

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

Module No. #01
Lecture No. #06
Speech perception – II

Namaste, friends. Welcome back, to this Lecture number-6, in the series on, course on Psychology of Language. What we have been doing this course is, we are looking at, the psychological aspect of a language, what it means, how it is developed, how it is perceived, and how a language is used, as a medium of communication. Before, we start this lecture, which is the sixth lecture, and progress on the concept of Speech perception, let us take a look back, at what we did, up till now, on what we have been doing in this course, up till now.

So, we started off this course, by looking at, the basic of language, the meaning of language, why it is needed. And, the best model, to do that, is to look at, Animal communication system. So, we started off by looking at, Animal communication systems. We looked at, the modes of Animal communication. And, we looked at, the reasons of why, animals communicate.

And thereon, we looked at, four different needs, why animals communicate, right from finding force, to finding friends, or for just finding food, or application, these are the, needs. So, we discussed that, as reasons why, animals communicate. And then, we looked at, certain characteristics of the, Animal communication system. We moved ahead, by looking at, the Human language, and basics of Human language.

And, what is the difference of Human language system, and the Animal language system. So, there we focused on, the fact that Human language, can very well transmit, a number of messages, various types of messages, and express different messages. But, that is not possible with, the Animal language system. And, that is one of the distinctions. But, there are several other distinctions, of the Human language system, from the Animal language system.

For example, being that, Human language systems, are governed by rules, and they are or have structured components. So, they have a syntax, way of expressing the language system. Also, they are productive in nature. And, Human language systems use, arbitrary symbols, so various

facts of the Human language system. Then, we looked at things like, Duality of Patterning, Recursive Structure, and several other characteristics of the Human language system.

We looked at, how the Human language system is arranged, in terms of the Phonemes, basic speech sounds, to the Morphemes, the Word, Sentence, Discourse, and so on, and so forth. So, how it is basically, built up, the phrase, the sentence, and so on, so forth. So, that is another thing, that we looked at. Further to it, we looked at, the evolution of language, how Human language system, actually evolved from animals. And, we looked at the, Continuity and Discontinuity Theories of Language Production, or how language developed.

And, both of them, have their own ideas of, how Human language evolved. And then, we looked at, certain proofs of the Proto language. Because, Human language system was believed to be, an upcoming of slow development from the prior humans, or the Monkeys, I would call them. So, we looked at evidences, that support this view that, it was a series of changes, that led to the present complex language, that we know of.

So, we verified the idea of Proto language, and we looked at evidences. For example, the evidence of a pig in which, is basically a language system, which has certain action words, and a handful of other words, which can express meaning between two language systems, where both the parties expressing the language, or talking to each other, expressing views with each other, cannot understand their, either's language.

So, Pidgin is basically an evidence for, the fact that, language develop in a progressive manner. So, that was what we did, initially. And, since we had laid the foundation of the language system, and what is language and the principles of language, and the basics of language, we moved on to looking at, the science of language. So, there we described, the famous N400 experiment.

And then, we described the scientific method, in detail, expressing what is a problem statement, how a hypothesis is formulated, and how this hypothesis leads to the generation of the or the hypothesis, how this hypothesis is generated, rather from a theory, and how that is that is tested. So, we looked at the Research Cycle, which is how, theory leads to hypothesis, which further leads to observations. And, these observations, lead to certain patterns. And, these patterns, verify the theory, may be verify or not verify the theory.

So, we looked at, that is a cycle of how, research is basically done. We looked at, experimental designs, which are used for conducting, experiments in language, in specific, and social sciences, in general. And so, we looked at the, within subject design, and the idea of between subject design. We also looked at, various component of the experimental process like, what are different kind of variables, the Independent variable, the Dependent variable. And, how they are manipulated. And, how they construct, the whole idea of doing research.

So, having described, the Independent variable, another idea, or another step was, describing the Dependent variable, which was, what do we measure, out of these experiments, that we do in language. And so, we looked at the idea of Latency, which is the reaction time, of how quickly, something is set upon, or how quickly, something is particular act is done. And, we also looked at Accuracy, if how correct a particular act is.

So, these two measures are the main measures, that is generally used in experimental treatments, or experimental problems in language. Further, we dedicated a section, on language and brain, and we looked at how, various regions of the brain are processing language. For example, the idea of Wernicke's area, the idea of Broca's area, and how these areas, have their say, in language. So, that is what we did, in the Second section, which is chapter numb Lecture number 3 and 4.

Now, in the previous lecture, that we did was on, looking at perception of sound. So, how does sound the how are sound produced, and what is speaking, basically. So, how the idea of sound is perceived. So, we looked at, certain parameters of sound, as a wave. So, we looked at the amplitude in frequency, which is the basic measure, or the basic parameters, of any sound wave.

Then, we looked at, the idea of the fundamental frequency, and the idea of overtones, what they are. Further to that, we looked at, how these sounds are Aperiodic and Periodic in nature. And, how this Aperiodic and Periodic sounds, are the sounds of vowels and consonants. Then, further to it, we looked at the structure of the ear, of how the ear perceives sound. So, we looked at, how the outer ear, the inner ear, and the middle ear, they combined with each other.

And, we looked in detail, idea of the cochlea, and the Basilar membrane, which perceives, or which notices changes in the sound wave, as they flow on the liquid on the Basilar membrane.

We looked at, the Tonotopic organization, of the Basilar membrane. And, how this Basilar membrane, which have hair cells, which pick up the sound base, how they are connected to, the primary auditory cortex, and the secondary auditory cortex.

And, how the organization of the Primary and Secondary auditory cortex, are similar, in way to the Basilar membrane, which is Tonotopic in organization. We then, looked at, how the speech stream looks like. So, when we speak something, what is it exactly, we try to look at that, and with the help of a Spectrograph, or the with the help of what a Spectrograph notices, we try to look at, what is the speech stream like, or what is speech like.

So, what is a Spectrograph. Spectrograph visualizes, the structure of the three streams. So, we looked at that. And thereupon, we tried to look at, what are consonants, and what are vowels. And, how these consonants and vowels, are distinguished, based on the Spectrograph. We also looked at, what the idea of Phonation, which is the vibration of the vocal fold. And, we also looked at the idea of, what is Prosody, which is the changes, subtle changes in the vocal cord, around the fundamental frequency.

(Refer Slide Time: 10:33)

Read My Lips

Multimodal perception

- Senses strongly interact to produce rich experience of world

McGurk effect

- Speech perception combines both auditory and visual information

```
graph TD; Hear[Hear b] --> Perceive[Perceive d]; See[See g] --> Perceive;
```

Now, we also looked at, certain other characteristics of the speech stream. For example, the idea of Formants or Sonorants. And, how these, Formants and Sonorants. So, Formants basically, are vowels. And, Sonorants are a kind of a vowel. And then, we looked at, Plosives and Fricatives, which are the representation of consonants, onto the speech stream. So, that is what, we were doing in, the last class.

