tells your articulators how to move.

Now the theory perception, there are a couple of theories of perception or maybe more that we will discuss in these, two lectures. But I'll begin with this, very interesting theory of speech perception that is called, 'The Motor Theory of Speech Perception'. Now I have been saying, if you remember the speech or language or speech production, so to speak is a rather sophisticated, specialized, motor activity. If you remember I said, it takes around hundred muscles, to be controlled, in order for you to articulate, exactly without an error. Whatever you intend to speak, at any point in time. So it is a motor act. So it's not really surprising that, there is a motor theory of perception. But let us see, what the motor theory of perception, has to say. So, the motor theory of perception basically proposes, that gestures rather than sounds, should be the fundamental unit of mental representation, as far as speech is concerned. Okay? So for example, as I said, what the speaker creates in the end, is the sound. However, people have expressed doubts, sound is not really a very faithful representation, of what is coming out of different speakers' mouths. Because, of the reason that, a different people have different kinds of voices, they have different types of vocal tracts and they can create, different entirely different kinds of sound signatures, even if they are speaking about the same thing or even if they are speaking the same sentence. So that kind of probably, you know, led to

some of these doubts and there is this interesting proposal, more specially by, Alwyn Lieberman, who says that, 'No let us not really go by someone, where the sound is not really a very faithful representation, let us go by motor acts', and by motor acts, he basically means, that you know, when you speak, you move your articulators, you know, the tongue and the lip and the teeth and the vocal cord and all of that, you move your articulators, to particular places, in very specific ways. I will, in the lectures that, are going on in this week, attach, some of the indicated videos of, that will help you understand, it should not matter of articulation or voicing really matter. And if you look at those videos, they are not our videos; they are basically somebody else's. But if you look at those videos, what you will understand is ,that, there is a lot of difference, in different kinds of speech sounds, depending upon, how they are produced. So the gestures that actually go in, for us to be able to create, a particular speech sound, is something that is very important. So much so, that Alwyn Lieberman believed, that it is these gestures, that we make the sounds, that will make the speech, identifiable, recognizable and discretely, different from each other. And that is why he said, the speech perception, basically should be based on our perception of these gestures, these motor acts. So he says when you speak, you attempt to move your articulators, in two particular places, in very specific ways. And each of these movements, basically that you do, constitutes what is called, 'A Gesture'. And we should kind of, try and analyze speech, in terms of the gestures, that basically were made to, create this speech. Okay? So the motor part, of the speech production system, it takes the sequence of words that you want to say and it comes up with a gestural score. We talked about that, in one of the earlier classes. This tells our articulators, how to move. So as I said, gestural score is a sort of a, executional level program. Okay? And this executional level program, basically tells you, okay, move articulator one to this point and you know, constrict the flow of air to this much and then leave it and then this is how, it goes. So this sort of a program is, is very important, in producing all the ends of speech. Refer Slide Time :( 8:12)

- Acc. to the theory, if you can figure out what gestures created a speech signal, you can figure out what the gestural plan was, which takes you back to the sequence of syllables or words that went into the gestural plan in the first place.
- So, by knowing what the gestures are, you can tell what was the set of words that produced that set of gestures.
- For e.g. the "core" part of the gesture to produce either "di" or "du" sounds is tapping the tip of your tongue against the back of your teeth (or your alveolar ridge).
- other parts of the gesture, like lip position are affected by coarticulation, but the core component of the gesture is the same regardless of the phonological context.

Now according to this theory, if you can figure out, what gestures were used to create a particular speech signal, you can actually figure out, what the gestural plan was, which will take you, back to the sequence of syllables or words, that the speaker was intending, to say. So you kind of not really you know, very faithfully, looking at the speech sound, that is created, that is obviously something that you hear, but what

