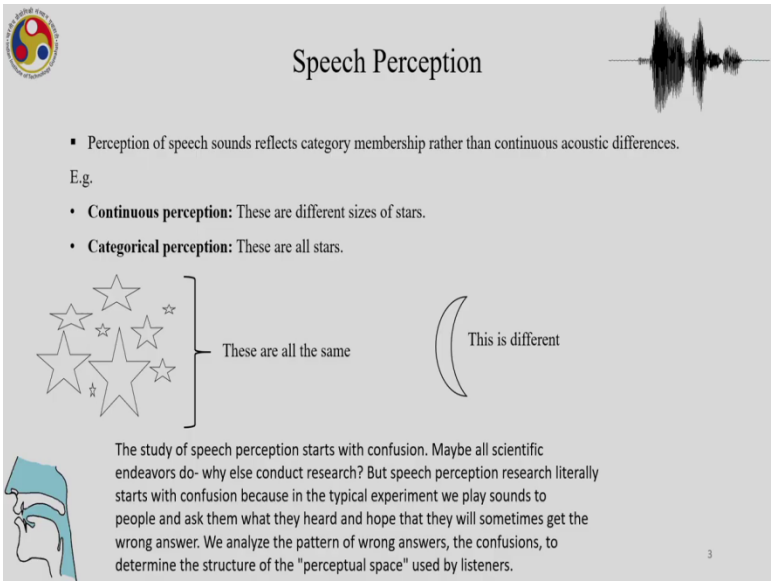
Phonetics and Phonology: A Broad Overview
Professor Shakuntala Mahanta
Department of Humanities and Social Sciences
Indian Institute of Technology, Guwahati
Lecture 14
Measuring Perceptual Distinctiveness, Multidimensional Scaling, Speech Perception Theories

Hello, and welcome to this NPTEL MOOCs course in Phonetics and Phonology, A Broad Overview.

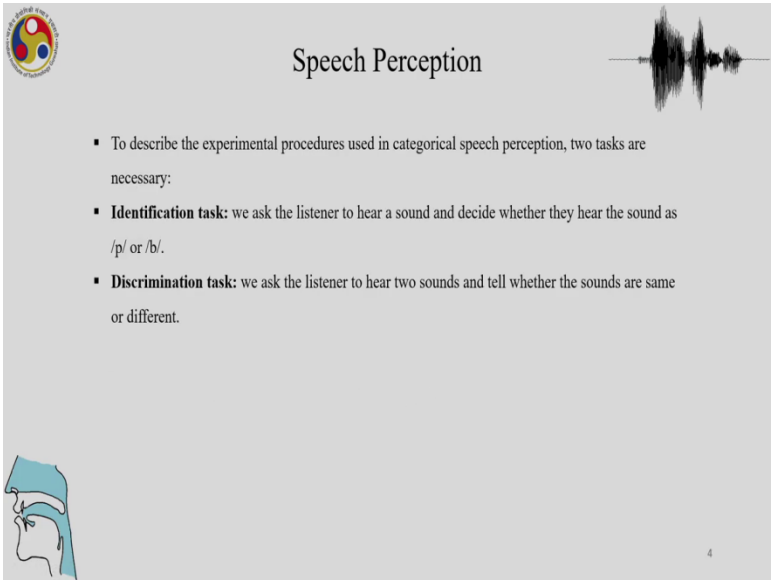
(Refer Slide Time: 00:40)



So, we are continuing with speech perception and as we already know, and we have seen in the last few classes, perception of speech sounds reflects a category membership rather than continuous acoustic differences. So, we learned about acoustic phonetic invariants and how there are invariant cues and that helps us to understand speech, human speech. And continuous perception is what you see here in an example there with the stars and then categorical perception.

So, there are various types of stars is all the same, but this is a categorical difference between the moon which is not the same as all the stars which are of various sizes, but we are conscious that these two are different categories.

(Refer Slide Time: 01:30)



So, speech perception then starts with something like a confusion, suppose. So, research in speech perception has to deal with all these problems in speech that one may be confused for the other, when we are hearing speech, we are aware of these possibilities. So, we can always hear one word to be different from the other and a lot of times, we miss hear or we do not perceive what was meant to be perceived in the same way.

So, speech perception starts with confused ability and if we take different classes of sounds, for instance, we take vowels, we take stops, we take fricatives then there is possibility that within that group, there may be some confusions. So, it underlines so, when that happens, there is always a possibility of substitution, substituting one sound for another sound. So, hence, we have to start with the hypothesis that predicts that consonants substitute one for the other.

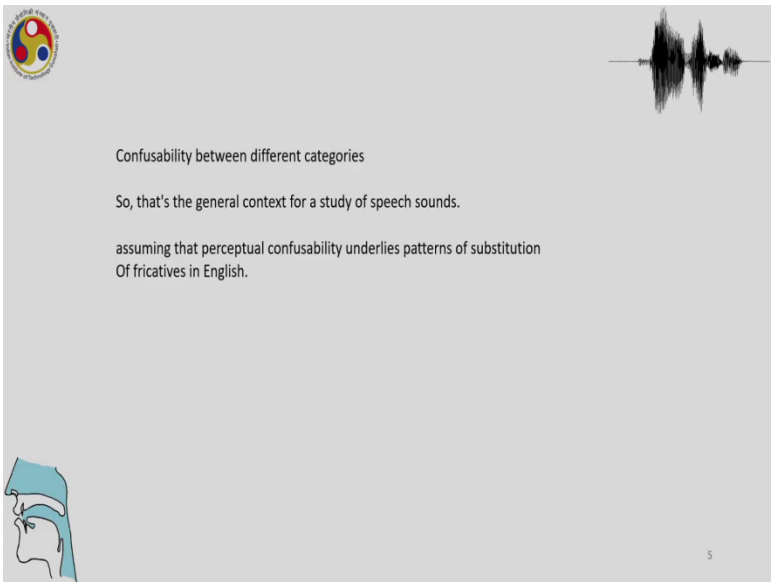
So, essentially, this hypothesis predicts that if there are consonants, then there is always a possibility that they will be substituted, one will be substituted for the other. So, let us begin to think about consonants. So, if we think about consonants within fricatives so, sometimes, so, in English it is often the case that, the is often heard as a or sa often heard as fa.

And what we see here in this slide, which talks about identification task versus the discrimination task, which we had seen before that, we asked the listener to hear a sound and decide whether

they heard the sound as pa or ba or we asked the listener to hear two sounds and tell us whether the two sounds are same or different.

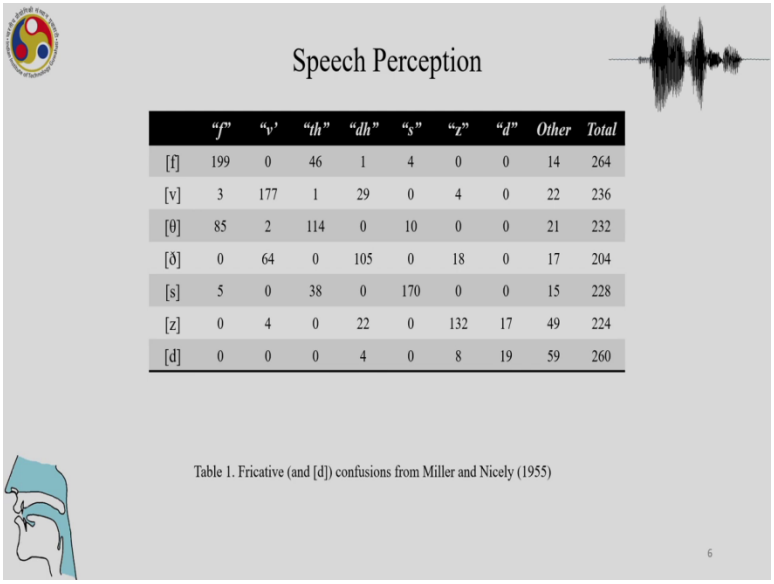
This is the identification task versus the discrimination task, which we are familiar with from our lectures on categorical perception.

(Refer Slide Time: 03:37)



So however, there could be also some confusability between different categories and assuming that there is perceptual configurability underlies a pattern of substitution of fricatives in English.

(Refer Slide Time: 03:48)



	"f"	"v"	"th"	"dh"	"s"	"z"	"d"	Other	Total
[f]	199	0	46	1	4	0	0	14	264
[v]	3	177	1	29	0	4	0	22	236
[θ]	85	2	114	0	10	0	0	21	232
[ð]	0	64	0	105	0	18	0	17	204
[s]	5	0	38	0	170	0	0	15	228
[z]	0	4	0	22	0	132	17	49	224
[d]	0	0	0	4	0	8	19	59	260

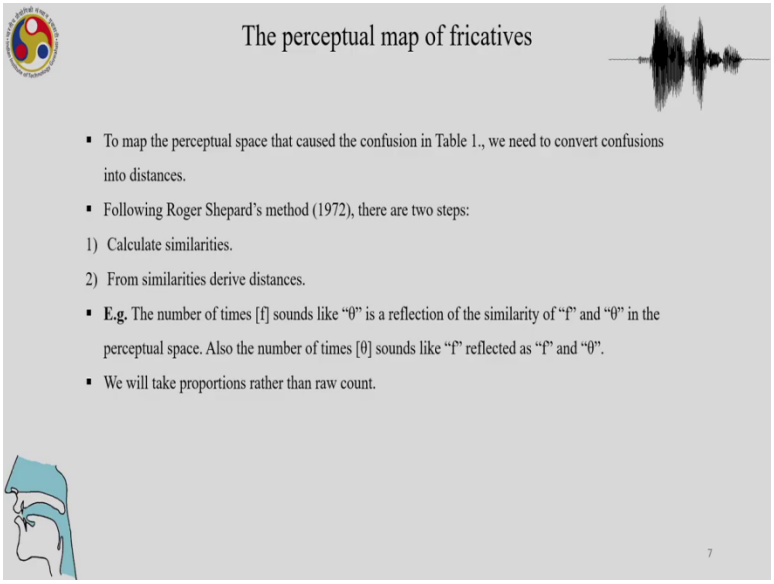
So, this is what we are seeing here with regard to fricatives, these are all the English fricatives. So now, what does this chart tell us? This chart tells us of all the times that f was heard as fa, was heard as sa, was heard as was dh, was heard as z or d, or all the times, va so, this is fa, this is va, va was heard as fa, va was heard as v, and also as s or z or s z etcetera.

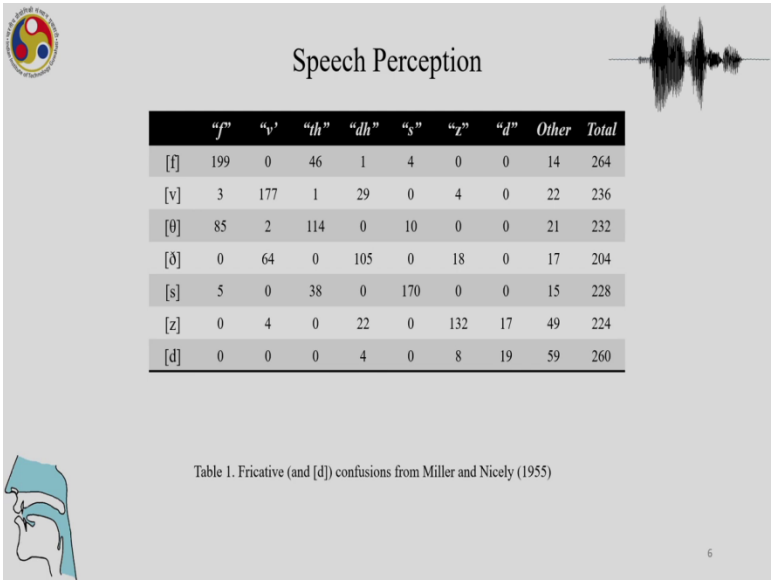
Now, what we see here is that this is a figure from Miller and Nicely and the confusions, the number of confusions of the fricatives and, d and given a set of when speakers were asked to listen to these words, and ask them what they heard. They often heard s as f so you can see the number of times s was heard as f. What this chart tells us is that these are the instances of how many times were there heard as one of these sounds?

So, this is a sort of a confused ability map that we have in front of us. And this is from the research done by Miller and Nicely in 1955. How do we come to this sort of a speech perception confusability map, there is a procedure for this. So, we have to collect a number of words, and then we have to record them. And then we have to ask speakers to identify the words. And then when they mistake one for the other, we note down the mistakes.

And that is how we come to this speech perception map.

(Refer Slide Time: 05:48)





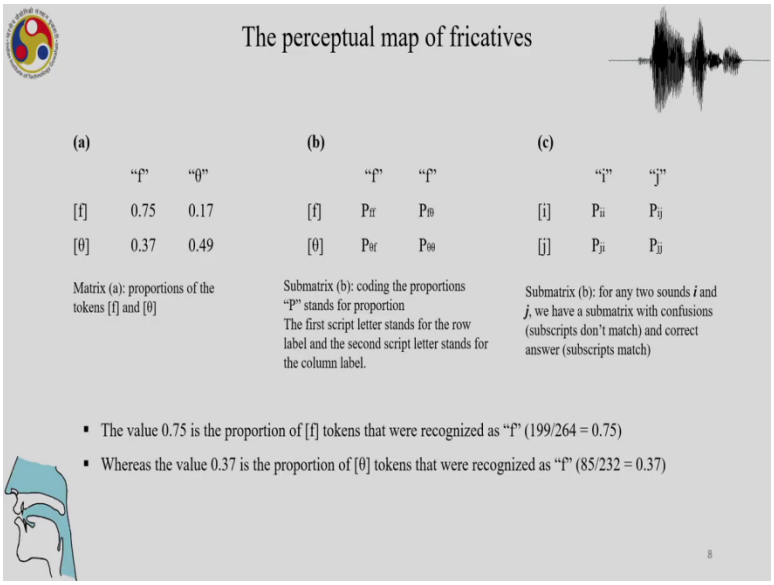
	"f"	"v"	"th"	"dh"	"s"	"z"	"d"	Other	Total
[f]	199	0	46	1	4	0	0	14	264
[v]	3	177	1	29	0	4	0	22	236
[θ]	85	2	114	0	10	0	0	21	232
[ð]	0	64	0	105	0	18	0	17	204
[s]	5	0	38	0	170	0	0	15	228
[z]	0	4	0	22	0	132	17	49	224
[d]	0	0	0	4	0	8	19	59	260

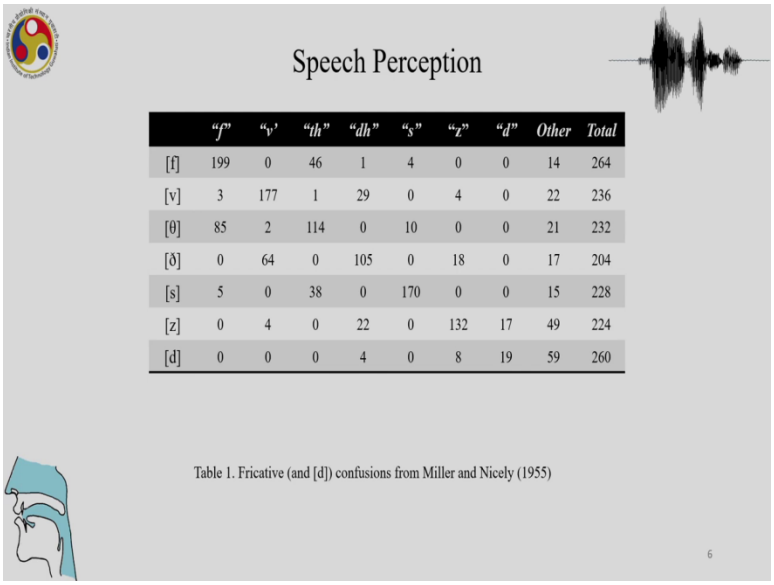
So, to map the perceptual space that caused the confusion that we see here that v was heard as f, s was heard as f, so this is the confusion that we are talking about. So, we need to convert these confusions into distances. Now, the mathematician Roger Shepard had devised a way for doing this. So, how to do this we calculate similarities, and from similarities, we derive the distances of one to the other.