Then, further to that, we looked at how, sound is perceived in categorical format. So basically, if you look at, speech stream or Spectrogram, you will look at a continuous, you will look at, you would not look at a continuous pattern. You look at patterns, which have higher amplitudes, and no amplitudes in between, and then again, higher amplitude. So, we looked at, how this this is interpreted, because speech is continuous in nature.

So, how does the Spectrogram, represent it. And so, the spa white space says, which you see in a Spectrogram, are actually consonants. And, these are called, stop consonant. So, that is what, it is. And so, speech actually, merges itself. So, then we looked at, how the speech is perceived, and we saw that, speech sound or speech is perceived, in terms of categorical perception, or in terms of categories. So, sound speech sounds are, forced into various categories. And, that is how, they are perceiving. That was what, we were looking at.

So, to demonstrate the speech is perceived, or categorical perception happens in a speech, a famous, Phoneme restoration effect was demonstrated. That is where, we ended the section. And, we saw, how Warren & Warren, did an experiment, and showed, how this Phoneme restoration effect is, basically demonstrated.

To end the section, in the last class, we described something called, McGurk effect, which basically says that, Speech perception, is not only through the auditory medium, but other mediums also like, visual medium is also used. And, we demonstrated the idea of, how somebody speaking, and somebody lip, speaking a particular word, how, when they are shown to the same person, how the person misreads.

So basically, Multimodal approach to perception saying that, auditory feedback is not the, or auditory stimulation is not the, only way of Speech perception, a multimodal technique is used by the human brain, to perceive speech. Now, what we are going to do in today's class, is look at how, speech is developed, in the Neonates. In small children, how speech is developed. And, towards the end of the section, we will look at several theories of Speech perception. So, that is what the aim is, presently in this lecture.

(Refer Slide Time: 13:24)

**Development of Speech Perception:
Language Learning in the Womb**

During third trimester, fetus can hear

- Mother's voice
- Environmental sounds

Change in fetal heartbeat indicates

- Fetus can discriminate stimuli

At weeks 35-37, fetuses can recognize

- Mother's voice
- Mother's language (spoken by another woman)
- Familiar nursery rhyme

So, development of Speech perception, and language learning, in the womb. So, how does it is it starts in the smaller children. So, a language learning, it generally begins, in the womb during the third trimester. So, by the third trimester, the Neonate, or the baby inside the Fetus, in inside the Mother's womb, is able to hear speech sounds, or is able to perceive speech sounds, which extends from week number 28, until birth, that is, week number 38 or later.

So, from the third trimester onwards, the baby is able to, actually hear speech sounds, of the Mother, or other people around, the room. So, during the as you can see, there is the same thing, that I have here. So, during the third trimester, Fetuses can hear, Mother's voice, and environmental sounds. And so, it tries on it. It is also able to distinguish, this Mother's voice, and environmental sound.

Now, a Fetus, at the third trimester, can hear and respond to sounds, in the environment, particularly the Mother's voice. So, that that is what, it can do. It is a number of experiments, which are there, which were done to prove, that the baby is able to hear, the Mother's sound, and other environmental noises. And, these can be done with, either the Head movement reflex, or the Nipple sucking reflex.

So, there are several experimental techniques, which are used for, showing that, this kind of response, or the Fetus is actually able to, hear the sounds. Now, one way to demonstrate, that the Fetus is able to hear the sound, is the change in Fetus heart rate. So, the Fetuses can

discriminate stimuli. So, when the Fetus hears, actually a stimuli, vocal stimuli, or speech stimuli, which is different from, or normal stimuli, which is different from, what he has been hearing, the heart rate increases.

But, when he has a stimuli, which is similar to, what he has been hearing, he decreases the heart rate. And so, that particular response is what, is the demonstration of the fact that, the Fetuses can actually hear the sound. Now, at birth, the new-born can already distinguish, its Mother's voice from other woman's voice. And its Mother's language from other language. Now, at week number 35 and 37, Fetuses can recognize Mother's voice, Mother's language, spoken by another woman, and primary nursery rhyme.

So, there were experiments done, in which the Mother's fundamental frequency, or the fundamental pitch was removed. It was filtered out. And, only the basic Prosody was there, or the basic Prosody of the Mother's tone was there, in the voice. So, higher frequencies were split out. Even then, the child was able to respond to, or the child was able to, um hear the Mother's voice, or distinguish the Mother's voice, in terms of the heart rate.

Also, if another woman, which is not the Mother, if she speaks in in in in a manner, similar to the Mother, the child actually showed a change in its basic heart rate. Which basically means that, she was able to able to realize the, or able to respond to the Mother's language. You know, in one experiment, Dr. Seuss Fairy-tale was told, was recited, to these 35, 37 years Fetuses, and so they were able to respond to these, these even if a different woman, other than the Mother, was playing this particular, or was reciting this particular nursery rhymes.

(Refer Slide Time: 17:23)

Born Yesterday

High-amplitude sucking technique ✓

- Measures frequency of infant's sucking on non-nutritive nipple
- Changes in frequency indicate discrimination of stimuli
- Used with newborns

✓ Newborns prefer

- Mother's voice
- Mother's language (spoken by another woman)
- Familiar nursery rhyme (heard in womb)

Clearly, newborns remember what they heard in the womb

So, another interesting thing, which used to measure, the fact that, babies can hear speech sounds, was the high amplitude sucking technique. So, in this technique, what was happen is, a non-nutritive a nipple was given to the baby, and he would suck on, the way he sucks on to the nipple, on hearing different sounds, familiar and non-familiar sound, demonstrated, how or whether the baby can, distinguish between the Mother's voice, or familiar voice versus un non-familiar voice.

So, measures, frequency of infant sucking, on non-nutritive nipple, and changes in the frequency indicate, discrimination of stimulus. And, this was used on, new-born. So, this is another interesting thing, which was there. And so, this was used to, demonstrate the fact that, babies, smaller children, were able to hear, this this sound, or distinction between, familiar and non-familiar sounds. Now, new-borns are known to prefer, Mother's voice. And so this this was demonstrated by, the high amplitude sucking technique.

Also, Mother's language, which is spoken by another woman, was also preferred by, most new-borns. So, as related to the Fetus, which actually demonstrated a change in heart rate, similar to them, new-borns also demonstrated, through the high amplitude sucking technique, they also demonstrated that, they preferred Mother's language, if spoken by another woman, more than any other any other text, which other woman was speaking.