you're trying to do, by listening to that speech sound, is, you're trying to make out what gestures were made, to come up with this kind of a sound. Once you kind of figure out the gestures, say for example, I have to say, 'Tata'. Okay? So what do I do? I am kind of creating, 'Ta', by keeping my tongue at the back of the teeth and I'm constricting the flow of air, only partially and then I'm kind of, you know, leaving it, for me to be able to say, 'ah'. R does not require any, you know, constriction of air, but, 'Ta', for example, requires. So I am kind of doing this, in a sort of a, 'Uh, Ta, Ta, uh, ta, uh', sort of a manner. Okay? So what you will really need to do is, in order to really specifically understand, what I am saying, is that, you kind of, make sure of, what is the gesture that I am making. And a motor theory says, if you can do, that you will be closer to understanding speech, as compared to, if you're only relying on sounds. Okay? So by knowing, what gestures, what these gesture are, you can tell, what was the set of words, that has produced, that set of gestures. And let's take an example, for the core part. Let's say, we take two sounds in, in hand, let's say, we take 'di' and 'du'. The core part, the basic the more the stem of this sound, is the, 'du'. Okay? And basically, how is the produce? Very similar to T, but slightly, in a, slightly different way. So it's by, tapping the tip of the tongue, against the back of the teeth, as I said for 'tu' as well. But you can kind of make sure is, how 'du', because 'du' in 'du' the teeth is slightly more further back. In 'Tu' that tongue, is a slightly further forward, still at the back of the teeth, though. Okay? So the idea is, you have to kind of figure this out, you have to, you know, spend your resources, in really figuring this out. Okay? So other parts of the gestures, like the lip position, which you make in terms of, if you're saying, 'ooh', as in do or D, as in e, are actually also affected by Co articulation. But the core component of the gesture, the 'du' part, is regard, is the same and it's a common, regardless of, accompanying phonological sounds. Refer Slide Time :( 10:45)

- Thus, rather than trying to map acoustic signals directly to phonemes, Alvin Liberman & his colleagues proposed that we map acoustic signals to gestures that produced them, as there is a closer relationship between gestures and phonemes than there is between acoustic signals & phonemes.
- In their words, "The relation between perception & articulation will be considerably simple than the relation between perception and the acoustic stimulus."
- Further, "perceived similarities and differences will correspond more closely to the articulatory than the acoustic similarities among the sounds."
- So, differences between two acoustic signals will not cause you to perceive two different phonemes as long as the gestures that created those two different acoustic signals are the same.

So, rather than trying to map the acoustic signals, directly to phonemes, Alban Liberman and his colleagues, proposed, that we should map acoustic signals, through the gestures, that produced it. Because they believe, that there is a much closer relationship between the gestures and the phonemes, because the gestures are directly responsible, for creating the core part, of those phonemes. Okay? In their words, they say, the relationship between perception and articulation, would be considerably simpler, than the relationship, between perception and acoustic stimulus. Just kind of spend a second thinking about this.

This is exactly what I was saying a minute ago, that, all of us have different kinds of vocal tracks, there are so many disturbances in the environment. It is often possible that, you will confuse, what you hear, from the person, however if you kind of, can get a reasonably, decent picture of, what the person was saying and you kind of decipher the gesture, you can be relatively sure of, what has been said, that's basically, what the point is, from Alban Lieberman and colleagues. So they say further, perceived similarities and differences will correspond more closely, to the articulatory, rather than, the acoustic and similarities, among the sounds. So they say, that you have to base your entire understanding of whatever you're hearing, on these articulatory gestures, because they are, closer in time, you know? You make this gesture, there is a sound that comes and that is what you hear, this part is not really, the middle part is not really very faithful. So, if you can kind of, you know, skip that part, in some sense or use that part to come to the, artillery gesture part, then you will have a much stronger theory of what was said. Okay? That's, that's basically what Lieberman is saying. So the differences between two acoustic signals, will not cause you to perceive, two different phonemes, as long as the gestures, that created those two different acoustic signals, are the same. So you kind of, if you're going by gestures, that will cause you less confusion, you'll probably be closer, you'll probably be better, in understanding, what the sound was, by inferring, what are the gestures that created that sound. Okay?

Refer Slide Time :( 12:49)

- Motor theory also seeks to explain how a person can perceive an acoustic stimulus as a phoneme in one context but as a chirp or a buzz in another context.
  - to explain that, the motor theory proposes that speech production is accomplished by a naturally selected module (Fodor, 1983).
  - this speech production module monitors incoming acoustic stimulation and reacts strongly when the signals contains the characteristic complex patterns that make up speech.
  - when the speech module recognised an incoming stimulus as speech, it preempts other auditory processing systems, preventing their output from entering consciousness,