So, the number of times f sounds like s is a reflection of the similarity of f and s in the perceptual space. So, that is, the reason as we said before, that there is confusion of any manner in a

perception is because there is a similarity in the two sounds. And also the number of times th sounds like f is reflected as f and th, and we will take the proportions rather than the raw count.

(Refer Slide Time: 06:47)



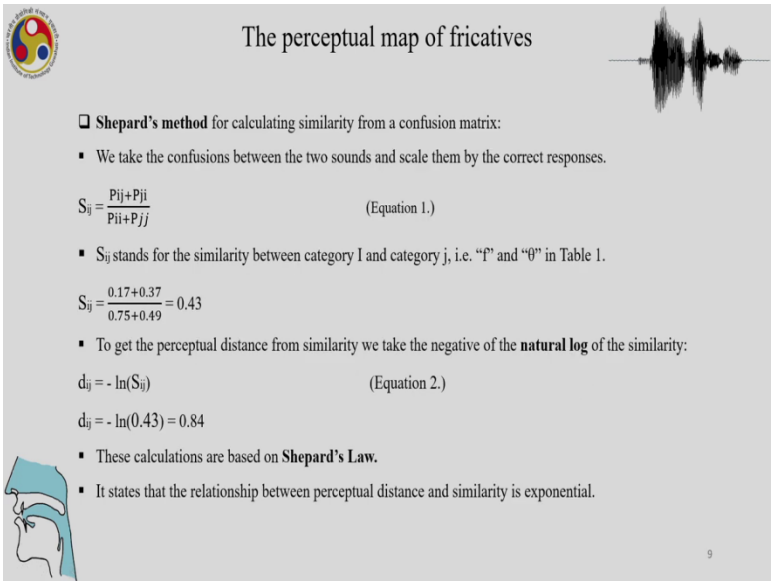


So, this is the perceptual map of fricatives. And this is the matrix which shows, if we count the number of times, not the raw times, we take the proportions. So, the value 0.75 that we see here is the proportion of f, or the f tokens that were recognized as f. So, out of the 264 times f that the speakers heard those out of that 199 times, it was heard as f. So, the proportion is 0.75, and 0.37 here is a proportion of th tokens that were categorized as f.

So, out of the 232 times that th given to the speakers some more of actually f. So, that is 0.37. So, this is how we code the proportions, the P here stands for the proportion, and the script, so stands for the row label, and the other stands for the column label. So, this is how we are mapping the perceptual map of fricatives. And now again, here, what is i and j. So, what you see here as f f of f s one is letter for the row label, and one is a letter for the script.

So basically, if we look at this map, again, we put the row label f and the column label v. And that is how we arrive at these two subscripts. The subscript that you see here pf pff f s, one stands for the row label, one stands for the column label. So, they are indexed as i and j. So, I stands for the subscript for the row label and one subscripts that, so subscript that does not match in the subscript that match.

(Refer Slide Time: 08:44)

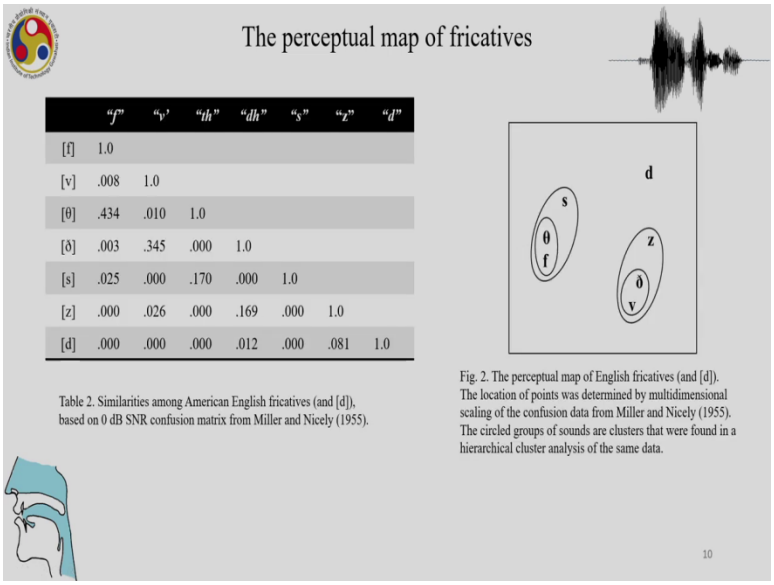


So Shepard's methods for calculating similarity is from a confusion matrix. So, this is how the confusion matrix is, is calculated. So, we take the confusion between the two sounds and scale them by the correct responses. So, here, Sij stands for the similarity between category i and category j, that is f and th in table one. So, we take these values of i and j, as we had just talked about the two subscripts and we put them here, and then we add them and divide by the total correct responses, and we get the value.

So, to get the perceptual distance from similarity, this is another calculation that we have the two equations, one is where we divide by the correct responses, and the other is where we take the negative of the natural log and then in maths, this is how we find the distance. This is a perceptual distance and this is the similarity and this is the perceptual distance. So, this is how we get the similarity how similar that they are and what is the distance.

And calculate the distance, we take the negative of the natural log, and this is the formula for the distance. So, these calculations are based on Shepherds law. And it states that the relationship between perceptual distance and similarity is exponential.

(Refer Slide Time: 10:17)



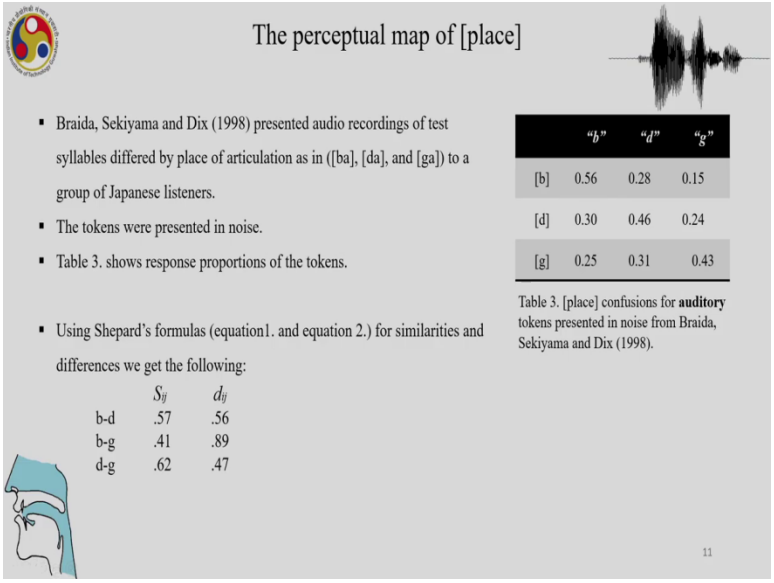
So, it can be exponential. So, as we can see from this is a similarity matrix. So, the similarities among American English fricatives, based on the confusion matrix from Miller and Nicely, so, the perceptual map of English fricatives, and d is a stop, so, it is outside of this map. And the location points was determined by the multi dimensional scaling of the confusion data. So, by scaling the data, these values that we have, this is what we come up with.

And now, two things here, we find clusters. And the circle groups are clusters. And there was a hierarchical cluster found, and one which is outside of this cluster. So, s and f were closer and z and v were closer than z. So, something you have to notice here is that, how the two voice and

voiceless groups are patterning separate, they are clustering separately. So, the s and f are very close together.

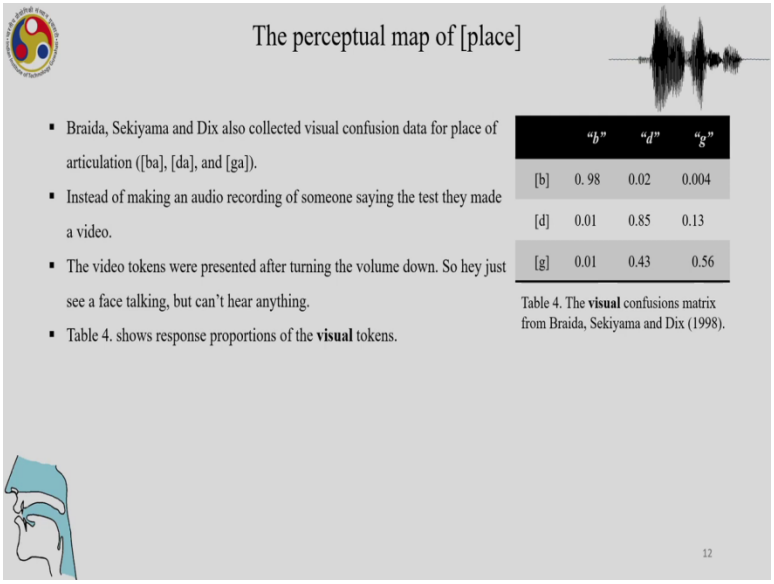
So, is f, so, are these two sounds v and z whereas z is patterning with v and z, but the clusters formed are s f and the v and d is outside of this cluster.

(Refer Slide Time: 11:35)



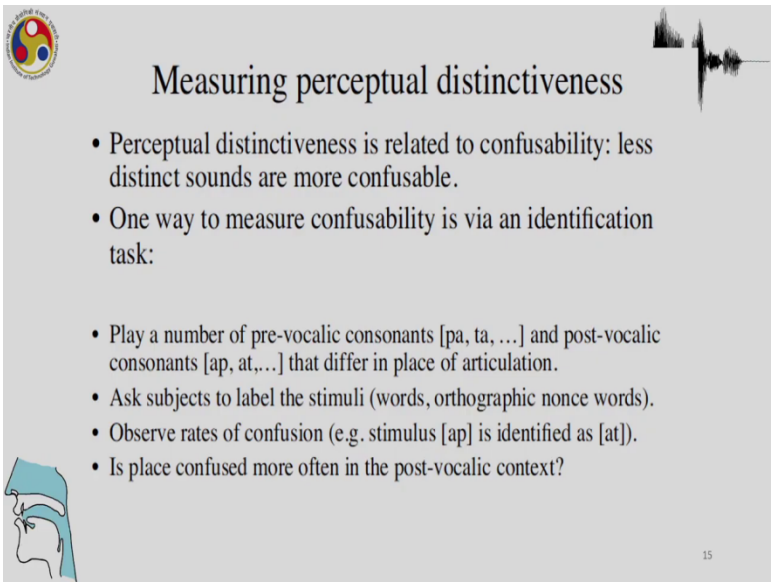
So, this is how we have the hierarchical cluster. Then look at some other perceptual data that we have, we find similar analysis where it finally shows that why something is confused for something most of the time, because they are very, very similar to each other the sounds and we find them when we calculate the distance between the two sounds.

(Refer Slide Time: 12:00)



So, the perceptual map of place stat that we have here also have similar findings with regard to the confusion matrix.

(Refer Slide Time: 12:11)

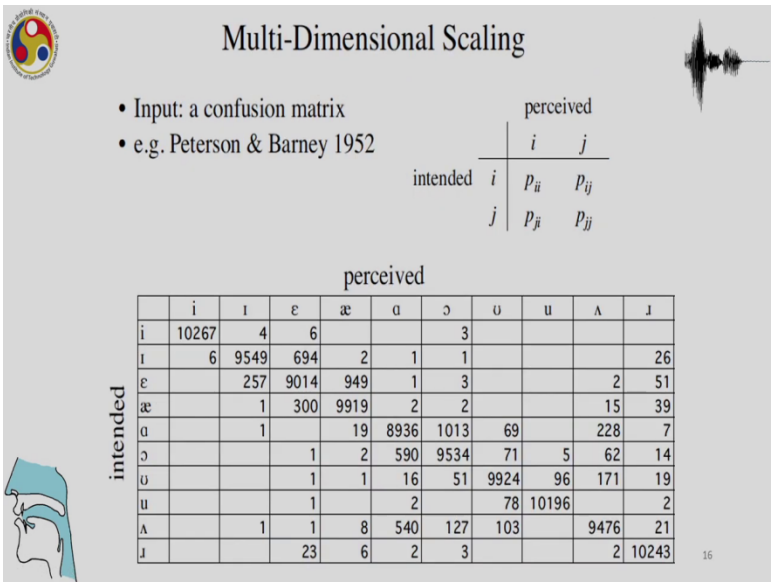


So, let us look at the perceptual distinctiveness of stops. So similar to what you have just seen, we will look at something similar. So, now repeating what we have been saying that perceptual distinctiveness is related to confusability, less distinct sounds are more confusable. One way to

measure confusability is via an identification task, play a number of consonants. So, in this case, we are talking about stops and asked the subjects to label the stimuli.

So, what did you hear pa ta, did you hear ap at, so participants will have to write down what they heard. And we then, we observe the rates of confusion as to what was heard as what. So, now, the question that we will ask here is that is, placed confused more often in the post vocalic context.