Also, family nursery rhymes, which they hear in the womb, when it was played back, after they were born, they preferred to that. And, that was demonstrated through, this high amplitude sucking technique. Now, clearly new-borns, remember what they heard, in the womb. And so,

this this basically demonstrate that, the baby, the Fetus, were actually able to hear, these sounds, or these the changes in sounds, are were able to, discriminate between these sounds.

(Refer Slide Time: 19:39)

Baby Talk

Infant-directed speech

- Manner of speaking to infants, attracts their attention, helps them learn language
- Higher fundamental frequency, greater range of pitch
- Exaggerated stress contrasts and vowel length
- Nearly universal phenomenon
- Also called *motherese* or *caregiver speech*

Prosodic bootstrapping

- Infants use intonation and stress patterns to infer phrase and word boundaries

Now, how do, the Mother pass on, this language ability into, the child. And so, that that that happens to something called, Baby Talk. So, what we have demonstrated, what we have established, up till now, is the fact that, the babies, or Fetuses, or new-borns, they can distinguish, or they can hear voice, of their Mother. And they can make distinctions between, the Mother's voice, or another familiar voices, with different voices. And, that we have shown, that not only at the level of the Fetus, but also when they are born, they show these distinguishing.

So, at the Fetus level, the change in heart rate is a measurement, of the likingness to a certain sound, or the distinguishing between, a certain familiar and non-familiar voices. When they are born, the new-borns, as old as one month old, by using the sucking technique, the high amplitude sucking technique, they also can demonstrate that, they can differentiate between, various changes into the, or they can discriminate between, familiar and non-familiar voices.

Now, the question is, how Mothers are able to teach, their children to speak. So, how do they perceive speech, or how do the Speech perception is, taught by the Mothers, to the small children. And, that happens by something called, Infant-directed speech. So, as you know, most Infant-directed speech are something called, Motherese. So, these are the manner in which, the child and the Mother, they speak to each other, or most Caregivers speak to the children.

So, they are basically, small utterances, with high pitch utterances, and so these are what is called, Motherese. So infant directed speech, is spoken with a higher fundamental frequency, a broader pitch range, and exaggerated information, and stress pattern. And so, this kind of directed speech, is the first demonstration, or the first way in which, the Mother interacts with the child. So, the manner of speaking to infants, attracts their attention, and help them to learn the language.

So, basically, language learning starts right when, the child is born. And, this starts with the Infant-directed speech with, either the Mothers, or the Caregivers, give to the children. And so, what are they? These are, speech with high fundamental frequency, in a broader pitch range. And so, they have exaggerated information with, different stress patterns. So, higher fundamental frequency, then greater range of pitch, also exaggerated stress contrast, and vowel length.

So, for children, we generally use, the stress on any word is exaggerated, and the vowel length also. When speaking of vowel, the way we speak to children, or small children, or just-born in infants, the way we see this, the vowel length is also increase. So, this is done, so that, the child is able to perceive the speech, or perceive the speech changes, the changes, that is happening in the speech.

And so, that that is the first learning, or that is the first technique of speaking, or making the children, learn the speech. So, nearly universal phenomena. And so, this way of speaking, with exaggerated stress contrasts, in vowel length, by the Mother to the children, is universal phenomena. So, all across the world, it is the same way, the Mother speak to the children. Now, right after birth itself, the child spends a lot of time, with the Mother.

And, there is a lot of eye contact, and a lot of speaking, from both the side. Although, the child may not be able to speak, but he expresses himself, in terms of boos and calls and, that is responded by, speech by the Mother. And, that is the first way, or the first technique, or the first step, in language learning, by the child. So, the child is able to, perceive those changes, perceive changes in the in the Phonation, or receive the Prosodic changes in the Mother's voice.

And, that is how, he starts learning, how to speak, or what speech is made up of, or how language is. One interesting map, when a child is born, he kept listening it, at the time of birth,

he can distinguish, probably all, between all languages. So, give him any language, and he will be able to discriminate between, two patterns of speech. But, as he grows, by the first month itself, he loses this capacity of understanding, or pointing out changes in, patterns of speech.

And so, that we will discuss later, in in the Speech Perception Theory, or the Theory of Speech perception. So basically, what we are trying to, I am trying to tell you here, is that the manner of speaking of the Mother, it has a lot of role, in in developing the language system. And so, what it is, how does it start. It starts by the Mother, basically producing high fundamental frequency sounds.

And, these high fundamental frequency sounds, are what is perceived by the, children. So, basically, this is called the, Motherese, or Caregiver speech. So, this pattern of speaking, of using elaborated syllables, and longer vowels, and this way of speaking is, Motherese, or Caregivers speech. Now, these Motherese and Caregiver speech, is the fundamental step, in teaching the child, the new-born, the basics of any language.

And so, there is there is a theory, which is called a Prosodic bootstrapping theory, which says that, this particular Motherese, or Caregiver speech, the way it is delivered to the children, or the way it is used in the children, it actually helps the children, in identifying speech, or in identifying, how a particular language is spoken. So, the Prosodic bootstrapping hypothesis, it proposes that, infants use intonation and stress patterns, to infer phrase and word boundaries.

So, how does the infant know, when a word has ended, or what is a word, and what is a phrase. And, the way, it does is, by using something called, Prosodic bootstrapping. So, he looks at intonations, which is how the Mother's voice is changing, and the stress that the Mother is putting, in in the Motherese, and based on these intonations and stresses, the child is able to demarcate, the word boundaries, and face boundaries, in the speech.

And so, that is the first step in learning, to speak the particular language. Now, there are there are a number of a number of evidences, which basically support the notion of Prosodic boot strapping. So, whether Prosodic boot strapping happens or not, the idea of Prosodic boot strapping, whether it happens or not, there is a lot of evidences, or there are, at least some evidence is out there, which suggests that, this Prosodic boot strapping, actually happens.

So, for example, the infants are sensitive to in indicators of face boundaries and Motherese. So, the one of the evidence is that, the infants are actually sensitive, or show sensitiveness, to indicators of face boundaries, and in the Motherese. So, this this basically sensitivity to face boundaries, suggests that, infants are able to bootstrap, the Prosodic speech, and prosodies in the speech, and so, are able to distinguish between, word boundaries. Also, there was a yappy study, which was done on, 5 months old German infants.

And, what was found out that, they were sensitive to, Acoustic hues, and face boundary, speech changes and vowel duration. And, according to this speed changes and vowel duration, the German children, 5 months old, they were able to understand, the spoken speech, or they were able to at least use, cues from the Acoustic Motherese, that that that was used. And so, that made them understand, what is the boundary of a board, where a word is there, and where a phrase has ended.