So let's move further. Motor Theory, it also seeks, to explain how a person can perceive an acoustic stimulus, as a phoneme in one context or as a chirp or a buzz, in another context. Okay? The same phoneme is said, but in one context, you perceive it, as a chirp or a buzz, 'zzzzzzzz', something like that or as a phoneme like, 'ba, ba,', but something like that. Okay. So to explain that, the motor theory proposes that, speech production is basically, accomplished, by a naturally selected module and module is typically, even this understands this, module theory was given by, Jerry Folder. If you really want to understand a little bit more about it, you can go into, so the cognitive psychology lectures, I have given. But module is, an understanding of that, different cognitive functions, are organized, in different, you know, modular systems and these modules are sort of, boxes, which are informationally, encapsulated

again, I don't want to throw jargon, but that basically means, that, if there is a job, say for example, is if there is visual perception and it's there, if there is learning and if there is, say for example, understanding, all of these three things, will be modular, in nature, the output of the perception process. So the perception process will be, self-sufficient in itself, it will just give an output, which kind of, you know, taken care of by a language module or say for example, speech perception module, as you know, Lieberman and guys say. The speech perception module is sort of, self sufficient in itself and it only interacts with the other, kind of cognitive functions, by virtue of input and output. Okay, now so they say, that a motor theory, you know, says that speech production is basically accomplished, by what is called, a naturally selected speech processing module and the speech processing module monitors, the incoming acoustic stimulation and it reacts strongly, when signals contain, the characteristic complex signatures, that make up speech. So there is this, particular module in your head, somewhere, just imagine that for a bit and that particular module is very closely monitoring, everything that you're hearing. And you hear so many different things. You hear motor sounds, you hear horn sounds, you hear the animal sounds, you hear, you know, all, all different kinds of sounds. This particular module is lashed down to, you know, reacting when it detects, that you were hearing speech. This is where the, you know, the module kind of, springs into action. Now the speech module, recognizes an incoming stimulus as speech, it pre-empts the other auditory processing systems, preventing their output from entering consciousness. So it's basically saying that, there is this speech processing module, that basically comes into action, when it detects, something as speech. Because it detects something coming as speech, it takes the input and it tries to process that input, differently, to how, other kinds of sounds will be processed. So the other kind of sounds could be basically, you know, things that will probably do a frequency analysis, they will probably, kind of you know, do a pitch and other kinds of analysis. With speech, you'll probably not do that or you probably do that, in a different way. So as soon as this particular module discovers, Okay, this particular incoming acoustic signal is, speech, it takes, it pre-empts every other process, it says, Okay, you don't really have to do anything here, I will handle it and it takes the input, to a different place. Again very, you know, metaphorically speaking and it kind of, trials in processes at, in a very simple, in a very different sort of manner.

Refer Slide Time :( 16:25)

- So, while the non - speech sounds are analysed according to the basic properties of frequency, amplitude, and timbre, and while we are able to perceive those characteristics of non-speech sounds accurately, when the speech module latches onto an acoustic stimulus; it prevents the kind of spectral analysis that general auditory processing mechanisms generally carry out for non - speech auditory stimuli.

- this *principle of preemption* explains why formant transitions are perceived as chirps or high - pitched whistles when played in isolation; but as phonemes when played in the context of other speech sounds.

Now, while the non speech sounds, are analyzed, according to the basic properties, as I was saying, basic properties of frequency, amplitude, timbre and while we are able to perceive those characteristics of ,non speech sounds accurately, when the speech module latches onto the incoming acoustic stimulus, that may be speech, it prevents this kind of spectral analysis, of that speech signal. Okay? That basically would, be done, by the general auditory processing module or mechanisms as you would say. This principle of preemption, it explains why formant transitions, are sometimes but she, perceived as chirps or high-pitched whistles, when played in isolation, but as phonemes, when played in the context of other speech sounds. So say for example, formants are basically frequency spectra, of incoming speech stimulus. So what you could do is, you could kind of analyze these formants and you could kind of, play them, using a particular computer program and you could hear, what these formants, what kind of sounds, these formants are representing. So the speed, the principle of preemption sort of tells us that, sometimes when you play these formants, you might hear them, if you are just playing them in isolation, you might hear them as, chirps or buzzes. But if you play them with the entire context, you might, you know, hear them as, particular phonemes or speech sounds.

Refer Slide Time :( 17:43)

- the preemption of normal auditory perceptual processes for speech stimuli can lead to *duplex perception* under special, controlled lab conditions (Lieberman & Mattingly, 1989).
- to create their experimental stimuli, researchers constructed artificial speech stimuli that sounded like /da/ or /ga/ depending upon whether the second formant transition decreased (/da/) in frequency over time or increased (/ga/).
- next, they edited their stimuli to create separate signals for the transition and the rest of the syllable, which they called the *base*.
- they played the two parts of the stimulus over headphones, with the transition going in one ear & the base going in one ear.