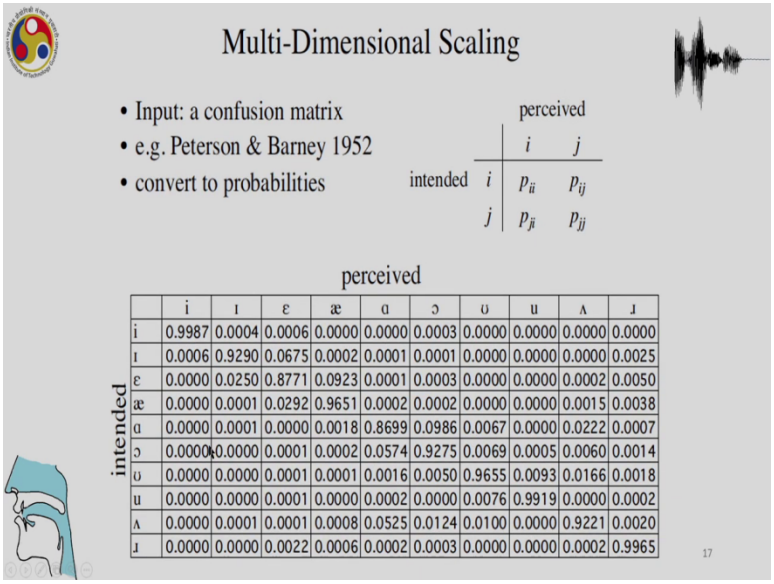
(Refer Slide Time: 13:05)



Now, this is similar to what you have just seen with regard to fricatives. So, suppose, here, now, we have the chart here on one side, or what you see vertically, here is the intended vowels and here we have, what was perceived. So, were they perceived correctly. So, i was perceived correctly most of the time, and then there was some confusions, however, if you compare i and e number of times, e was confused for a is much more than the 694.

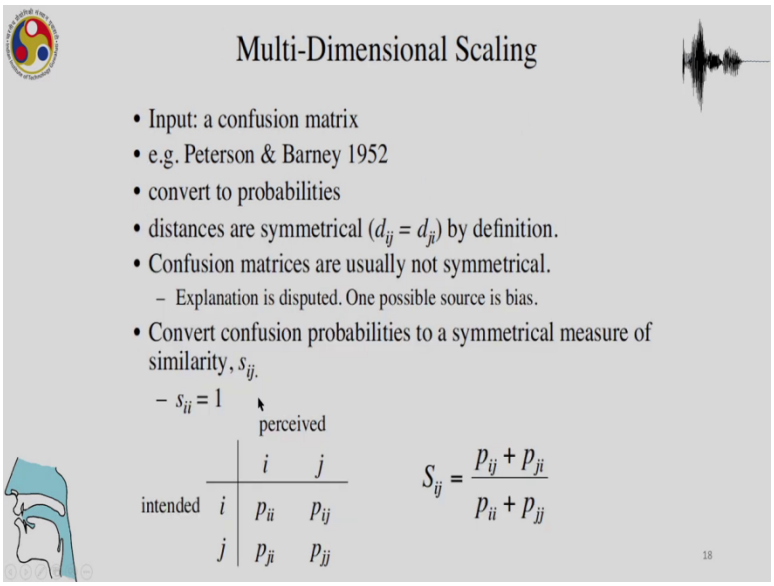
And similarly, when we have vowels like aa so was very often confused for au like 1013 times. So, this is what something that you see here, what was intended.

(Refer Slide Time: 13:52)



And then now, we as we said, before, we calculate the proportions as to how many times i was perceived as e and how many times i was perceived as a.

(Refer Slide Time: 14:05)

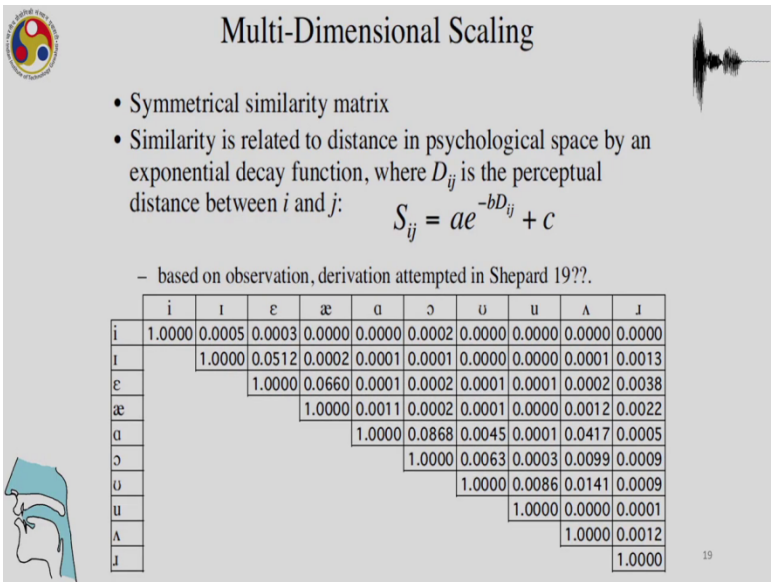


Now, we have the proportions. And then we calculate from the proportion after we find, that we use our similarity formula here, where we divide with the number of times that something was supposed to be perceived as i or a and then as to what was the actual value, they ended up

perceiving them as, and then we also find out distances similar to what we have just seen. So, distances, confusion, matrices are usually not symmetrical.

So, we do not find that if i is perceived as a more and that is i will not be perceived as i or e is perceived a lot of times as a, it is not the same thing so they are not symmetrical. So, similarity is related to distance in the psychological space by an exponential decay function.

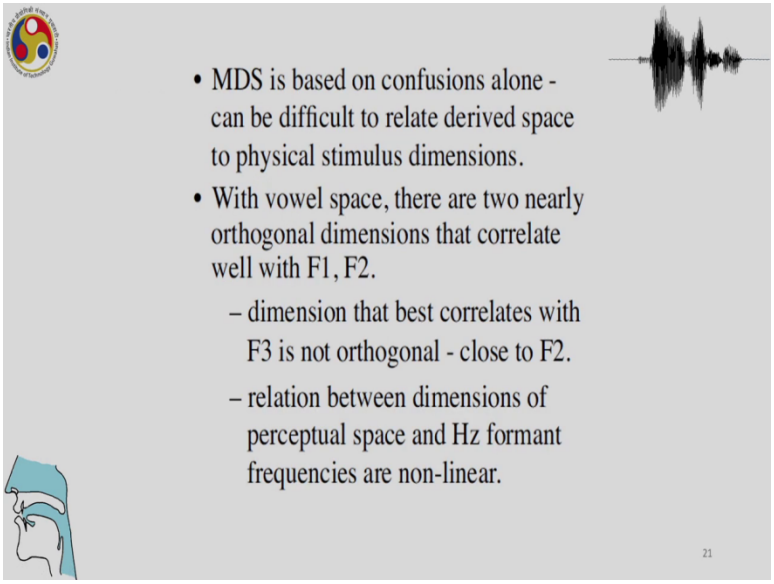
(Refer Slide Time: 14:56)



	i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɹ
i	1.0000	0.0005	0.0003	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000
ɪ		1.0000	0.0512	0.0002	0.0001	0.0001	0.0000	0.0000	0.0001	0.0013
ɛ			1.0000	0.0660	0.0001	0.0002	0.0001	0.0001	0.0002	0.0038
æ				1.0000	0.0011	0.0002	0.0001	0.0000	0.0012	0.0022
ɑ					1.0000	0.0868	0.0045	0.0001	0.0417	0.0005
ɔ						1.0000	0.0063	0.0003	0.0099	0.0009
ʊ							1.0000	0.0086	0.0141	0.0009
u								1.0000	0.0000	0.0001
ʌ									1.0000	0.0012
ɹ										1.0000

And this is a bit more complicated than what we had found that there is this exponential decay functions.

(Refer Slide Time: 15:07)



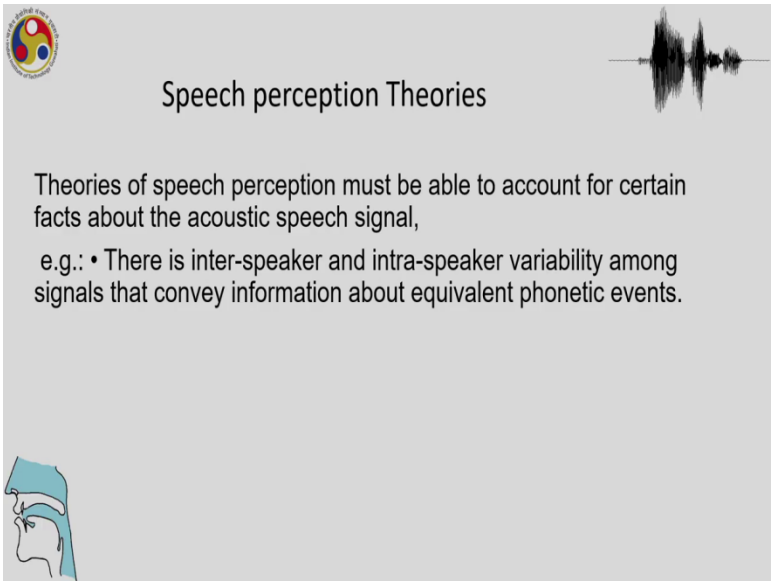
Now, the whole point of doing this extensive lecture on multi dimensional scaling is not to make you understand these formulas that are given by Shepherd, but help to make you understand something that there is perception related to confusability , these things can be mathematically understood. And then there are various mathematical equations as to how we can derive those similarities and distances between sounds.

And although you will, something to remember is that you will not be evaluated for these distances and similarities, this is just a lecture showing that we can do these things with the help of mathematical equations. And then we come to the end of our presentation today on perceptual confusions with some input on the vowel space. So, with regard to vowel space, there are two nearly orthogonal dimensions that correlate well with F1 and F2. So, with vowels, there is additionally the formant values that we have.

And the dimension that best correlates with F3 is not orthogonal, it is close to F2. So, F3 as we know is not always the most important formant taken to understand vowels. So, relations between dimensions of perceptual space and hertz formant frequencies are again nonlinear. So, we cannot equate the formant frequencies and what we have from the formant frequency values in hertz, how in the perceptual space, they are mapping those formant values, it is not linear that it does not fall in place one on one.

The point of making this presentation to you is to show that vowels have complexities, which also have to be dealt with differently from the consonants.

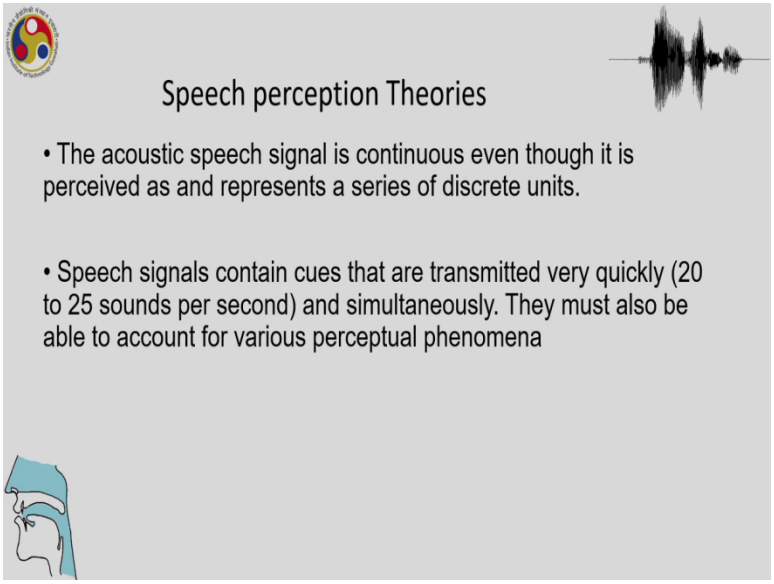
(Refer Slide Time: 17:11)



So, coming now, the brief presentation on confusability that is there in perception, let us now have a brief look at the speech perception theories, which are there in the literature and how we understand perception with regard to these taking into account the various things postulated by the speech perception theories. So, theories of speech perception must be able to account for certain facts about the acoustic speech signal.

And there is inter speaker and intra speaker variability among signals that convey information about equivalent phonetic events.

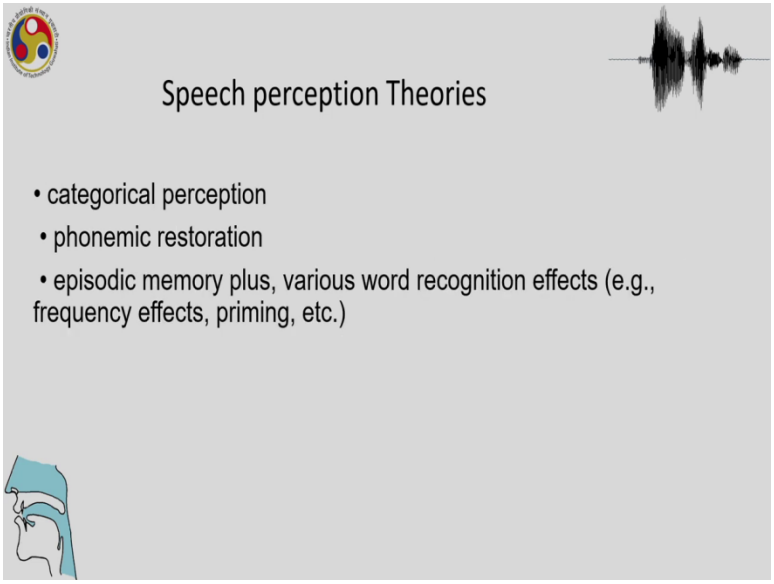
(Refer slide Time: 17:55)



So, the acoustic speech signal is continuous even though it is perceived as and represents a series of discrete units. So, that we have seen that in the last few lectures, that even though speech is continuous, there is no break between those the different parts of a speech that we produce, but even then, we perceive them as discrete units. So, that is how speech is like, we do not perceive them as the sounds, that make up the words and sentences would not perceive them as continuous we perceive them as discrete.

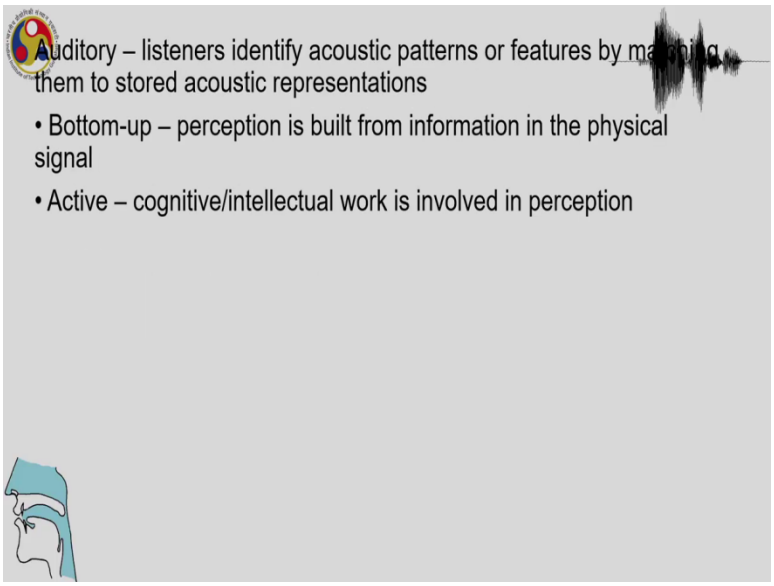
So, speech signals contained cues are transmitted very quickly about 20 to 25 sounds per second and simultaneously, and they must also be able to account for various perceptual phenomena. So, speech perception theories have to deal with these issues of speech that it is continuous, but it is perceived discreetly, that speech is transmitted very quickly simultaneously, and all these things must be accounted for by speech perception theories.