And, how this actual speech is composed of, in what pattern, it is arranged, and where should the brakes be. So basically, at the point of time, when they do not know, what a word is, or what a phrase is, the certain cues from the way, the Mother speaks to the child, these cues are utilized by the child, to understand, what boundary is. And so, this is this is a basically a very interesting fact, that children, as small as 5 months old, are able to distinguish, these kind of patterns, or this kind of syntax, or rules, into the spoken stream.

(Refer Slide Time: 28:51)

I've Got Rhythm

Metrical segmentation strategy

- Rule-of-thumb for segmenting speech stream
- Assume words begin on stressed syllable (for English)
- Used by infants and adults

Conditioned head turn technique

- Infant trained to turn head when it detects change in stimulus

Another interesting thing, is about the English language. So, the English language, has a Metrical Segment Strategy, which basically means that, most English languages, have stress on

the first syllable, and not on the second syllable. Unlike, some other languages out there, like French, which has the stress, generally it is a stress on the second syllable. And so, Metrical Segment Strategy, is another strategy, which is used, both by the infants and adults, and not by the adult, but the infants, to understand the speech stream, or to understand the boundaries.

So, according to the Metrical Segmentation Strategy, both the infants and adults, they tend to segment the speech stream, at the onset of stress syllables, and at least in the English language. So, they know that, most words in the English language, starts by a stress, on the first syllable. And so, that is how, they are able to segment. So, Metrical Segmentation Strategy says that, these are rules of thumb, for segmenting speech streams.

So, if there is a speech stream, with somebody, if something is being stressed, is being spoken, how this is broken down into his basic properties, or into basic segments, is done by something called the, Metrical Segmentation Strategy. They assume that, word begins on stressed syllables. For example, English. Because, French words, they can actually have a stress on the, second syllable also.

Also, used by infants and adults, to segment this speech stream. Now, it is believed that, English has a characteristic rhythmic pattern, of alternating stress and unstressed syllables. Now, the stressed syllables, have long or complex vowel sounds, while the unstressed syllables in English, have reduced or shortened vowels. So basically, this idea, that English has a rhythmic pattern of stress and unstressed syllables.

And, the idea that the stressed syllables, are long or complex words, complex vowel sounds, while unstressed syllables have, reduced shortened vowels. This is another interesting thing, are used by the children, or the Neonate, the small children, or new-borns, to understand the speech segmentation, how to divide the speech, which is there. Now, understand that, the child when he is born, he does not have the idea of, the Phonemes, Morphemes, and that kind of a thing.

So basically, what he does is, he picks up these strings, or picks up these ideas, or generates these ideas, from the way, the Mother is talking to the children. And so, this kind of idea, or this kind of probability is that, these are the rules, which are used by the smaller children, or the children who are 6 to 8 months old, or maybe around 6 months old, so smaller children, to use, understand the speech stream.

So, Morgan, 1996, he did an experiment to find out, whether this Metrical Segmentation Strategy, is true or not. And so, he presented two syllable sequences, with stress on, either the first and the second syllable. So, the idea was that, whether smaller children, or so smaller children, or older children, they can distinguish between, whether children can actually distinguish between, this Metrical Segmentation Strategy, or can understand, the difference between, first stressed syllable, and the second stressed syllable.

So, what Morgan did was, he produced words, or he presented two syllable sequences, with stress, either on the first word, on the first syllable, or the second syllable, to 6 and 9 months old infants. And, they found out, the 6 months' infant, they did not respond to, the differences between, the first stressed syllable, and second stressed syllable. Whereas, the 9 months old, were able to distinguish between the, first stressed syllable, and second stressed syllable.

So, when the second stressed syllable was produced, they did not give this head response, or they did not respond to it, or they were not able to distinguish between it. And, that was the demonstration of Metrical Segmentation Strategy, as used by infants and adults, to segment the speech system. Conditioned head-turning technique. Now, infants trained to turn head, when he detects a changes in stimulus.

So, another technique that is used, to since children are small, so how would they show responses, when they are able to distinguishes change, in speech stream. And so, one another technique, which is there is, Conditioned head turning technique. So, till the point of time, that they cannot do this head turn thing, the high amplitude sucking technique is used. But then, if they can turn their head, a conditioned head turning technique is used, to demonstrate that, the child is able to distinguish between, certain aspects of the speech streams

(Refer Slide Time: 33:49)

I've Got Rhythm

- At 9 months, infants prefer strong-weak stress pattern over weak-strong
- At 6 months, no preference

1st 2nd
Strong weak

1st 2nd
weak strong

Toddlers frequently impose strong-weak stress pattern

- Spaghetti → sketti
- Amanda → Manda

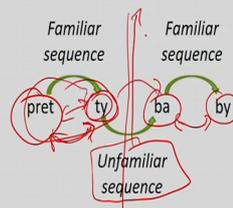
So, as I said, now in modern area, Morgan's experiment, 9 months' infants prefer, strong weak stress pattern, over weak strong stress pattern. So, when it was first syllable stressed, and second syllable non-stressed, they preferred this, in in in relation to, when they had the first syllable non-stressed, and the second syllable stressed. Also, at 6 months, the preferences were not there. So, toddlers frequently imposed, strong weak stress pattern for, Spaghetti to this sketti, Manda. So, that kind of, variation is there.

(Refer Slide Time: 34:25)

What's Coming Next

Transitional probability

- Likelihood a particular event will occur next given the current event
- At 8 months, infants can use TP to segment speech stream (as tested with conditioned head turn)



Pretty baby
ba by pret ty
ba ty pret ty

Another interesting thing, that is used by infants, to understand the speech pattern, or to decipher the speech pattern of the Mothers, or other Caregivers, is something called, the idea of Transitional probability. So, what is Transitional probability. Transitional probability, is the probability, that given a particular sequence, what is the probability that, the next sequence will follow it, or given a particular stimuli, what is the probability, that related stimuli will follow.

So, likelihood of a particular event will occur, next given the current event. So, Transitional probability, they provide a fairly reliable cue, to what boundaries, that even young infants are sensitive to. So basically, what definition is. The definition says that, the likelihood that a particular event will occur, given the fact, the event will occur, next given the current event is already occurred, is what is called, Transitional probability.

And so, Transitional probability, what it provides is that, is a fairly reliable cue, to what boundaries, that even young infants are sensitive to. So basically, these Transitional Probability, or the probability of the fact that, a certain syllable will repeat, after some other syllable, or a certain pattern will repeat, after some other pattern, is used by infants, or is again used by infants to, segment the speech stream.

Now, 8 months' infant can use, the Transitional probability, to segment speech stream, as tested with condition, head turn response. So, 8 months old infants were able to use, this Transitional probability mechanism, to demonstrate, to segment the speech stream. For example, if Pretty Baby is something, that the child is able to, or child hears a lot of time. So, if this pret is spoken, then there is a very high chance that, ty or t will be repeated.