Now this preemption of normally auditory perception processes for speech stimuli and sometimes leads to what is referred to as, duplex perception, under special and controlled slab conditions. Now duplex perception is, when you can hear the same signal, same acoustic signal, at times, as a chirp or a buzz and at different times as, a particular phoneme. So Lieberman & Mattingly, 1989, they did this very interesting experiment. What they did was, they tried to create their experimental stimuli, using artificial speech stimuli, that either sounded like 'ga' or 'da'. Depending upon whether the second formant transition, decrease in frequency, over time or increase. So you can say for example, imagine this, as this is a set of four men going, it could either increase, you know, to have the speech sound or it could, I, come like this and either decrease. So what they basically do is, they take this part, in the flat part away and they take the ascending part or the descending part, away. What they do is, they edit their stimuli in such a way, that they create separate signals, for the transition parts, this one's and the rest of the syllable, which is called the, 'Base'. So this one is called a base, this is called the transition, this one is called the



base, this is called the transition. This is what the special analysis, looks like. So what they do is, they separate the base and they separate the transition. Okay? And base is, so for example, I'm talking about, da and ga, so the base of da and ga will also be different, the transition of da and ga will also be different. So you kind of, get four, different components. What do you do is, you play them to participants in separate ears. So the idea is; they played the two parts of the stimulus over the headphones, with the transition going in one ear, suppose the transition comes to my right ear and the base comes to my left ear. This is how they do it. What did they find out of that?

Refer Slide Time :( 19:35)

- the question was, how would people perceived the stimulus?
  - it turned out that people perceived two different things at the same time. at the ear that the transition was played into, people perceived a high - pitched chirp or whistle. But at the same they perceived the original syllable, just as if the entire, intact stimulus had been presented.
- Liberman & colleagues, argued that simultaneously perceiving the transition in two ways - as a chirp & as a phoneme - reflected the simultaneous operation of the speech module and general purpose auditory procesing mechanisms.

The question was, the basically we were trying to ask the question, how would the participants, they perceive the stimulus? Will they combine these two and listen them as a proper phoneme? Will they listen to them as separate, you know, frequency things, which could be, you know chirps or buzzzz's or something like that? Now what they find out, is that, people perceived two different things, almost at the same time. At the ear, the transition was played into, people perceived a high pitched chirp or a buzz, so this part, where ever, say for example, in the right ear, I will perceive, a high-pitched chirp or a buzz. But at the same time, they also perceived, the original phoneme, which is basically, because they probably in, in some sense of time, they combined this and this and they heard the actual phoneme, that was played and split also. So this is an example, of what I was saying, as duplex perception. Okay? So, Lieberman colleagues, argued that simultaneously perceiving the transition as two ways, as the chirp and a buzz and as a phoneme, it reflected the simultaneous operation of, both kinds of speech modules, you know. The general auditory module, which, which you know, analyzes all kinds of vehicle incoming acoustic stimulation and the speech module, which basically only is specialized, for analyzing and you know, understanding the speech sounds. Alright?

Refer Slide Time :( 21:04)

- duplex perception happened because the auditory system could not treat the transition and base as coming from the same source (as they were played in two different ears).
- because the auditory system recognised two different sources, it had to do something with the transition that it would not normally do., i.e. it had to analysed it for the frequencies it contained and the result was hearing it as a chirp.
- but simultaneously, speech processing module recognised a familiar patten of transitions and formants & as a results the auditory system reflexively integrated the transition & the base and led to the experience of hearing a unified syllable.

So, moving further, duplex perception happened, as I said, because the auditory system, could not read the transition and base, as coming from the same source, because the auditory system recognized, two different sources, it probably had to do something with the translation, that it would not normally do. So it analyzed that, as a non street sound, in terms of frequencies, that is why you hear the chirp and the buzz. But simultaneously speech processing module, recognize a familiar pattern of transitions and formants and as they combine, the base and the transition. And that is, what resulted, in the auditory system, reflexively, perceiving this as the phoneme that was a split. Okay? So that leads to the perception the unified syllable. Now moving further,  
Refer Slide Time :( 21:44)