(Refer Slide Time: 18:57)



And there is categorical perception, there is phonemic restoration and episodic memory plus various word recognition effects.

(Refer Slide Time: 19:09)

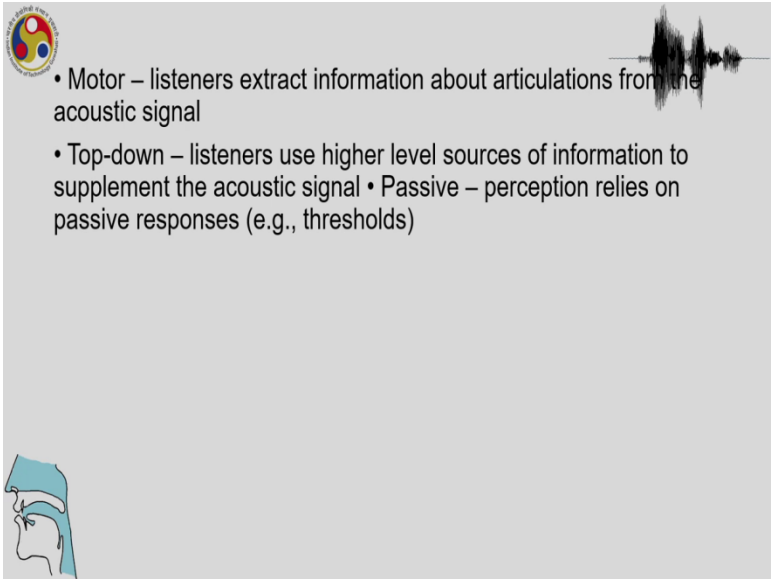


And also there are various ways in which speech perception can be understood. So, one is auditory listeners identify acoustic patterns by assigning them to stored acoustic representations. So, then we have bottom up so, speech is perception is built up from the main information in the acoustic signal, in the physical signal. So, these are questions that speech perception theories will

have to deal with, speech always built up from the signal or is it not bottom up, it is top down is it from it is, is it that we constantly map with our knowledge.

And then active versus passive. So, cognitive intellectual work is involved in perception or that nothing like that is involved.

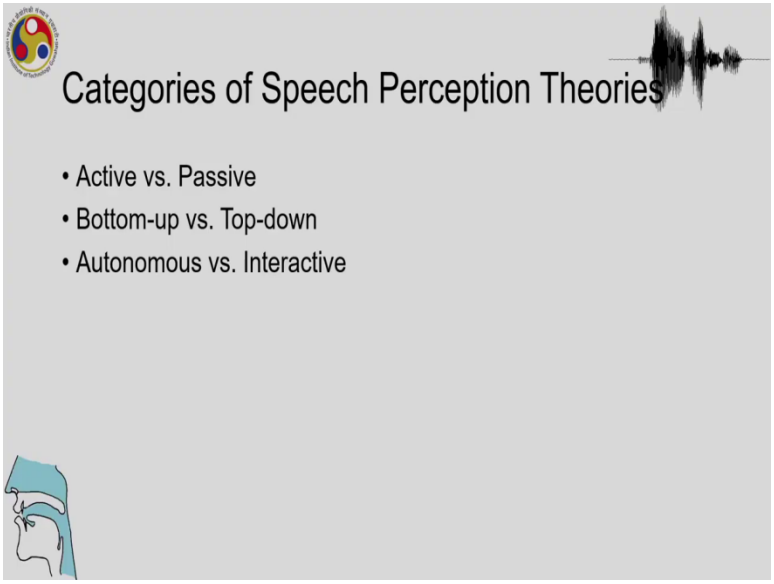
(Refer Slide Time: 19:56)



So, motor properties, so listeners extract information articulation from the acoustic signal. That is what motor theory says, so that is another aspect that speech perception will have to take into account. And also, top down listeners use high level sources of information to supplement the acoustic signal.

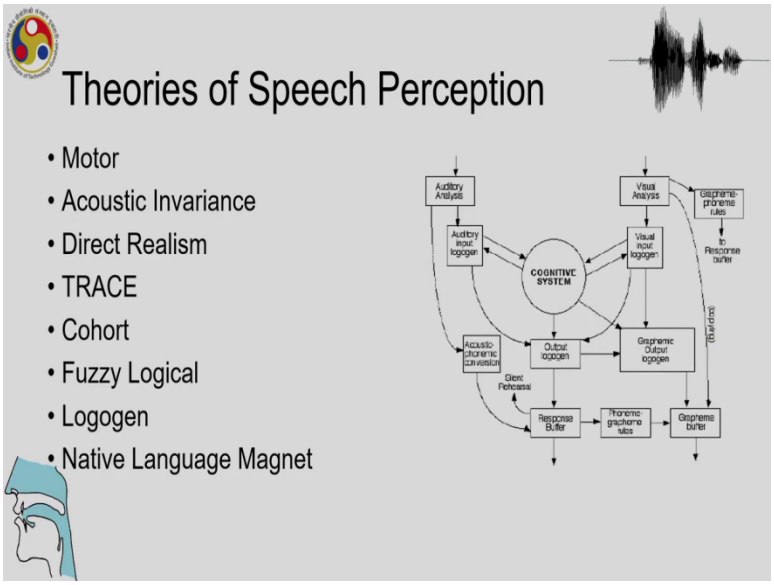
And again, is it active that is cognitive work is involved? Or is it that passive perception relies on passive responses? So that it is completely passive, there is no active involvement of our intellectual abilities involved in perception. So, whether it is active or passive, whether it is bottom up or top down, whether it is auditory or motor, these are the different things that speech perception theories will have to take into account, while dealing with speech perception or coming up with an idea, which sort of closely tells us in a certain way that we can understand speech perception, the way it happens.

(Refer Slide Time: 21:00)



Now, if we see all these speed categories or speech perception theories, we will see that most of the things that have been dealt with by speech perception theories are around these ideas, active versus passive, bottom up versus top down, interactive versus autonomous or passive versus active auditory motor. So, these are the things that speech perception theories are dealing with most of the time.

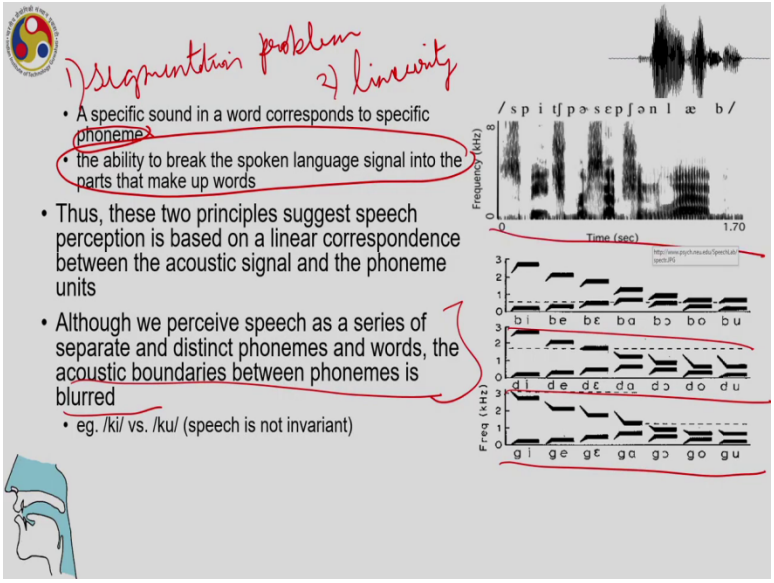
(Refer Slide Time: 21:22)



So, these are the different theories of speech perception that we have, we have motor theory, we have theory of acoustic invariants, we have direct realism, we have trace theories, we have cohort, we have other like native language, magnet or fuzzy logical or logogen. So, there are many theories, we will see a couple of them.

So, here we have the cognitive system, and then we have all these different things that we have to count for and different theories are accounting the, for the cognitive system in different ways using auditory analysis, or visual analysis or grapheme phoneme rules or acoustic phonemic conversion, what are the different things which speech perception theories are concerned about? Let us have an overview of these things before we wrap up speech perception, and what are the different ways in which speech perception is seen through the lens of these different theories.

(Refer Slide Time: 22:22)



So, let us recap the things that a speech perception theory will have to deal with. So, there is a segmentation problem. One, there is a segmentation problem, which we talked about in all the previous lectures, and there is the linearity issue. Segmentation problem, a specific sound in a word corresponds to a specific phoneme, and the ability to break the spoken language signal into the parts that make up words.

So, this is the segmentation issue that the signal has, the signal is continuous, and but we have the ability to break it up into the constituent parts signal, there is no break in the signal, but that

the achievement of human perception is that we are able to break it into the component words, in component sounds, that is our segmentation problem.

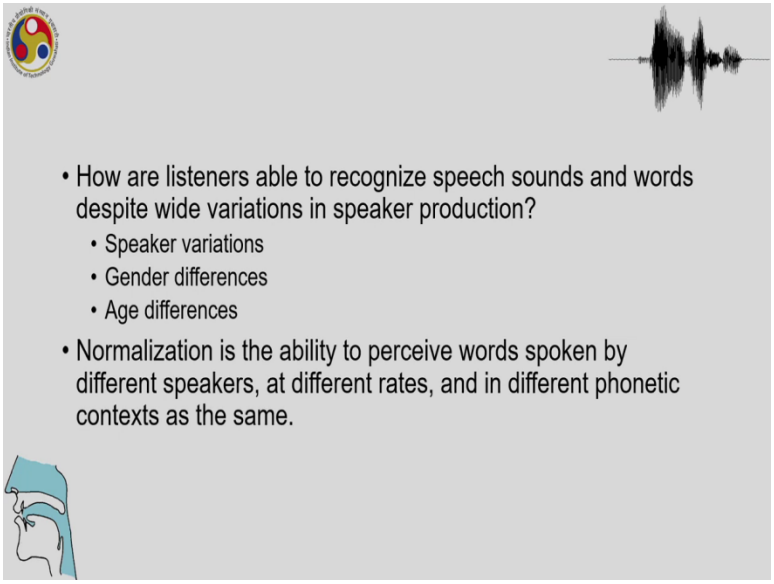
And then linearity principle is that specific sounding word corresponds to a specific phoneme. So, this is happening in a linear fashion, we are not jumbling up the sounds and still producing the words, these things are happening in a linear fashion, we are still able to break down into discrete parts, and this is how perception is happening. Although while we are speaking, it is a continuous stream of sounds.

So, these two problems issues principles, like the segmentation problem or the linearity principle suggests that speech perception is based on a linear correspondence between the acoustic signal and the phonemic units. So, this is happening continuously and in a linear fashion one by one, we are breaking up the words and the sounds and making sense of what is being spoken. So, although we perceive speech as a series of separate and distinct phonemes and words, the acoustic boundaries between phonemes is blurred.

And although we are perceiving the speech as a series of separate distinct phonemes and words, the acoustic boundaries between phonemes is blurred and this is what we keep on saying that, the speech signal has very minimal information, but then the speech perception ability is that despite the minimum information, despite all the variation that is possible in the acoustic signal, a listener is able to perceive and discreetly break up the sounds component sounds of a word and sentence and linear in a linear fashion map one to the other, even though the acoustic signal is rife with all sorts of variation, is rife with all sorts of information, which may not be always invariant.

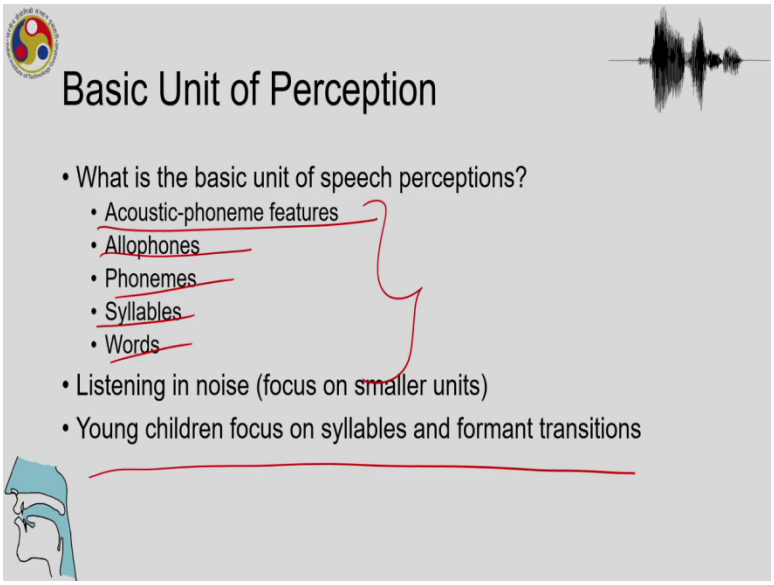
We heard that in the previous lecture that not always invariant, for instance, the form and transitions as we can see here the subject change. So, the speech signal is also continuous.

(Refer Slide Time: 25:30)



So, how are listeners able to recognize speech sounds in words despite wide variations in speaker production? So, speaker variations and gender differences, age differences etceteras are the things which contribute to the variation in the speech signal. So, normalization is the ability to perceive words spoken by different speakers at different rates and in different phonetic contexts as the same.

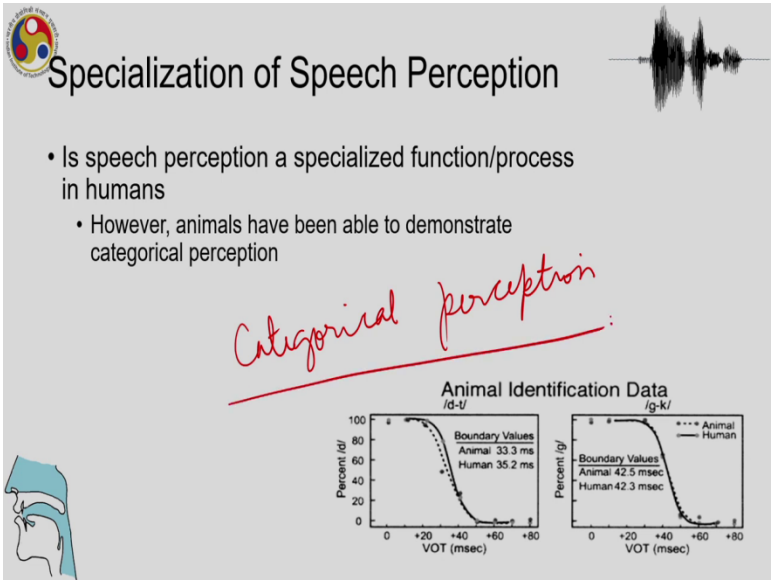
(Refer Slide Time: 26:00)



And then some other things that speech perception theories will have to take into account, first thing is the, what is the basic unit of a speech perception. So, is it acoustic phoneme features, is it allophones, phonemes, syllables, words what is the basic unit of speech perception and then listening in noise, so, we are listening and we are perceiving speech in a lot of noise. So, there may be other people speaking, there may be noise, industrial noise, there may noise from vehicles, there may be noise from nature, environment etcetera.