And, so the child, this pret is there, the child always responds to, or waits for, or shows this condition had response to show that, ty will follow. And so, pretty is the word. But, after pretty, you know the fact that, baby will be repeated, is very less. And, so the child is, does not show, any response to that. Because, they do hear, so many other words, related to it. So, it is not that, pretty is something, that they always hear.

Because if the syllable practice is produced, they always know, or they turn their head to show that, ty, which is, pretty will be, what will be spoken next. But, the fact that, pretty will be followed by baby is, the Transitional probability is very less. And so, the child does not show this condition head response, or condition head responds to it. But, the moment, because it could be, pretty dog, pretty eagle, pretty egg, and so on and so forth. So again, any anything could be there.

But, as soon as, ba is spoken, the child shows a conditioned head reflex, to show that, Baby will be followed. And so, when the ba syllable is there, the b syllable will repeat it. That is shown,

by this Transitional probability. So, the fact that, if Pret is spoken, and ty will follow it, this is called the transaction the probability the Condition Transitional probability. And, if ba is spoken, then by will follow, this is what is, Transitional probability. But, the fact that, if pretty is spoken, then Baby follows it, is an unfamiliar sequence. And so, in this case, the child does not respond, head-turning response.

(Refer Slide Time: 38:11)

Narrowing Down

Perceptual narrowing

- Transition from broad to narrow perceptual categories
- At birth, infants perceive all possible speech categories
- By 1 year, they only perceive categories in their language

Distributional learning

- Track frequency and location of various sounds in speech stream
- Aids infants in organizing perceptual categories of their language

Handwritten notes in red:
 pretty leads to pret and ty leads to by
 pret and ty are grouped together with a bracket

So, on some other techniques like, statistical learning techniques are used by the child, to segment the speech stream. What we have been doing all along is, we are trying to tell you, or how the child makes these identification of the speech stream, or segments the speech stream. When he is hearing a speech, how does he know, where to stop, and where to start, or what are the boundaries of the speech. So, where is the word end, and where is the phrase boundaries, and that kind of thing, how does a child understand that.

That is what, we are trying to, or that is what, I am trying to tell you. So, in infants, the Transitional probability from pret to ty, is very high. Likewise, from ba to by, is again very high. However, the Transitional probability from ty to ba is very low. So, from the syllable Ty, the fact that, Ba will follow, because it is pretty, right, and Baby. So, pret and ty, and ba and by. So, the Transitional probability between ta ty and ba is very low.

Whereas, the Transitional probability between, pret and ty, and ba and by, is very, very high. So, infants are born, with the ability, to discriminate more speech sounds, and through a process called, Perceptual narrowing and acquiring, the phonetic categories of the language, that they are learning, during the first year of life. So, another thing, that the infants use, in segmenting

the speech stream, is something called, Perceptual narrowing. So, what is this Perceptual narrowing, is a transition from broad to narrow perceptual categories.

So, what is it. They are born with the ability to discriminate, much speech sounds. And, so a process called Perceptual narrowing, they acquire the phonetic categories of the language, they are learning, during the first year of life. So, they use this Perceptual narrowing, for categorical perception, or putting speech sounds, into certain categories. At birth, infant perceive, all possible speech categories. Now, by 1 year, they perceive categories in their language.

And, as I said, as I have been telling before, when a child is born, he is able to um categorize differences, in any language of the world. But, as he grows, or as he learns, as he talks to his Mother more and more, he does this perceptual narrowing, or he attains this perceptual narrowing, and then, he can only distinguish the speech sounds, or speech stream, or differentiate between speech streams, on the in his, Mother tongue.

So, experiments were done with, English-speaking children, or Hindi speech, or Hindi h speech stream. And, children were um very small, or just-born children, were able to discriminate between, two sequences. But, when this this child became, 1 month old, he was not able to discriminate between, changes in speech stream, in Hindi. But, they were able to do that, in this particular process of, how did they do it, is basically called, perceptual narrowing.

They also use something called, the distributed learning technique. So, the perceptual narrowing of phonetic categories is aided by a process, which is called, um Distributional learning. So, what is this Distributional learning. It is a way of tracking, or the frequency, and localization of various sounds, in the speech stream. So, the Distributional learning, as a process of, tracking changes in the speech stream, and narrowing down on to, or localizing various sound, in the speech stream.

So, Distributional distribution learning, is tracking frequencies, and localization of various speech sounds, in the speech stream. And, aids infants in organizing perceptual categories, of their language. So basically, this both perceptual narrowing, and distributional learning, are two techniques, which the infant uses, to understand the speech stream, which is being spoken to him, and to distinguish in various characteristics of the speech, and understand the speech.

And, basically perceive the speech in to categorical perception of the speech. So, this is what happens, or how the child learns in to, or learns this speech stream, or basically, the speech, when it is aided a lot by Motherese, of how a Mother speak to the children.

(Refer Slide Time: 43:33)

Theories of Speech: Lack of Invariance

Lack of invariance

- No reliable relationship between phoneme and acoustic signal

Problem for all theories of speech perception

- How do we extract original speech sounds from speech stream?

Three approaches

- Motor theory ✓
- General Auditory Framework ✓
- Direct Realism ✓

Now, towards the end of the section, we looked at several theories of speech, and how these theories argument each other, and complement and dis-complement each other. So basically, there are several theories of speech, which is out there. And so, these theories of speech, is what we are going to learn, next. So, we consciously believe, that speech stream is a sequence of speech sounds, grouped into words and phrases.

So, believe that, when somebody is speaking, it is it is basically a continuous flow of words and phrases. But, in reality, we know that, this is only an illusion .so, speech is not continuous at all. This is there is no sorry the speech is not non-continuous at all; it is a continuous thing. The observation that, there are no reliable relationship, between Phoneme and the Acoustic signal, is known as the, Lack of invariance.

So, the way, the Phoneme is being produced, and the Acoustic signal, which is generated out of that Phoneme, there is no relation between the two. And, this is basically, what is called the, Lack of invariance. So, no reliable relationship between, Phonemes and the Acoustic signal. So, what signal, what speech signal is generated, and what Phoneme is being used, there is no relation between these two. And, this is called the, Lack of invariance.

Now, the problem, that theories of perception, need to explain is, how we are able to extract, the original speech sound, from the speech stream. Now, if the Acoustic signal, which is being produced, and the Phoneme, which is there, if there is no relation between them, the theories of perception need to express, how we perceive speech stream, or how we are able to extract, these things from the speech stream.

And so, problem for all theories of Speech perception is that, how do we extract original speech sound from the, speech stream, or the speech signal. So, there are three theories, which have been proposed. The Motor Theory, the General Auditory Framework, and the Direct Realism. And so, we will go into these, one by one.