- Acc. to the motor theory, *categorical perception* is another product of the speech perception module.
- categorical perception happens when a wide variety of physically distinct stimuli are perceived as belonging to one of a fixed set of categories.
  - for example: every vocal tract is different from every other vocal tract & as a result the sound waves that come out of your mouth when you say *pink* are very different that the sound waves that come out of my mouth when I say *pink*, and so on.
  - nonetheless, your phonological perception is blind to the physical differences and perceives all of those signals as containing an instance of the category /p/.

according to the Motor Theory, the categorical perception, is another product of the speech perception module. What is categorical perception, if I may ask, if you go back, if you remember the last week's

lecture, I talked about categorical perception? What is categorical perception? Categorical perception is our ability to, perceive different phonemes, as categorically different, from each other. You know? Perceiving 'Pa' and 'Ba', as different phonemes, that's what categorical perception is. Now categorical perception happens, when a wide variety of physically indistinct stimuli, perceived as belonging to one of, fixed set of categories. So these, these, stimuli's, are basically happening, are coming from this category. So 'Pa' and all variations of 'Pa', are in one category. 'Ba' and all variations of 'Ba' are in another category. Let's take an example. Every vocal tract is different, as I have been saying, from every other vocal tract and as a result, the pattern of sound waves that come out of your mouth, when you say, pink or my mouth, when I say pink, are very different. How does the system understand? How do we both hear pink, even if you are saying it or I am saying it? Nonetheless, so this is basically what happens, the phonological perception system is blind, to the physical differences and perceives all of these variations as, one category. Okay? If you remember the whole experiment with, the Kikuyu children, you might go back and refer to that, if you don't remember it, right now. Now, it may be noted that, because all of our voices, have different qualities, than each other,

Refer Slide Time :( 23: 14)

- It may be noted that all of our voices have different qualities than each other, but we categorise the speech sounds from each of us, in much the same way. This is because, all of those different noises map to the same set of 40 phonemes (in English).
- In addition, although the acoustic properties of speech stimuli can vary across a wide range, our perception does not change in little bitty steps with each little bitty change in the acoustic signal.
- We are insensitive to some kinds of variation in the speech signal, but if the speech signal changes enough , we perceive that change as the difference between one phoneme and another (Liberman et al., 1957).

but weaker, but at the same time, we categorize the speech sounds, from each of us, in much the same way. This happens, because, any sound that you would make, say for example, in English, will broadly need to be mapped on to this very limited space, of 40 phonemes. So you will probably, lump together, so many variations and tell that this is the category, this variation belongs to. You know, it's the whole concept of phonological prototypes, being created. Now in addition the acoustic properties of speech stimuli, can vary across a wide range. Okay? Our perception does not released change in very little steps and small steps, with there each of it very small changes, in this thing. So it does change, but it does not really change very slowly. Okay? So you're insensitive, that is one of the reasons, that we are insensitive, to some kinds of variations in the speech signal, that are happening. But, suppose the speech signal changes, enough, you know, from 'pa, pa, pa, pa, pa, pa', it changes to 'ba', you know. If the variation is large enough, then obviously we will, start perceiving the two sounds, as different stimuli. Okay?

Refer Slide Time :( 24: 24)

- An example:
  - the difference between /b/ & /p/ is that the /b/ is voiced while the /p/ is not.
  - other than voicing the two phonemes are essentially identical; in that they are both *labial plosives*, meaning that we make these sounds by closing our lips & allowing air pressure to build up behind our lip dam and then releasing the pressure suddenly, creating a burst of air that rushes out of the mouth.
  - the difference between the two phonemes has to do with the timing of the burst and the vocal fold vibrations that create voicing.
  - for the /b/ sound, the vocal folds begin vibrating while your lips are closed or just after; but for the /p/ sound, there is a delay between the burst and the point in time when the vocal folds begin to vibrate. This gap is the *voice onset time*.