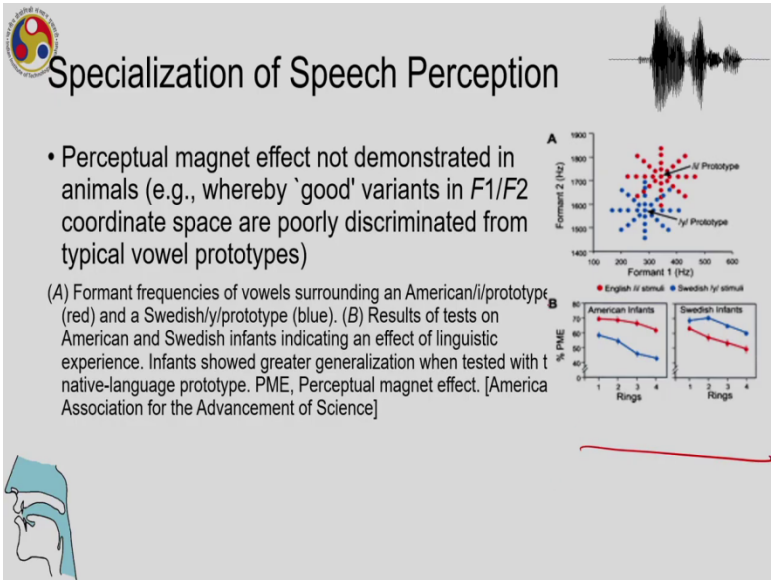
It is yet we still perceive. So, and then they can be very general noise, coming from virtually, almost nothing at all just distance etcetera. So, there is always, there is always the possibility of noise in speech and yet we are able to perceive. So, that noise element will have to be accounted for by speech perception. And then how do children start perceiving when they are trying to respond to what they have listened to. So, it is seen that they focus on syllables and form and transitions.

(Refer Slide Time: 27: 07)



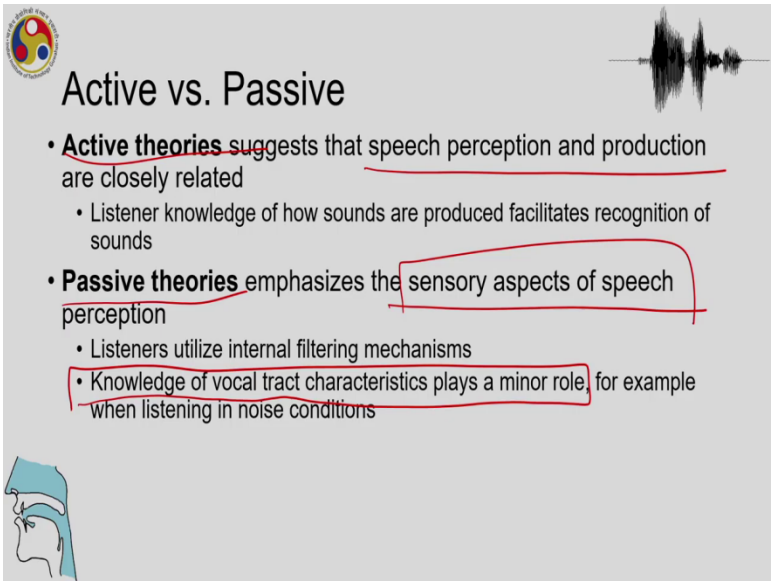
And then specialize, speech perception in speech perception a specialized function in humans, humans are able to do, what we know now as categorical perception. And we know that not just humans, we have seen that in the previous lecture, even animals are capable of categorical perception.

(Refer Slide Time: 27:32)



However, the perceptual magnet effect, we will study what is the perception magnet effect is not very prominent in animals. So, poorly discriminated from typical vowel prototypes. So, this is a diagram showing the responses to perceptual magnets by of prototypes by American infants and Swedish infants. And then we can see that they respond poorly to perceptual magnets.

(Refer Slide Time: 28:01)

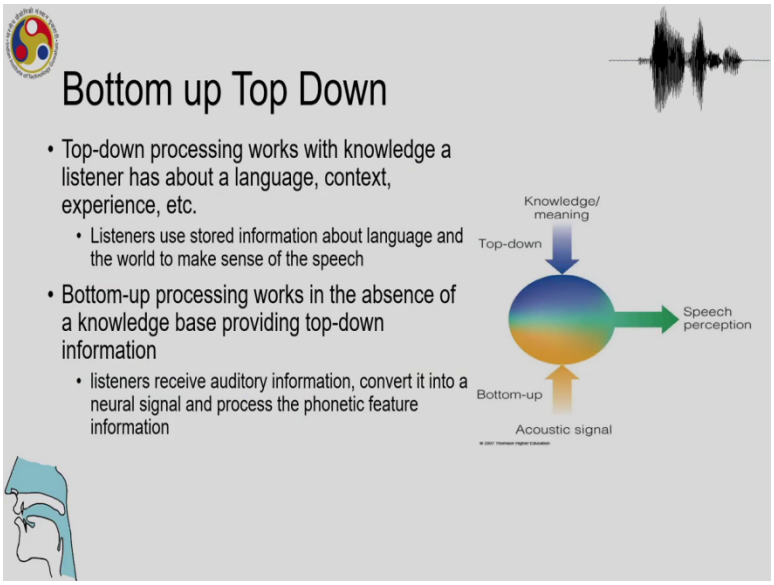


Again, we have quintet ourselves with active versus passive theories, a bit before and we know that active theory suggests that speech perception and production are closely related. And listener

knowledge of how sounds are produced facilitates recognition of sounds. And passive theories emphasizes the sensory aspects of speech perception. And listeners utilize internal filtering mechanisms and knowledge of vocal tract characteristics play a minor role, for example, by listening in noise conditions.

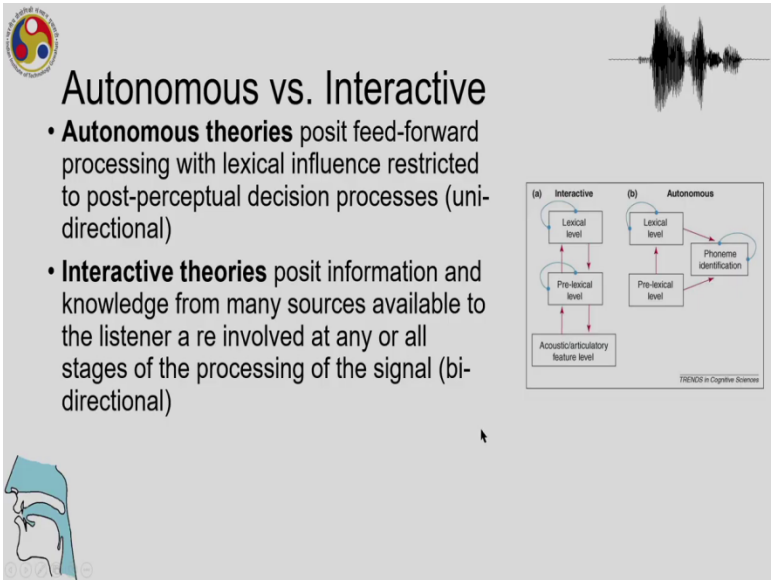
So, active versus passive theories, we have acquainted ourselves with that and then in one suggests that speech perception production are closely related, in the other it emphasizes the sensory aspect. So, of speech, it knowledge plays a lesser role, the role of sensory aspects of speeches highlighted in passive theories. So, active theories put a greater emphasis on our intelligence, human intelligence in perception.

(Refer Slide Time: 29:05)



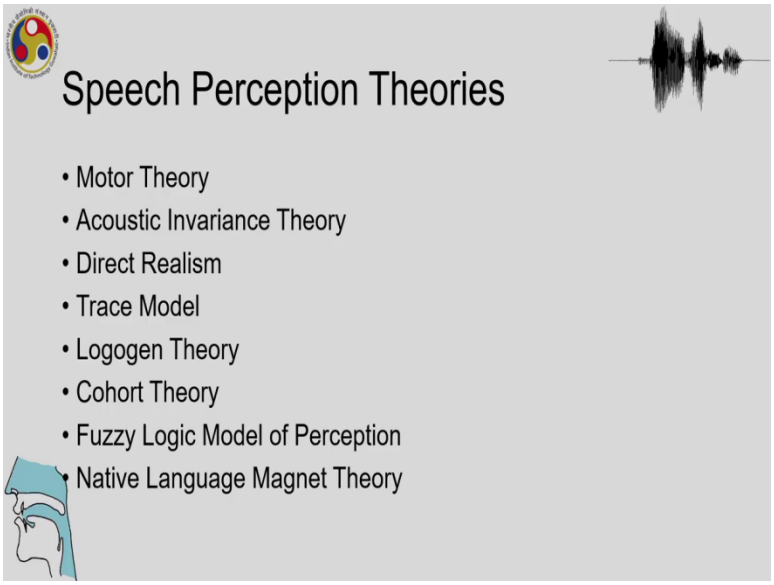
And so, again bottom up, top down we have seen this before and top down processing works with a knowledge of the listener has about language context etcetera. And bottom up processing works in the absence of a knowledge base providing top down information. So, listeners receive auditory information converted into a neural signal and process the phonetic feature for information.

(Refer Slide Time: 29:30)



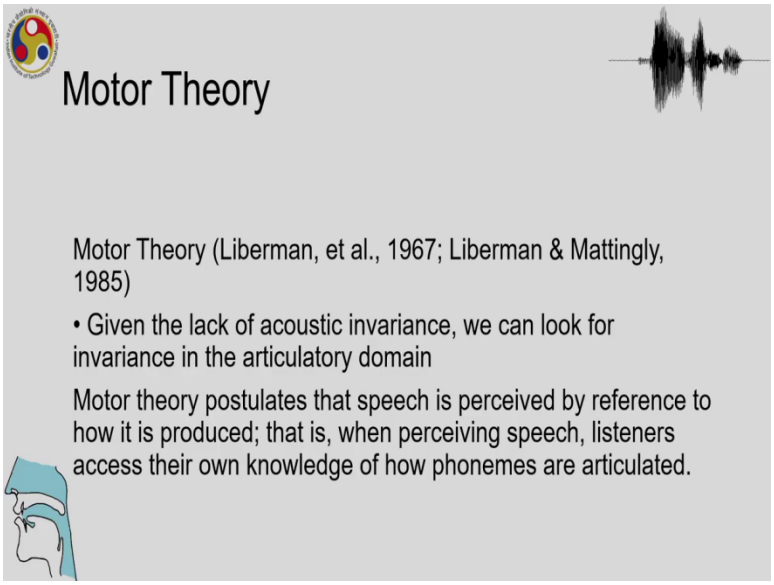
And then we have the autonomous versus interactive theories. So, we will not talk about autonomous versus interactive so much as in the difference between feed forward processing and because you are not talking so much about processing, but it is good to know that such theories are also there.

(Refer Slide Time: 29:51)



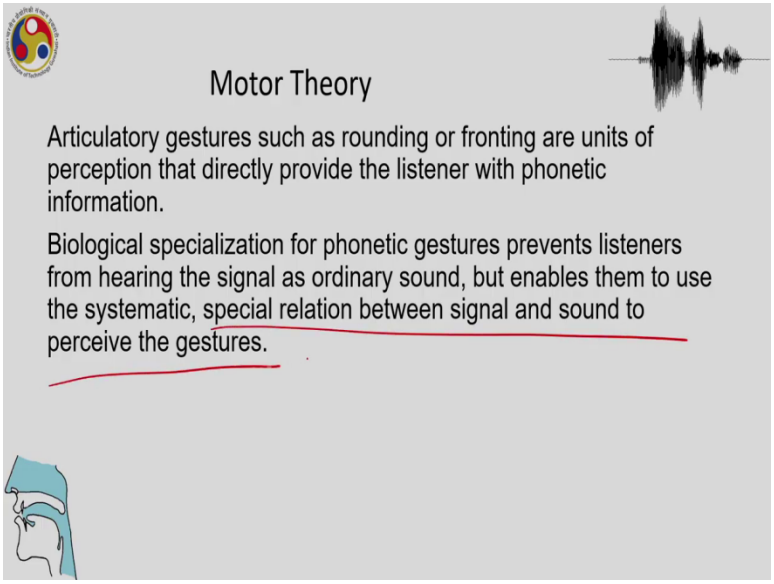
So, these are the different speech perception theories as we have just mentioned, we have motor theory acoustic invariance theory, we have direct realism, trace models, we have logogen, we have cohort, fuzzy logic model of perception native language magnet theory.

(Refer Slide Time: 30:06)



So, what is motor theory? Motor theory given the lack of acoustic invariants we can look for invariants in the articulatory domain and motor theory postulates that speech is perceived by reference to how it is produced, and that is when perceiving speech listeners access to your own knowledge of how phonemes are articulated.

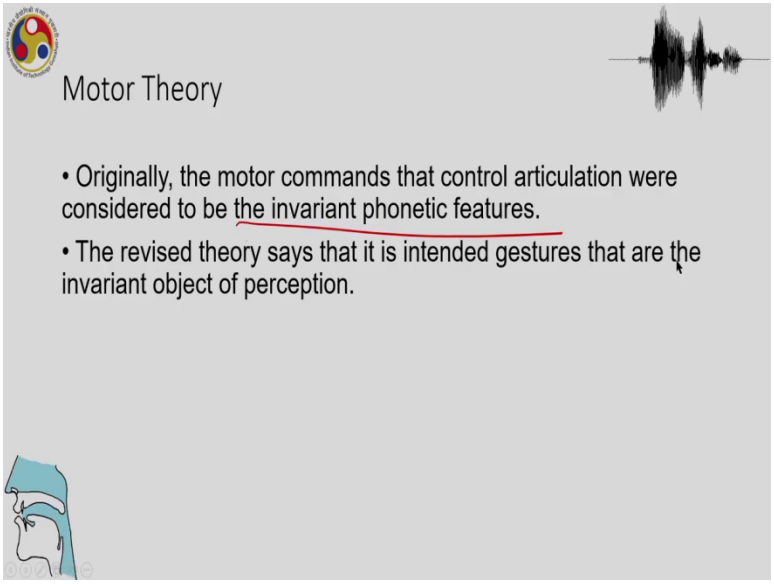
(Refer Slide Time: 30:29)



And motor theory shows that articulatory gestures such as rounding or fronting etc. These are the actual units of perception. That is what motor theory tells us, and it directly provide the listener with phonetic information. So, the biological specialization for phonetic gestures prevents listeners from hearing the signal as ordinary sound, but enables them to use that for systematic special relation between signal and sound to perceive the gestures.