(Refer Slide Time: 44:33)

Motor Theory

Nativism

- View that behavior is mainly shaped by natural selection, encoded in genes

Language acquisition device (Chomsky, 1959)

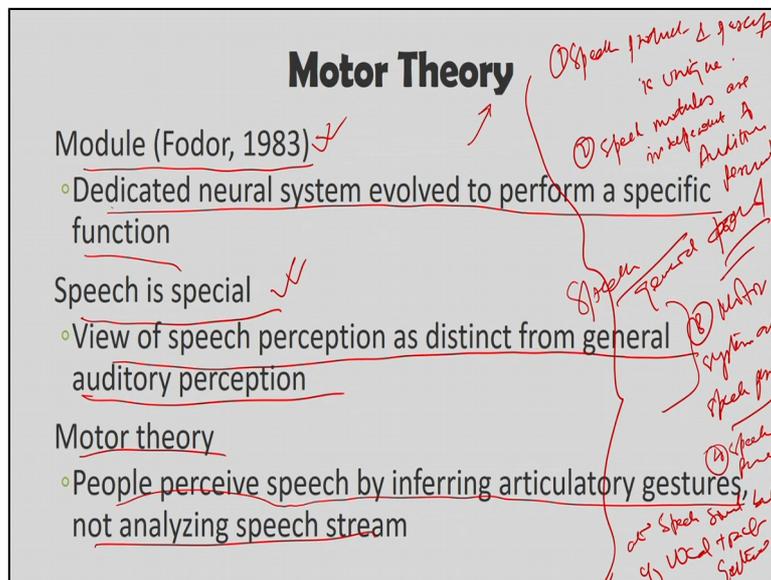
- Specialized processing units in the brain
- Guide rapid development of language in infants

Let us look at the, Motor Theory. So, Motor Theory, it uses a nativistic approach, and views that behavior is mainly shaped by, natural selection, encoded in genes. So, it believes that, speech starts with genes. And so, it has a nativistic view. It sits in the gene, somewhere. And so, from there, the Speech perception is, genetic in nature. And so, it is basically, it is supported by, Chomsky's Ideas of Language acquisition device. And Chomsky, he believes that, there are specialized processing units in the brain, they guide rapid development of the language in infants.

So, he believes that, it is the, every child comes with, or every person comes with, the Language acquisition device. And, in the childhood itself, there are specialized in use out there, which basically um let the infant learn the speech. So, Alvin Liberman and his colleagues, at the

Haskins Laboratory, built a pattern playback machine, that the use of which, led to the development of the, Motor Theory of Speech perception.

(Refer Slide Time: 45:30)



So, it was also the idea of Module, by Fodor, in 1983, which led to the increment in the Motor Theory. So, dedicated neural systems, evolved to perform, a specific function. So, generally speaking, it was Alvin Liberman and his colleagues' idea of, using the playback machine, to develop this Motor Theory. And, what is the basic idea of this Motor Theory. The idea of Motor Theory is that, the way we speak is, it is trapped by the changes in the vocal cord. So, Motor Theory, what it proposes is, it proposes that, people perceive speech.

How do you perceive speech? We perceive speech, by inferring the movements of the vocal tract, that produced this sound. So, by looking at the vocal tract, which is producing this sound, we perceive speech. So, instead of analysing the speech stream, into its Phonemes, so instead of looking at speech, in terms of its phonemes, the perception of speech actually progresses through, or it manifests through, it happens through, tracking the way, the vocal tract is responding, when somebody is producing speech.

In other words, we comprehend speech, by imagining, how we would make the sounds, we are hearing. So, Speech perception happens by, another person, or somebody, who is hearing somebody speak. How we perceive speech, is by imagining, how he would have produced, that speech. And so, by that manner, Speech perception happens. So, influenced by the Theories of Chomsky, Fodor, Motor Theory argued that, the speech is special in in in manner.

So, that is the basic of the Motor Theory. But, as I said before, Chomsky's idea was, there is something called the, Language acquisition device. And so, this Language acquisition device, is the way, or the reason, how the smaller children, or infants, develop speech. Also, Fodor, 1983, he proposed the idea of a Module, and he says that, the brain has dedicated neural systems, for learning speech. And, that is how, speech is basically perceived.

And so, based on, both their ideas, it was believed that, speech is a special quality of human beings. And, view of Speech perception as, distinct from, general auditory perception. They believe that, speech is different from, general auditory perception. So, viewing of other auditory sounds, is different from, when you hear somebody speak. Now, what it says is that, speech is processed, by innate dedicated modules, that are separated from, general auditory perception.

Now speech is special, was based on, several reasons. Now, Motor Theory, the people perceive speech, by inferring articulated gestures of, not analysing the speech stream. So basically, the idea, that speech is special, it is dedicated, or it rests on three basic reasons. First, speech, perception, and production, was viewed as a, unique human ability. So, three reasons, speech, production, and perception, is unique to humans, first.

Second, Speech perception modules work, independent of auditory perception. So, speech modules, they are independent of auditory perception system. And, the third reason is that, Speech perception was proposed by the, motor system. It was believed that, Speech perception, motor system aids, Speech perception, or Speech perception, happens through the motor systems.

And so, these three reasons, were quoted, rather for the fact that, also D, another reason, that was given is that, the objects of Speech perception, was not speech sound themselves, but rather the intended vocal tract gestures. So, Speech perception was not looking at speech sounds, but on vocal tract gestures. So, how does the vocal tract, actually move, that was the reason of, or that was the way, the Speech perception should be done.

And so, this is the this is the framework on which, the Motor Theory was based. Now, ample research, has been shown, that non-human animals, they perceive speech sounds, in much the same way, as humans do, forcing, motor theories to forsake the position that, speech is special. And so, this idea, that speech is special, is basically nullified, or it is being taken back, by the

by these evidences that, non-humans would perceive speech sound, in much the same way as, humans do.

(Refer Slide Time: 50:11)

General Auditory Framework

General auditory framework

- Speech perception operates by same mechanisms that have evolved for perceiving environmental sounds
- ② Human infants can discriminate speech sounds before they can speak
- ③ Nonhuman animals can distinguish speech sounds
 - Chinchillas categorically perceive *da* and *ta* (Kuhl & Miller, 1978)
 - Japanese quail perceive *d* in *dah*, *dee*, *doo* as same despite coarticulation effects (Kluender et al., 1987)

So, that is the basic of the, Motor Theory. Now, to distinguish the Motor Theory, or to basically cover up, the fact of Motor Theory, that speech is special, another theory was proposed, which is called the, General Auditory Framework. The General Auditory Framework, what does it say. Speech perception operates in the same mechanism, that have evolved for the perception of environmental sound.