Let's take an example. I am taking this 'ba' and 'pa', example again and again. So the difference between /b/ and /p/ is /b/ is voiced, while 'pa' is not. What is voicing? Voicing is basically, when your vocal cords, vibrate after you said something. Okay? Other than voicing, the two phonemes are essentially identical, both are produced using, exactly the same gesture. So in that they are both, labial plosives. Again, the video is talking to you about, manner of articulation, pace of articulation, you kind of get, sure of these concepts, a little bit better. Labial plosives meaning that, we make these sounds, by closing our lip and allowing the air pressure, to build up, behind our lip dam and then releasing it, slowly. Okay? The difference between the two phonemes has to do with, the timing of the burst, the timing, when you allow the air to come out and the gap between which your, verbal portion, will start migrating. So I say /b/, versus, I say, /p/. Okay? There is a difference between, the time, that the vocal cords, will start vibrating. For the /b/ sound, the vocal cords, begin vibrating, while your lips are closed, they already start vibrate, so 'Buuu', the vibration is already building up. And /p/ there is a slight delay, you know. A slight delay of around 20 milliseconds or a little bit more. So /p/. Okay? This delay, basically is referred to as, 'The voice onset time'. So in other words, you could say, the difference between, /b/ and /p/ is that of voice onset time. Okay?

Refer Slide Time :( 25: 51)

- the VOT is a variable that can take any value whatsoever, so it is called a continuous variable. but even though not can vary continuously in this way, we do not perceive much of that variation. for e.g. we can not greatly hear the difference between a bot of 2ms and 7ms or between 7 ms & 15ms.
- instead we map a range of votes on the same percept. Those different acoustic signals are called *allophones* - different signals that are perceived as being the same phoneme. so the experience with a range of short VOTs is as /b/ & long VOTs is as /p/; the difference point being 20ms.

So this VOT is a variable, that can take any value, whatsoever. So it is called a continuous variable. But even though, it cannot vary, but even though it cannot vary, in continuously, in this way, we do not perceive that sort of variation. So there can be different kinds of voice onset delays between different speakers, saying this. Okay? So we cannot really, what happens is, we cannot really, greatly hear the difference between a, VOT of, 2 milliseconds or 7 milliseconds or between 7 milliseconds, in 15 milliseconds. So what happens, is that, we hear a range of VOT's, on the same preset. Okay? Those different acoustic signals are called, 'Allophones'. Allophones basically are, different signals, that are perceived, as being the same phoneme. So different versions of /p /p /p /p /p /, all of that, with slightly different, you know, VOT's, will all be perceived as, the same category. These are, these will be referred to as, 'Allophones'. Okay. However, if the difference basically, is slightly larger, if the difference is basically around a, 20 milliseconds mark or more, then what your happens is, that you start perceiving, the and the same kind of gesture, as a different phoneme. So till a particular point when the variation is between, /b/ is from 2 milliseconds to around 80, 90 milliseconds, it's still Percy, gets perceived as, /b/, but if the difference is more than 20 milliseconds, starts getting perceived as /p/. Well that's the point that I was trying to say. Okay, so that being said, let us move on to another very important effect, that is basically, there and it demonstrates, a few properties of speech perception, is this effect called the, 'Mc Gurk Effect'. Okay?

Refer Slide Time :( 27: 29)

- **The McGurk Effect:**

- **Acc. to the motor theory of speech perception, understanding speech requires you to figure out which gestures created a given acoustic signal.**
- **the system therefore uses any sort of information that could help identify gestures.**
- **while acoustic stimuli offer cues to what those gestures are, other perceptual systems could possibly help out, and if they can, motor theory says that the speech perception system will take advantage of them.**

I would strongly advise you to go to the YouTube and type, Mc Gurk effect and you'll see some of the videos, that are very interesting. People have made a lot of videos, you know, showing the Mc Gurk effect and the Mc Gurk effect is, basically same, something, that, I'll, I'll talk to you about it in a bit. But that's basically, when you hear something else and you see something else, your brain kind of, you know, makes something, completely else, out of it. So, we, we'll come to that. Now before moving that, let's just talk, a little bit about, the motor theory of perception. So it says that understanding a speech requires you to figure out, the gestures, that have created their acoustic signal, that's what I have been saying. Now the system wave, basically, what does it do? To you know, understand the gesture. We've not talked about

that, we have been saying that, Okay, you have to understand the gesture, in order to understand the sound. But how do you understand that, it shall. What are the sources of information, that will help you understand? Okay, this gesture would have created this sound. Okay? So, like acoustic stimuli offer cues or in themselves to what gestures might have created them, other perceptual systems could also possibly point out. Okay? So and if they can, say for example, the motor theory, say that a speech perception system, will take advantage, of information about possible gestures, from anywhere that it can. Alright?