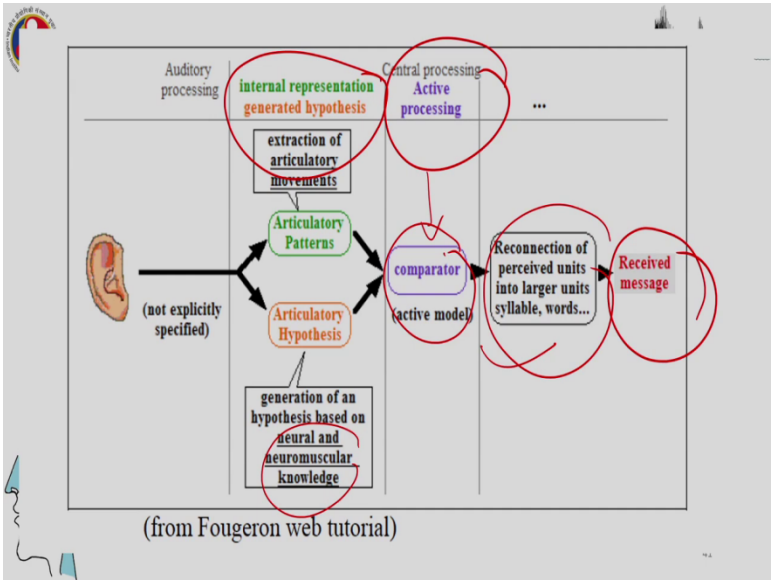
So, the articulatory gestures are the real perceptual units. So, and then there is a biological specialization for phonetic gestures, which prevents listeners from hearing the signal as ordinary sound. So, we hear speech signal as linguistic and we always map it to something that is linguistic and has to be perceived in a certain way, because it is the articulatory gestures are encoded in our perceptual abilities in such a way that they are the units of perception. So, there is a special relation between signal and sound to perceive the gestures.

(Refer Slide Time: 31:46)



And motor theory commands that control articulation were considered to be the invariant phonetic features. And the revised theory says that, that is not it is intended gestures that are the invariant object of perception. But regardless, we have to remember that the gestures the articulatory gestures are the object of perception, the gestures are the units of perception, that is what the motor theory tells us.

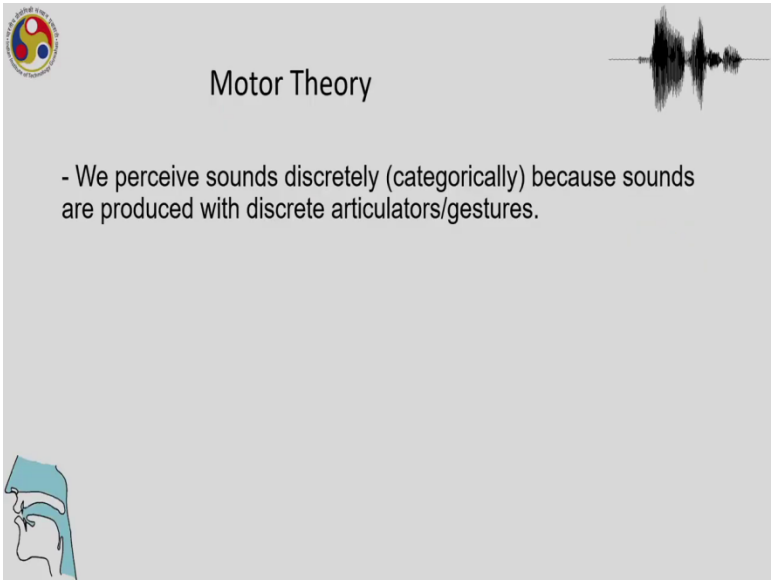
(Refer Slide Time: 32:15)



And these are the various stages involved in motor theory, there is an internal representation, and there is a processing which is involved there, and then this is the received message. So, we have various things in the articulation the articulatory pattern, articulatory hypothesis, and then these are extracted for articulatory movements.

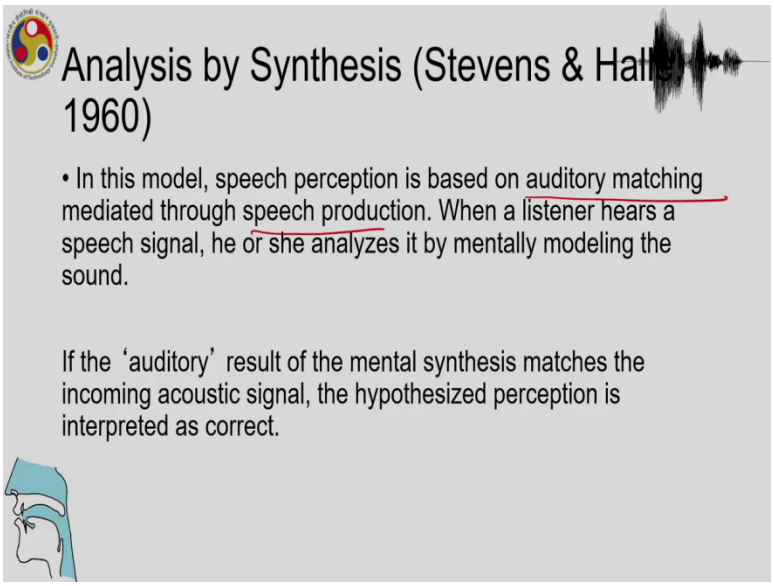
And then this generation of hypothesis of the sound ones based on the neuromuscular knowledge, this is the active processing which is model and then the reconnection the connection the perceived units into larger units of symbol of syllable, the syllables in words etc happen in the motor theory. This is the postulation of speech, this is the postulation of motor theory with regard to speech perception.

(Refer Slide Time: 33:07)



So, we perceive according to motor theory with speech, perceived speech discretely categorically because sounds are produced with discrete articulators and gestures.

(Refer Slide Time: 33:18)

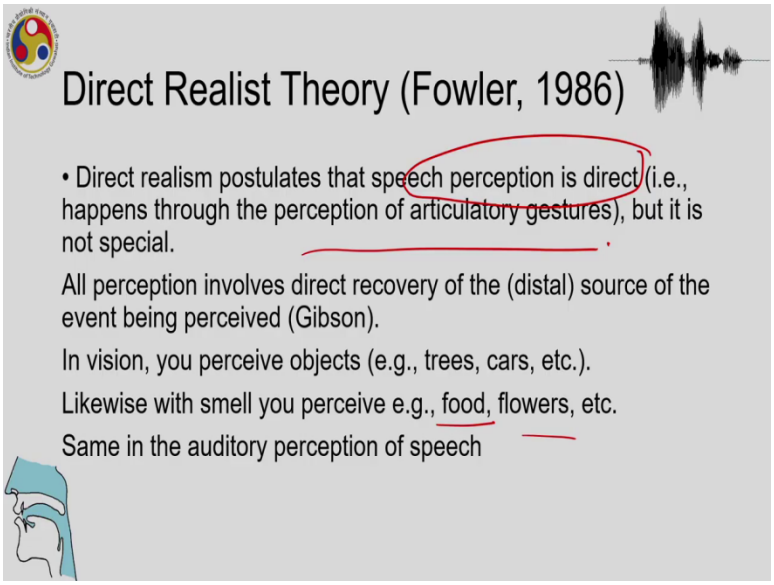


In this model, speech perception is based on auditory matching mediated through speech production, when a listener hears a speech signal, he or she analyzes it by mentally modeling the sound. If the auditory result of the of the mental synthesis matches the incoming acoustic signal,

the hypothesized perception is interpreted as correct. So, analysis by synthesis is therefore different from motor theory.

In this models, the auditory matching is true speech production. So, this is one of the auditory audition theories where the auditory part is important. So, when the listener hears a speech signal, he or she analyzes it by mentally modeling the sound. So, when you hear something, you model it and if the auditory result of the mental synthesis matches the incoming acoustic signal, the perception is interpreted is correct. So, a lot of emphasis is put on the auditory part of speech perception.

(Refer Slide Time: 34:19)

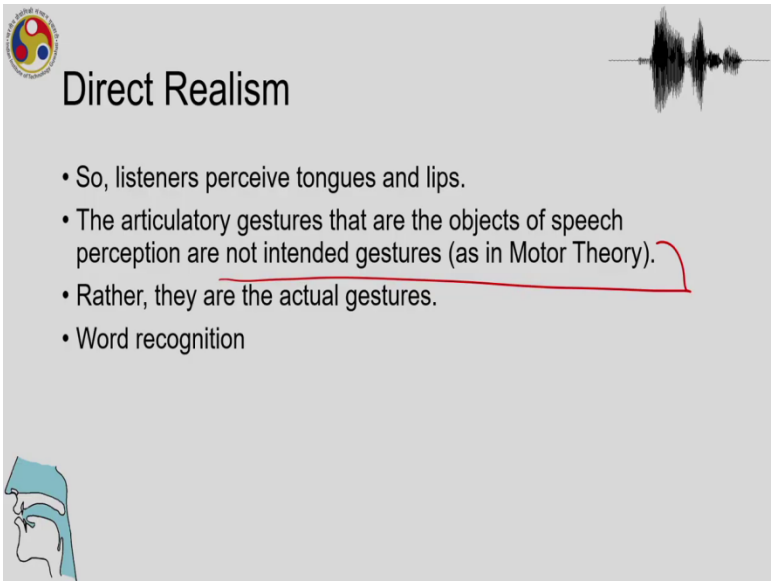


And then we have the direct realist theory. So, direct realism postulates that speech perception is direct that is, happens through the perception of articulatory gestures, but it is not special. All perception involves direct recovery of the distal source of the event being perceived. And in vision, you perceive objects likewise with smell you perceive food, flowers, etc. And same in the auditory perception of speech.

So, in direct realism theory, speech perception is direct. What does it mean? It happens through perception of articulator gestures, but the gestures are not special because there is no intervening process there. So, it is a sensory perception. So, like food flowers, if you get the smell, you

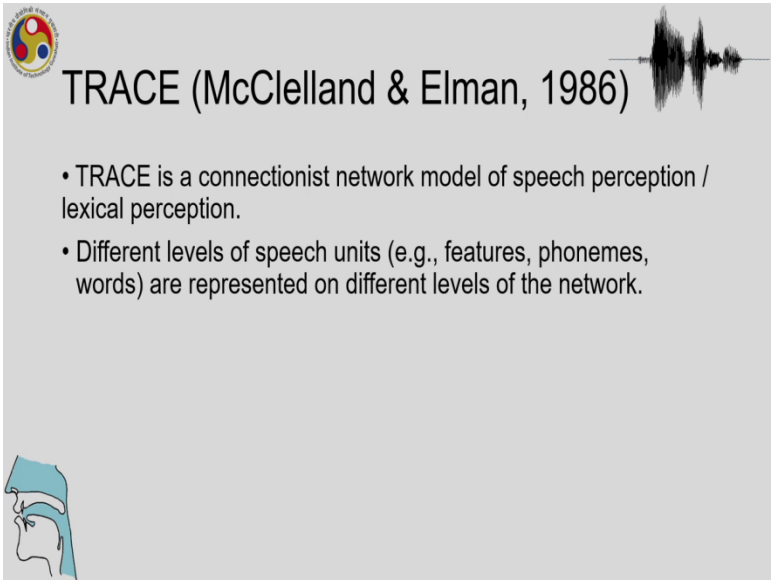
perceive that is food you get the smell you perceive that flowers. And the same thing is with auditory perception of speech.

(Refer Slide Time: 35:17)



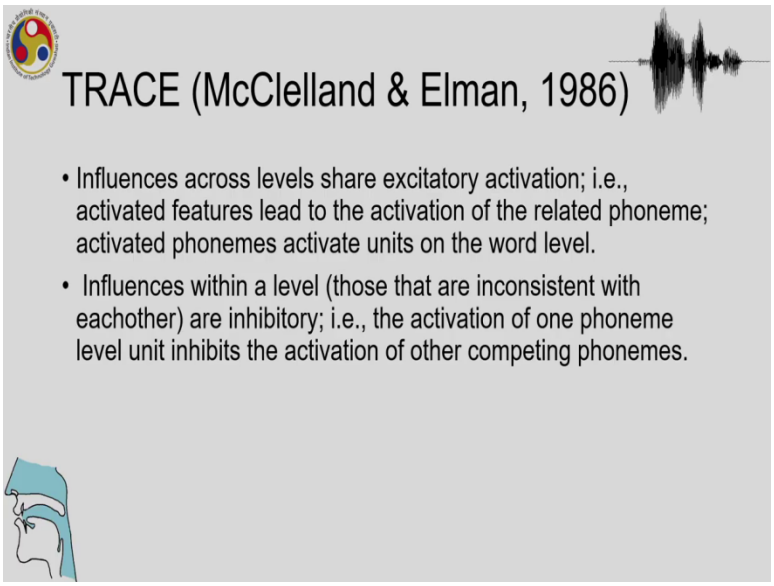
So, listeners perceive tongues and lips and the articulatory gestures that are the objects of speech perception are not intended gestures. So, unlike motor theory, here indirect realism theory, the articulatory gestures are where the articulatory gestures or objects of perception here it is not. So, here, it is not about the gestures, it is about tongue, lip, etc, the different parts of our vocal abilities to produce sound, those are perceived, and that is how we perceive sounds. Rather they are the actual gestures so and that is how direct realism is understood.