So, there are several, what is the proposal? The proposal is that, Speech perception operates by the same mechanisms, that have evolved for perceiving environmental sounds, which are not special. Speech perception is not special from, some other ways of perception. Now, there are several evidences, which go against the Motor Theory, which has been provided by the General Auditory Framework, which goes against the Motor Theory.

The first evidence, against Motor Theory is that, human infant speech processing. How humans infant process speech, that has been quoted, as the basic reason of how the, or General Auditory Framework, differs from the Motor Theory. Now, human infants, are born with the ability to discriminate, nearly, all possible speech sounds, yet they do not begin speaking, a year later. So, as has been said that, they are able to discriminate, all speech sound, but they do not speak by first year.

And so, they basically, that is the that is the reason, that these General Auditory Framework people, they propose against the Motor Theory. But then, General Auditory Framework, believes that infants have this ability, because Human languages, only make sound distinctions, that the that the auditory system is already sensitive to. So, the way, the human auditory system perceives, is the same way, is used, for, perception of speech. And so, infants also use, the same system.

Now, the motor theories also cite human infant speech processing, as support for their own theory, especially the claim that, Speech perception involves, innate modules. Motor Theory, they believe that, Speech perception production system, is hardwired to the brain, and guide the development of language specific skills, in both perception and production, during the first year of life.

So, the same idea, or the same evidence, that is used by the General Auditory Framework people, that infant speech processing, that they believe that, infant speech processing supports their theory, the same thing is used by the Motor Theories. And, what the Motor Theory says that, there are innate modules, in the Speech perception, in the in the infants, which held them in perceiving speech.

Whereas, evidence that the general framework people are using is that, although infants are able to, perceive changes in speech, or perceive various differences in speech, by before one year of age. Only by one year, they start speaking. So, before one year, they can they can do, they can understand, this changes, and they respond to it also, when they see a change. But, by using Perceptual narrowing, they start speaking by first year of life.

And so, that is that is one evidence. So, human infants can discriminate speech sound, before they can speak. Now, the second evidence, which is which is used by the General Auditory Framework, against the Motor Theory, comes from the observation, or Speech perception, in non-humans, animals. For example, the Chinchillas. And so, non-human animals, they can distinguish, speech sounds.

So, that is another second evidence. So, this is my first evidence, and this is my second evidence, of General Auditory Framework against the, Motor Theory. So, Chinchillas categorically perceive speech, DA and TA. So, they can do this, categorical perception also.

Japanese Quail, perceive D in Dah, Dee, Doo, as the same despite Coarticulation effects. And so, this this is another thing, which is there.

(Refer Slide Time: 53:48)

General Auditory Framework

Speech perception is so complex that we use any available cue

Fuzzy-logical model of perception

- Perceptual decisions made by matching relative goodness of sensory inputs to prototypes in memory

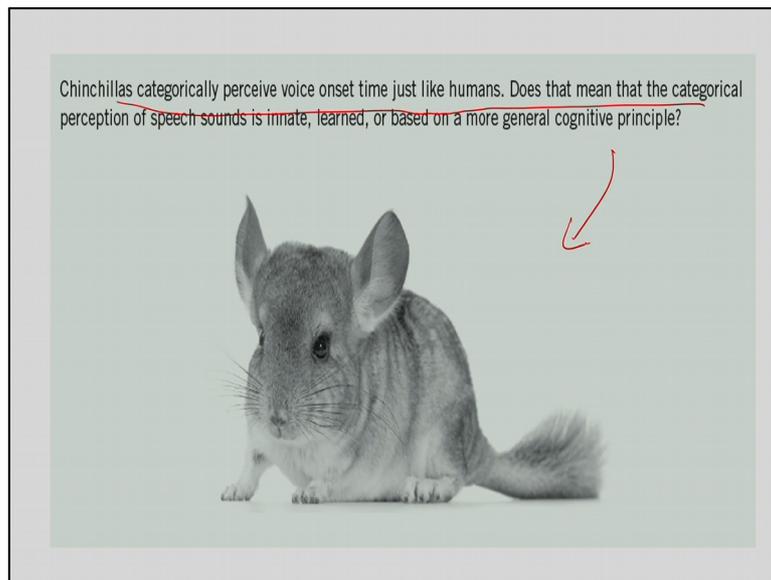
Now, Speech perception is so complex, that we use, any available cues. Now, researchers working in the General Auditory Framework, takes the proposition, the speech is such an amplitude task, that human takes, advantage of whatever information is available, how imperfect, to tackle this problem. So, Speech perception is not innate. It is not processed by a dedicated system.

What it says is, the Speech perception, the General Auditory Framework says that, it happens, Speech perception happens, in the same way, as the auditory perception. But then, the humans take advantage of all available, imperfect cues, from the environment, in the speech on itself, to distinguish, or tackle the problem of Speech perception. Now, one example of this approach, is something called the, the Fuzzy-logical model of perception. And so, what is this model of perception.

It proposes that, we arrive at perceptual decisions, by matching the relative goodness of various sensory inputs, against their values of particular prototype, stored in memory. So, when we are doing perception, what we do is, we look at various examples of the object in question, and we compared them to Prototypes, and then we do a matching, how good a match is, a particular stimuli, with the perceptual Prototype, which is stored.

And, that gives an idea, of how perception takes place. So, Lack of invariance, is not a problem because, contextual cues are always, taken into account. So, this General Auditory Framework, does not suffer from this, idea of, Lack of invariance. So, perceptual decisions made by matching relative goodness of sensory inputs to Prototypes in the memory. And, so there is a, how good the sensory input is, based on that, the perception is happening. So, information from other modalities, are also used, if available. So, not only the Speech perception system, visual and other modalities also, held the General Auditory Framework.

(Refer Slide Time: 55:41)



So, Chinchillas categorize, perception voice onset time. And so, this is an example, to show that, non-humans can also, do this kind of speech perception, and so, our speech is not special.

(Refer Slide Time: 55:51)

Direct Realism

Direct realism

- Sensory input sufficiently rich, allows us to completely recover object of perception
- Hence, we have direct awareness of world, no inferences necessary

Rejects "speech is special"

Motor system involved in speech perception

Handwritten notes:
- "Direct" (circled)
- "Lewin (1949)"
- "A."
- "Direct" (circled)
- "speech" (circled)
- "motor" (circled)
- "voice" (circled)

And, the last theory, that we are going to do today, is called, Direct realism. So, Motor Theory has made, a comeback with Carol Fowler, the idea of Direct realism, which is again, Carol Fowler is again from the Haskin Lab. Now, this new version of Motor Theory, is called the, Direct realism. And, it is based on, Gibson's 1979 Theory, that we have direct awareness of the world, because the sensory input is sufficiently rich, for us to completely recover an object of perception.