Refer Slide Time :( 28: 55)

- Infact, two non - auditory perceptual systems - vision & touch - have been shown to affect speech perception.
- The most famous demonstration of *multi - modal perception* is the McGurk Effect (McGurk & MacDonald, 1976).
- The McGurk effect happens when people watch a video of a person talking, but the audio portion of the tape has been altered. for e.g. the video might show a person saying /ga/ but the audio signal is of a person saying /ba/. What people actually perceive is someone saying /da/.

So what happens, the two non auditory perceptual systems vision and touch, have also been, you know, shown, to affect speech perception, because they tell us, some information, about the gesture that might have gotten. Okay? So the famous demonstration of this is, the Mc Gurk effect, as I was saying. And Mc Gurk effect also phones into, you know, falls into this category of, multi modal perception. What is it exactly? It happens when you, you know watch a video of a person talking. But the person, but the audio portion of the tape, has been altered. So the person in the video, is probably saying, /g/, but the audio signal, that is being played, is of /b/. So the video that you listen is /g/, but the audio signal that you actually hear is /b/. So what happens is, there is different information coming from the vision, you are seeing a different kind of gesture and audition, you're hearing a different kind of sound. What the mind does is, it obviously gets confused, as to, Okay, I'm seeing this, but I'm listening something else. What does the mind do here? The mind basically combines these two sources of information and the people actually he, end up hearing /d/, that was not said, it was not in the audio file, it was not in the video file. But basically what the brain does is, it trying, tries to combine these two discriminate sources of information and it ends up with you, being able to say, /d/. Okay? That is the, 'McGurk Effect'. I'm sure I've not really given you a decent demonstration of this. I'm sure you should certainly go to You Tube and look at some of these videos. Okay.

Refer Slide Time :( 30: 35)

- If the visual information is removed (when the observing individual shuts his/her eyes), the auditory information is accurately perceived and the person hears /ba/
- The McGurk effect is incredibly robust: It happens even when people are fully warned that the auditory & visual information do not match; and it happens even if one tries to pay close attention to the auditory information and ignore the visual.
- The McGurk effect happens because our speech perception system combines visual and auditory information when perceiving speech, rather than relying on auditory information alone.

Yeah. So if the visual information is removed, the auditory information is perceived accurately. So if you close your eyes and you again listen to this, then you will be able to actually hear, 'gah', Okay? So the McGurk effect, is also incredibly robust, it happens even, when people are fully warned, that the auditory and visual information, will not match and if they are really trying to not hear the /da/, but many times people try and do this, but they always end up hearing 'duh'. Because the system is very robust, it kind of tries to combine these two sources of information. Okay? So why does the McGurk happen? The McGurk effect happens, because a speech perception system combines the visual information and the auditory information, when perceiving sound, rather than relying on, only the auditory information or only the visual information. So that is how, it came up in multimodal perception. Most of our speech comprehension is, actually rather multimodal, as long as, you are actually having, you know, you're at least having a visual of the person, who's speaking.

Refer Slide Time :( 31: 36)

- Of course the auditory information by itself is sufficient for perception to occur, but the McGurk effect shows that the visual information influences speech perception when that visual information is available.
- The McGurk effect is an example of multi-modal perception because two sensory modalities, hearing & vision, contribute to the subjective experience of the stimulus.

Of course, the auditory information by itself, is sufficient for perception to occur, but a mega McGurk effect shows, that the visual information, influences the speech perception. Okay? As soon as you close your eyes, you'll not really hear the confusion; you will hear the exact thing, that is being played in the audio file. But however, what really happens is that, the system is very keen on taking on information, from other sources. So it does, as long as your eyes are open. It will inadvertently combine the visual information with the auditory information, leading to the McGurk effect, if these two are not matching. This is an example of, multimodal perception. Because two sensory modality, is hearing and vision, contribute to the subjective experience of the stimulus.

Refer Slide Time :( 32: 21)

- Another way to create a variant of the McGurk effect is by combining haptic information with auditory information to change the way people perceive a spoken syllable (Fowler & Dekle, 1991).
- This kind of speech perception occurs outside the laboratory from time - to - time in a specialised mode called *tadoma*.
- Hellen Keller & other hearing & vision - impaired individuals have learned to speak by using their sense of touch to feel the articulatory information in speech.