(Refer Slide Time: 36:01)



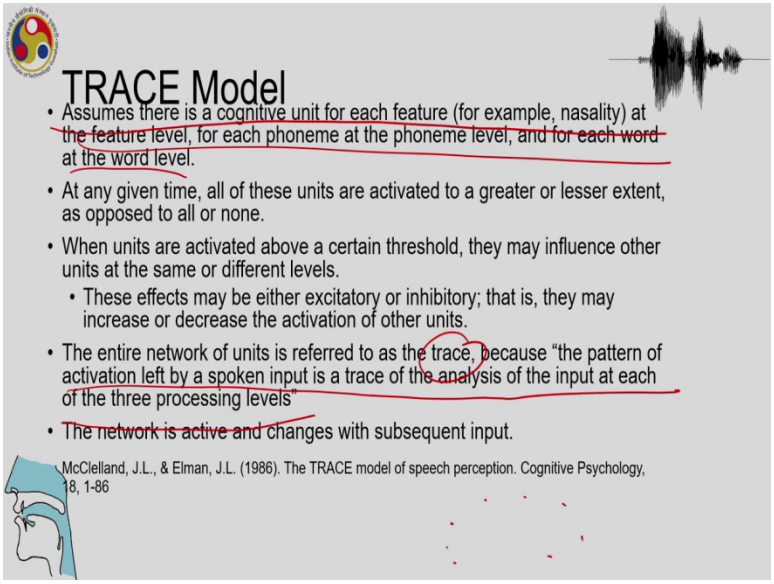
So, a trace model it is a connectionist network model of speech perception lexical perception, and different levels of speech units are represented on different levels of the network.

(Refer Slide Time: 36:10)



So, it influences across levels and shares excitatory activation that is activated features lead to the activation of the related phoneme, and activated phonemes activate units on the word level and influences within a level those that are inconsistent with each other are inhibitory. That is the activation of one phoneme level unit inhibits activation of other competing phonemes.

(Refer Slide Time: 36:40)

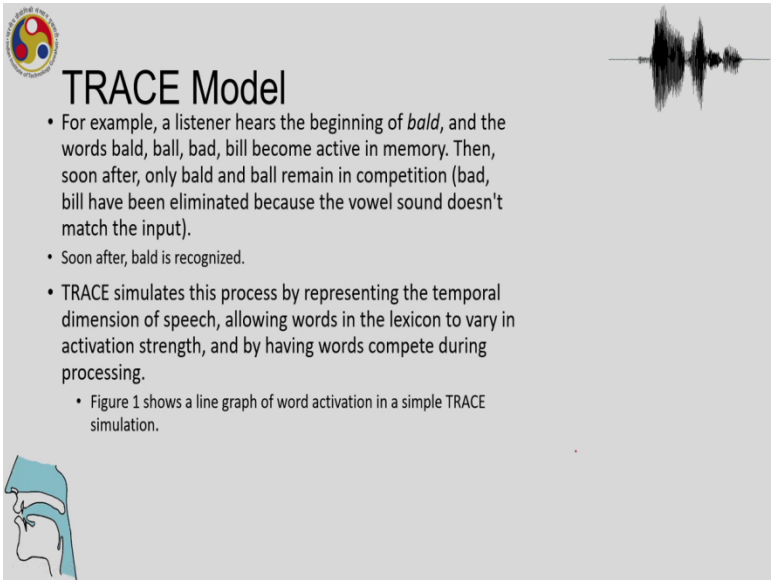


And Trace model assumes there is a cognitive unit for each feature. Each feature, for example, nasality at the feature level for each phoneme at the phoneme level. And for each word at the word level. At any given time, all of these units are activated to a greater or lesser extent, as opposed to all or none. When you need to activate it above a certain threshold, they may influence other units at the same time.

And these effects may be either excitatory or inhibitory. And the entire network of units is referred to as trace, because the pattern of activation that is left by spoken unit is a trace of the analysis of the input at each of the three processing levels. So, there are three processing levels. But the spoken input is the trace, it leaves a trace in the three levels. And that is why it is called the trace model.

So, the cognitive unit for each feature and at the feature level for each phoneme and the phoneme level. And for each word at the word level. So, this cognitive unit, and the entire, the network of units that we have in trace model. So, it is a trace, basically and that is why the cognitive unit is a trace and that is why it is called a trace model.

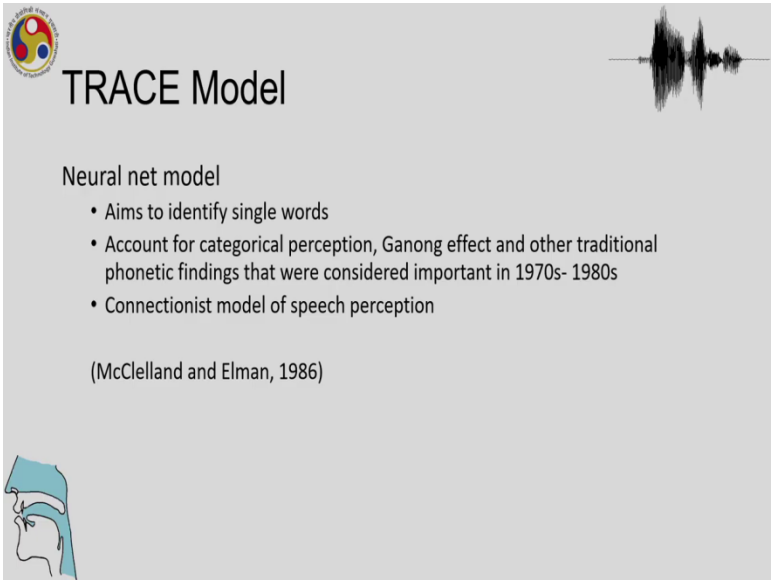
(Refer Slide Time: 38:05)



So, for example, a listener hears the beginning of bald and the bald and the words bald, ball, bad, bill become active and memory, then soon after only bald and ball remain in competition and bad and bill have been eliminated, eliminated because vowel sound does not match the input soon after bald is recognized. And therefore trace simulates this process by representing the temporal dimension of speech allowing words in the lexicon to vary in activation strength and by having words compete during the processing.

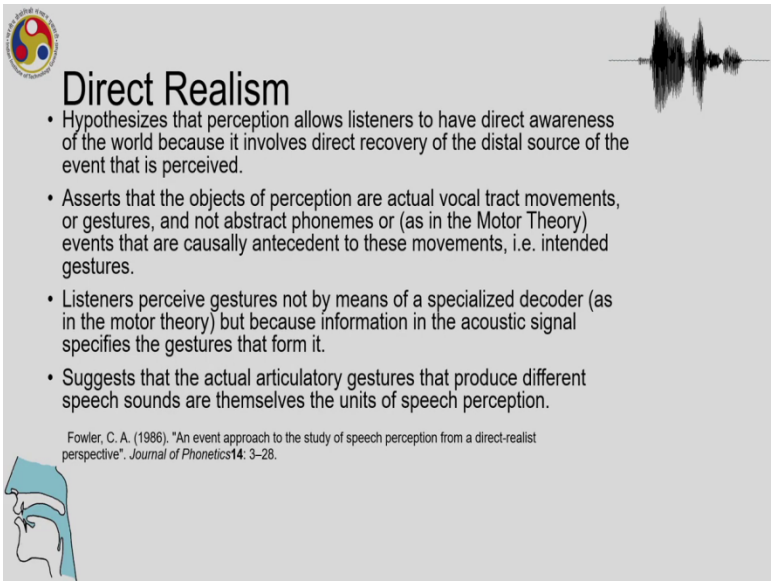
So, there is activation happening at each level and whatever is activated is basically understood as one of the words which might be the, the resultant output of the system, because the activation strength, the possibility of each word, resulting as the output or resulting as a word which will be perceived will be different.

(Refer Slide Time: 39:02)



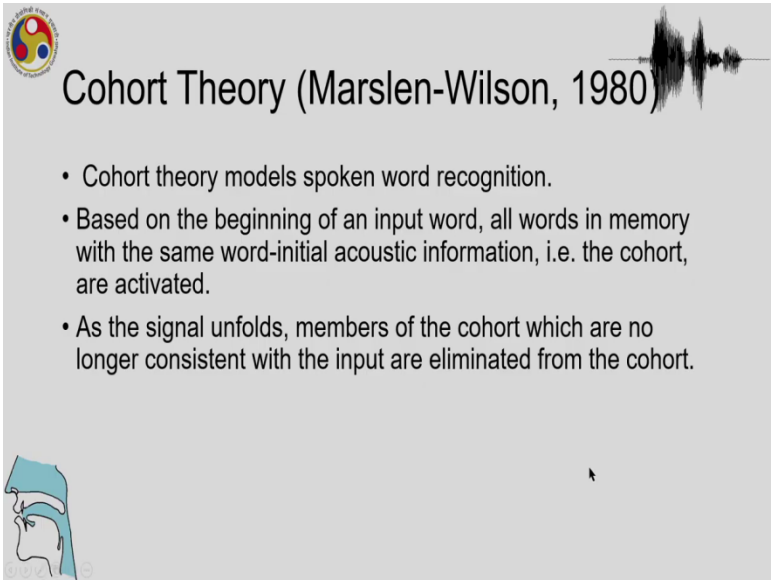
So, therefore, trace model is a neural net model. And it aims to identify or an activate words, and its accounts for categorical perception and Ganong effect and other traditional phonetic findings that were considered important in the 1970s. So, it is a connectionist model of speech perception.

(Refer Slide Time: 39:21)



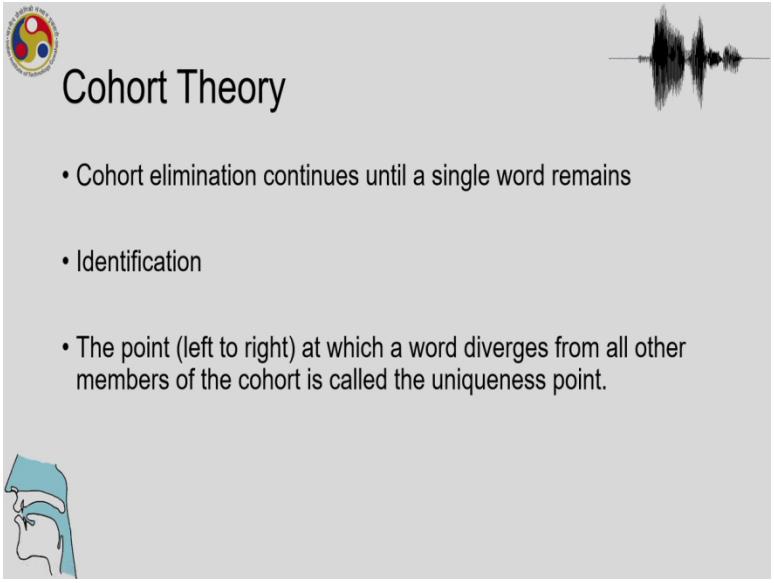
Unlike direct realism, so we have already looked at direct realism.

(Refer Slide Time: 39:26)



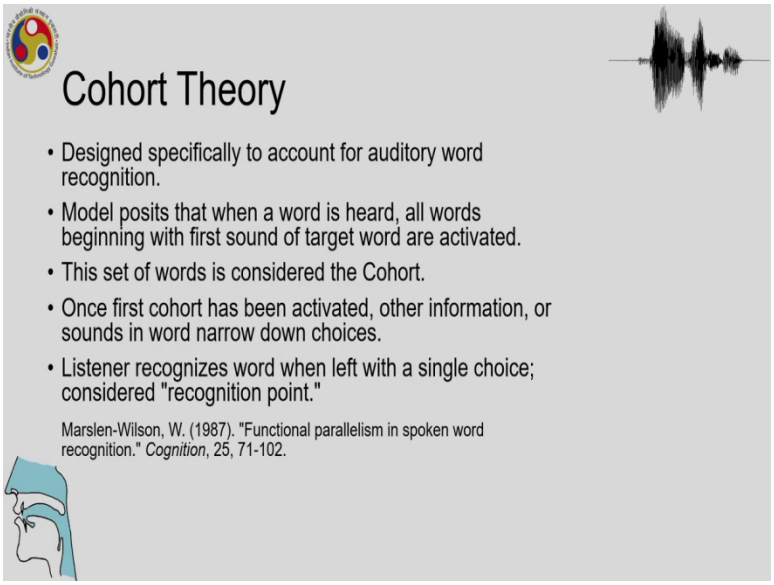
So, we will look at cohort theory a bit now. A cohort theory models, spoken word recognition, it is based on the beginning of an input word and all words in memory with the same word ensure acoustic information, that is the cohort are activated. So, there is a cohort which is activated and as the signal unfolds members for cohort which are no longer consistent with the input are eliminated.

(Refer Slide Time: 39:57)



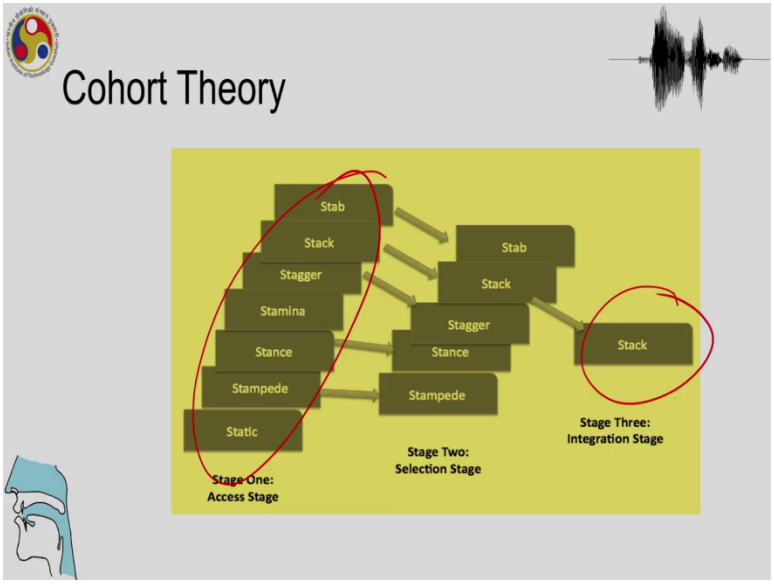
So, cohort elimination continues until a single word remains, and the point at which a word diverges from all other members of the cohort is called the uniqueness point.

(Refer Slide Time: 40:05)



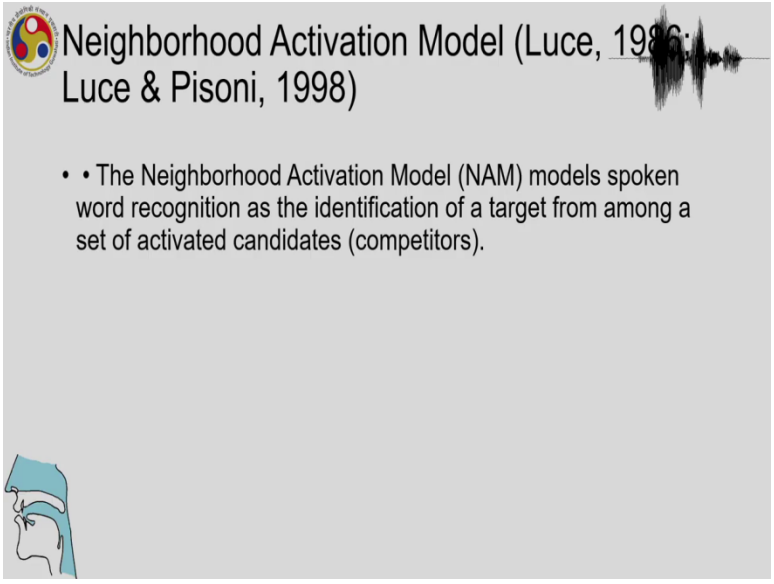
And cohort theories was designed specifically to account for auditory word recognition. And model posits that when a word is heard, all words beginning with the first sound of target words are activated. And this set of words is considered the cohort. Once our first cohort has been activated other information or sounds in the word narrowed down, are, choices are narrowed down. So, listener recognizes word left with a single choice that is called the recognition point.