Now, if you remember, Gibson's idea, Gibson believed, or gave the idea of Direct perception, and believes that, the light, which is falling onto the eyes, that has all the information, that that is necessary for us to perceive. Similarly, this Direct realism, also believes that, the sound which is falling onto the ear, that has enough information, for us to perceive speech. So, directly Alison says that, sensory inputs is sufficiently rich, allowing us to completely recover, the objects of perception.

And, it is basically a Gibsonian view, as used in perception, 1979 Direct perception, if you do a direct comparison, if we want to compare, how this Gibson and group, fits into this idea. Now, one way in which, the Direct realism differs from Motor Theory is that, it rejects the idea that, speech is special. It believes, the speech is not special, it is general in case. And, so the speech, which is falling on your ears, has all the information, which is necessary for you to perceive.

You do not do this, categorical perception, or perceptual narrowing, or that kind of things are, not there. Also, direct realism, they believe that, speech processing, it involves, perceiving gestures, not Acoustic signal. So, Speech perception is not processing Acoustic signals, rather it is processing gestures. So, based on gestures, the perception of speech is happening, not on terms of, how Acoustic signal is perceived. The Direct realism also holds, on to the claim that, the motor system is involved, in perception of speech.

So, Direct realism believes that, motor system, so because, these gestures are done, by the motor system, or the vocal tract. And so, they believe, that the motor system is involved, in processing of speech. Now, the idea that, hence we have direct awareness of world, no inferences needed. Now, the very fact that, evidences are provided to this the this, Direct realism is through the concept of something called, motor neurons, sorry, Mirror neurons.

And, what is this Mirror neurons? So, Mirror neurons, which have been found in monkeys, for the first time, they provide biological evidence, for the tight linkage between, perception and action. And so, what are these Mirror neurons. Mirror neurons are neurons, which are in the brains of the primate, that fire, when they perform a particular action. And, they also fire, when a particular person, sees an act, that he has done before, or he wants to do.

And, they fire so, whether he is observing an act in being action, or whether he is acting, in both the cases, the motor neurons actually fire. And so, this this Mirror neuron is a biological evidence, to the fact that, the idea that speech is not, the idea that, there is nothing special in speech. So, it rejects the speech is special thing, and motor in system, involving Speech perception. First time, we come to know, that motor systems are involved in, perception of speech.

(Refer Slide Time: 59:19)

Direct Realism

Mirror neurons

- Neurons in primates that fire when perceiving or performing task
- Links perceptual and motor systems

Embodied cognition

- Cognition rooted in body's interactions with environment

So, Mirror neurons. These are neurons in primates, that fire when perceiving or performing a task. So, not only perceiving a task, it is also performing a task. And so, if we do an act, or we perceive an act being done, the Mirror neuron fires. So, that that is basically an evidence, a biological evidence, to the fact that, Speech perception is through Direct realism, and links, perception and motor systems, together.

Also, the Direct Realistic Theorists, they view Mirror neurons, as providing the mechanism, falling in perception and action, not just in speech, but in perception also. Also, the Direct realism, aligns itself, to the notion of Embodied cognition. So, it basically, it forwards the idea

of Embodied cognition, which basically says that, cognition rooted in body's interaction, with the environment. And so, what it is.

It a point of view, arguing that, cognition is rooted in the, body's interaction with the world around it. So basically, how the cognition is, not only in the in the minor brain, it is it is the way, we interact with environment. And so, this idea of Embodied cognition, is very basic to the idea of Direct realism. The essential purpose of brain, is to provide the organism, with information about the environment, in order to guide the behavior.

So, what the brain actually does, it provides information, to the to the organism, about his environment around us, the way the organism interacts with the environment, in terms of the information, which has been provided with the brain, that provides the way, we perceive speech. And, so this is the idea of Direct realism. So basically, that should bring us, to the end of this section. Now, what we did today, is we looked at, how Speech perception happens, in neonates.

So basically, small children, or just-born children, how do they perceive speech. And so, we saw several mechanisms, and we saw, several ways of responding. And, the basic idea, that develop is that this this small children, at a very early age, even one month from the time they are born, they are able to distinguish, these speech sounds. And, that happens because, the Mother speak to them in a very particular manner, in the Motherese.

And so, these Motherese, the way the Motherese is spoken, that gives the child to, make a perception of, what are the word boundaries, and phase boundaries, and that kind of a thing, and so, that helps them. And so, this is this is not theoretical because, the sucking reflex, the high sucking reflex and the cut head turning reflex, these two reflexes, basically demonstrate the way, demonstrate the fact that the child, is able to discriminate, or the infant is able to discriminate between, different speed sounds.

We also looked at the idea of how, this the idea of how the idea of we also are able to see, how this idea of perceptual narrowing, how this perceptual narrowing, and distributed learning, they are used by the children for, understanding speech, or variations in speech. We also saw how transition probabilities signal, the fact that, speech is perceived by children.

Next, we focused on, the various theories of Speech perception. They started with Motor Theory, which believes that, Speech perception is not only perceiving speech, it basically speech perception happens the way, the vocal cord is vibrating. And so, the vocal cord vibration, is primary to perception of speech. We also saw the idea in Motor Theory, that speech is special.

And, is basically furthered by Chomsky's idea of the Language acquisition device, which believes that, the child has innate modules, or predetermined modules in the brain, which helps in processing speech. Also, Fodder supports this theory, or this idea of Speech perception. Then, we come to the auditory framework, the General Auditory Framework believes that, the speech is not special.

So, Speech perception happens in the similar way, in which other non-animal, or other auditory sounds are perceived. And, he gives a lot of evidences, this theory, the General Auditory Framework, uses a lot of evidences, to prove that, speech is not special. And, it is general in nature. And, Speech perception happens, by a phonological method.

Now, there is a third concept of Speech perception, which is called the Direct realism, which believes that, again the motor systems is involved, in Speech perception. And, the speech which is coming to the infant, or anyone, for that matter, it has enough information in it, to be perceived, or to give you the idea of perception, and it rejects the idea that, speech is special.

And, it uses the concept, or it use the evidences of, Mirror neurons, the biological Mirror neurons, to support its idea that, speech is not special. But, it is general in characteristics, but then, the speech stream, which is coming to you, has all the information, which is necessary. Further, it also uses something called, Embodied cognition. It develops the Direct realism. It develops the idea of Embodied cognition.

Which means that, brain creates, or it makes us familiar, with the environment. And, our interaction in the environment, is what creates, cognition around us. That is a that is that is that is a review of, how speech is perceived, and speech is made and perceived. When we meet next, we look at, how speech is produced. So, the production of speech, by humans, that is what, we are going to focus, in the upcoming two lectures. But, until we do that, it is goodbye from here, and thank you.