Another way, another variant or another way that McGurk effect can happen is, by combining the haptic perception, with auditory information. So I said in the beginning, vision and touch, you can actually, also get clues to, what is being said, by the tactile modality, by the touch modality. So what happens is, this is something that is used, basically with the people, who you know, who were not, were not being able to see perfectly. This kind of speech perception is basically referred to as, Ted Ouma. Okay? And it happens in specialized laboratory conditions, some, from time to time. And the idea is, that basically when a person is speaking, you keep your hand, you keep your palm, in front of their mouth. So you kind of get the gesture that the person is going to say. So Helen Keller and other hearing in vision impaired individuals, have learned to speak, by using the sense of touch, to feel the articulatory information, in speech. Because you need to know, what gestures are being created, to create what. The hearing is alright, but the speech is not. So you kind of really need to get that. Probably in some cases, your hearing is not alright itself. But the tactile modality is a good clue to, what you know, gestures were needed, were needed to produce, this kind of sound.

Refer Slide Time :( 33: 45)



- Acc, to the motor theory, information about speech gestures should be useful regardless of the source, auditory or otherwise.
- That being the case, information about articulatory gestures that is gathered via the perceiver's sense of touch should affect speech perception.
  - to test this: Carol Fowler had experimental participants feel her lips while they listened to a recording of a female speaker speaking a variety of syllables.
  - Blindfolded and gloved, experimental participants heard the syllable /ga/ over a speaker while CF simultaneously mouthed the syllable /ba/.

So according to the motor theory, information of speech gestures, should be useful, regardless of the source, auditory or otherwise. That being the case, information about articulatory gestures, that is gathered by the other perceiver sense of touch, should affect speech perception. Okay? So initially for normal individuals, who have their hearing and their vision both normally functioning, vision is a very good source and vision is a very good information source, about the gestures. For people who are hearing or visually impaired, so does touch work, for them, in the same way that was the question. And to test this basically Carolyn, Carol Fowler, she had an experimental participants, feel her lips, while they listened to a recording, of a female speaker, speaking a variety of syllables. So the lips says 'da' and hearing says, sorry, the lip says, 'ba', the hearing says '/d/' and give what you hear is, a combination of these two, saying /d/, that's the McGurk effect, that really happens. So, blindfold and in gloved, experimental participants, heard the syllable, /ga/over a speaker, while Carol Fowler actually merged, /ba/ and what they do is, they hear, /da/. Okay? So the McGurk effect is, clearly there with them, as well Okay? As in the visual version of the McGurk effect what person is actually perceived is the combination of, the type, information from the tactile modality, versus, information from the, auditory modality. Okay? So they actually, again hear the syllable, /duh/.

Refer Slide Time :( 35: 17)

- the motor theory explains both versions of the McGurk effect, the visual one & the haptic one; as stemming from the same basic processes.
- The goal of the speech production system is not a spectral analysis of the auditory input; rather, it is figuring out what set of gestures created the auditory signals in the first place.
- Motor theory handles the visual & haptic effects on speech perception by arguing that both the modalities can contribute information that helps the perceiver figure out what gesture the speaker made.
- Under natural conditions, the visual, touch & auditory information will all line up perfectly, meaning that all secondary sources will be perfectly valid cues; in conditions as we saw that was not the case.

Now the motor theory explains versions, the visual and the tactile version, of the Mc Gurk effect, as stemming from the same basic process. They say the same thing is happening here, as well. The goal of the speech production system is, not a spectral analysis of the auditory input, it is trying to figure out, what sort of gestures created, that kind of sound in the first place. So what they do is, they combine what information, about gestures, you can come about, by vision or by you know, audition and or by touch. Both, everything is kind of, combined and they basically come up with the, output, on the basis of this combination of information. Motor theory handles the visual enough and haptic effects on speech perception, by arguing, that both modalities do contribute to the, information that helps the perceiver, figure out, what is being said. Under natural conditions, the visual touch and auditory information will all be available, not you really touch, because you know, we're not really touching people, when they're speaking. But say for example, but, all of them are available. If you, for example, decide to really check, all three information will be available and will basically be consistent and can be combined, should lead to a reliable, estimate of what the gestures were and so motor Theory switch perception will work perfectly well. Okay? However, the Mc Gurk effect is an artificially created effect, because the information from the visual and the auditory and the or the touch and the auditory modality's' were not consistent by design, that is why, you miss here something, that is why; you miss here, what is referred to as the, Mc Gurk effect. Okay? So this is all what I had to say, about the Motor Theory of Speech Perception. This was the fourth lecture, of the week. In tomorrow's lecture, I will talk to you about, some of the other Theories of Perception. Thank you.