(Refer Slide Time: 40:41)



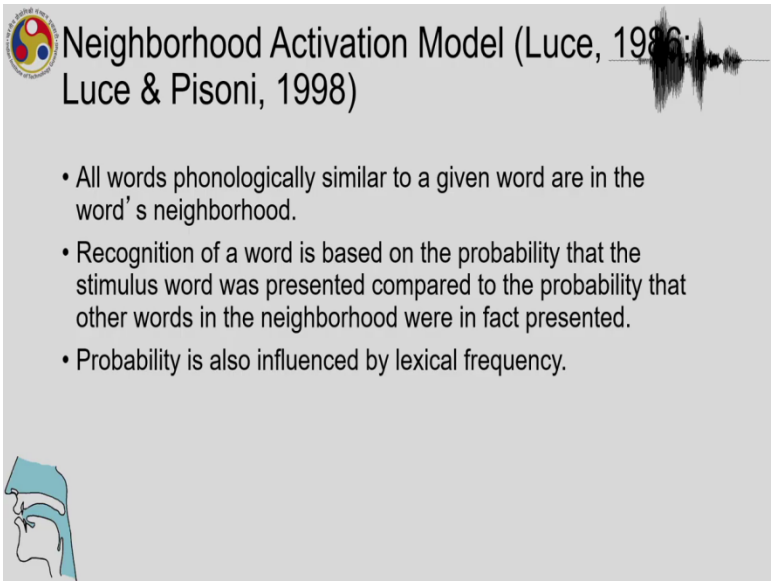
So, this is a cohort theory. So, in stage one, suppose this was the word stack. So, all these words will be activated. So similar to stack, static, stack, stab, stagger, stamina stampede, and then there might be elimination of some words because they are different. So, stab and stack are pretty close. And stamina is now excluded, because it is very far from stack, and then stampede and all these are better matches to stack and all these are activated and finally, they are all eliminated to the fact that the one which is perceived is remaining.

(Refer Slide Times: 41:32)



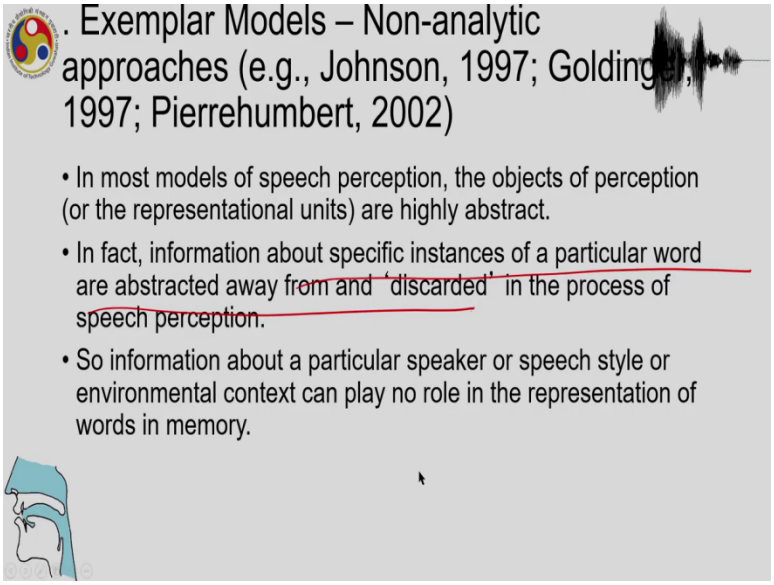
And that is the elimination point. Similar also to cohort and in similar models, where different words are activated is the neighborhood activation model. And the neighborhood activation models spoken word recognition as the identification of a target from among a set of activated candidates.

(Refer Slide Time: 41:50)



So, these are the competitors. And all words of knowledge very similar to a given word are in the words neighborhood and recognition of a word is based on the probability that the stimulus word was presented compared to the probability that other words in the neighborhood were in fact also presented. So, probability is also influenced by lexical frequency, and that is the neighborhood activation model.

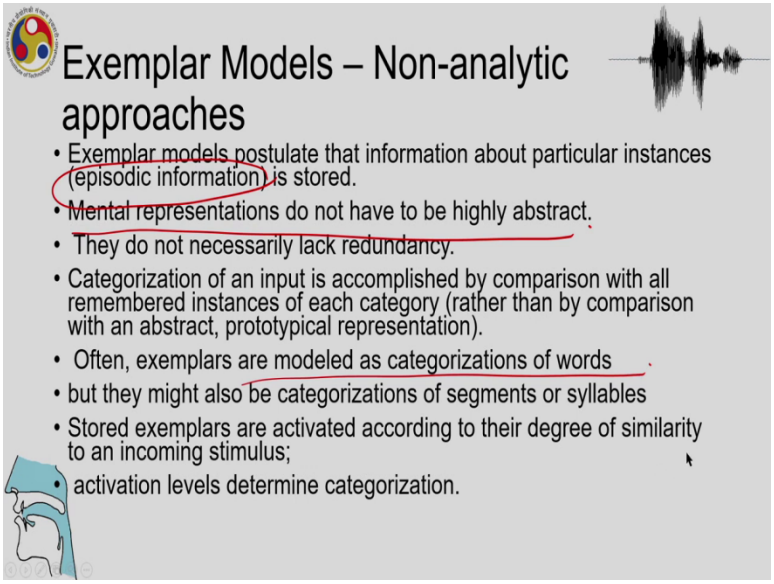
(Refer Slide Time: 42:14)



And then there are other exemplar models like non analytic approaches. In most of these models speech perception of speech perception, the objects of perception are highly abstract and in fact, information about specific instances of a particular word are abstracted away from and discarded in the process of speech, perception. So, information about a particular speaker or speech style or environmental context can play no role in the representation of words in memory.

So, information about particular speaker speech style does not play a role according to exemplar models. And it is a very highly abstract sort of model where information of the specific instances are abstracted away and discarded in the process of speech perception. And only the abstract information remains. So, that is why it is called an exemplar model.

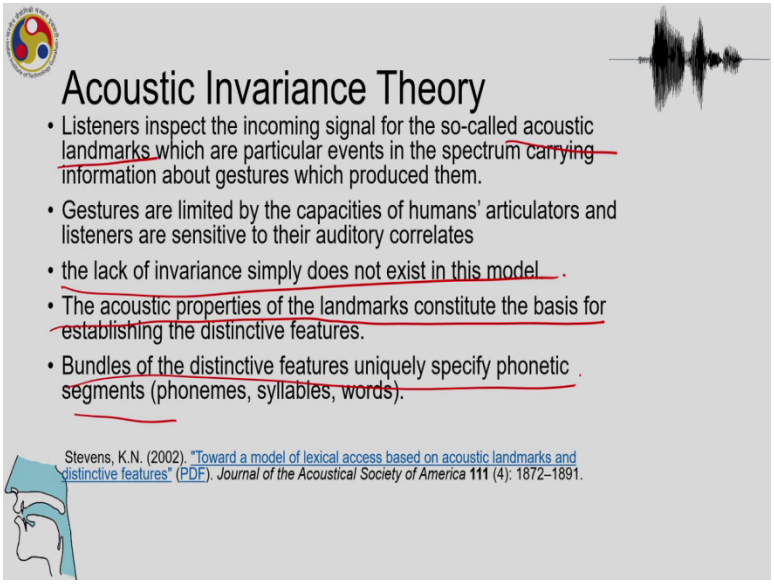
(Refer Slide Time: 43:14)



And exemplar models postulate that information about particular instances, the episode information is stored. And these are mental representations that do not have to be highly abstract, but they do not necessarily lack redundancy and the categorization of an input is accomplished by comparison of all remembered instances of each category.

And often exemplars are more or less categorizations of words, but they might also be categorizations of segments, or syllables. Stored exemplars are activated according to the degree of similarity and activation levels determine categorization.

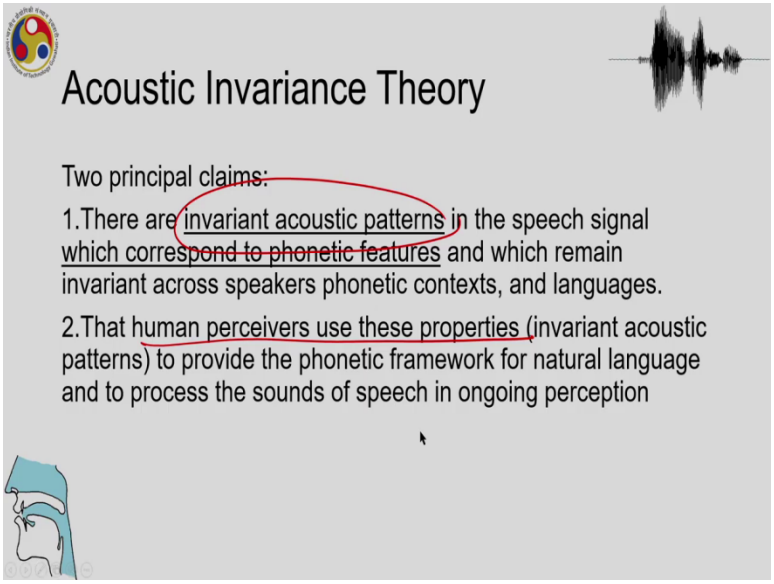
(Refer Slide Time: 43:48)



So, we have acoustic invariance theory where listeners inspect the incoming signal for the so called acoustic landmarks, which are particular events in the spectrum carrying information about gestures which produce them. And gestures are limited by the capacities of human articulators and listeners are sensitive to their auditory correlates. The lack of invariance simply does not exist in the model, and the acoustic properties of the landmarks constitute the basis for establishing the distinctive features.

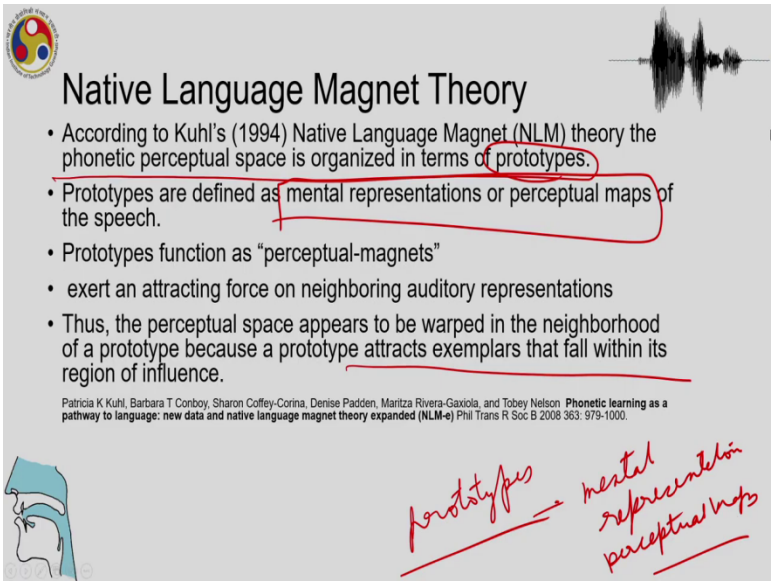
And very importantly, the acoustic properties are the most important here the acoustic invariance theory, and there is no lack of invariance in this model. And bundles of distinctive features uniquely specify phonetic segments to the incoming acoustic signal has acoustic landmarks which the listener is listening to. And that is what the acoustic invariance theory tells us.

(Refer Slide Time: 44:55)



The two principal claims of the acoustic invariance theories that they are invariant acoustic patterns in the speech signal, which correspond to phonetic features, and that humans perceive these properties. These acoustic invariant acoustic patterns are perceived by humans to provide a phonetic framework for natural language and to process the sounds of speech in ongoing perception.

(Refer Slide Time: 45:19)

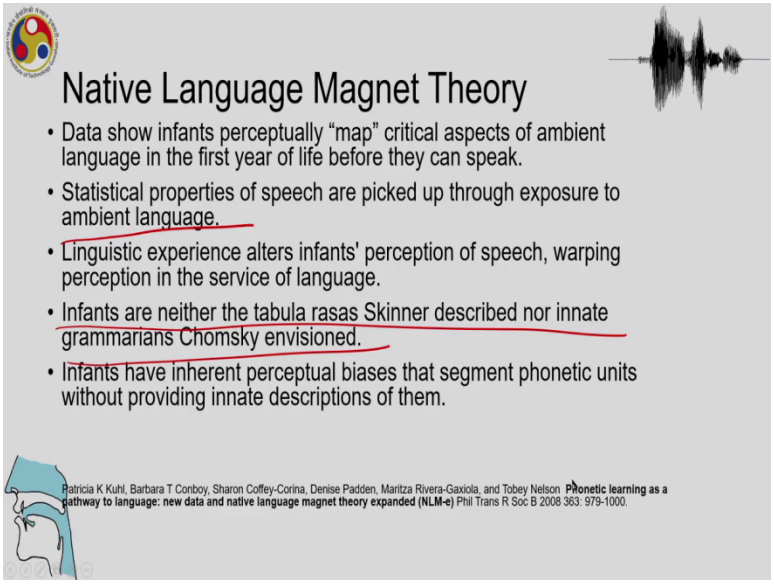


Finally, we have the native language magnet theory. And according to Kuhl's native language magnets theory, the phonetic perceptual space is organized in terms of prototypes. Prototypes are defined as mental representations or perceptual maps of the speech, and prototypes functioned as perceptual magnets and exert in attracting force on neighboring auditory representations. Thus, the perceptual space appears to be warped in the neighborhood of a prototype.

Because a prototype tracks exemplars remember the exemplar theory, where we have different exemplars. Now abstract features that fall within its region of influence. So, the perceptual the phonetic perceptual space is organized according to prototypes. Prototypes is extremely important in native language, magnet theory, and prototypes are defined as mental representations or of all perceptual maps of the speech.

So, what are prototypes? These are mental representations or perceptual maps. So, and that is how we create prototypes of our native languages.

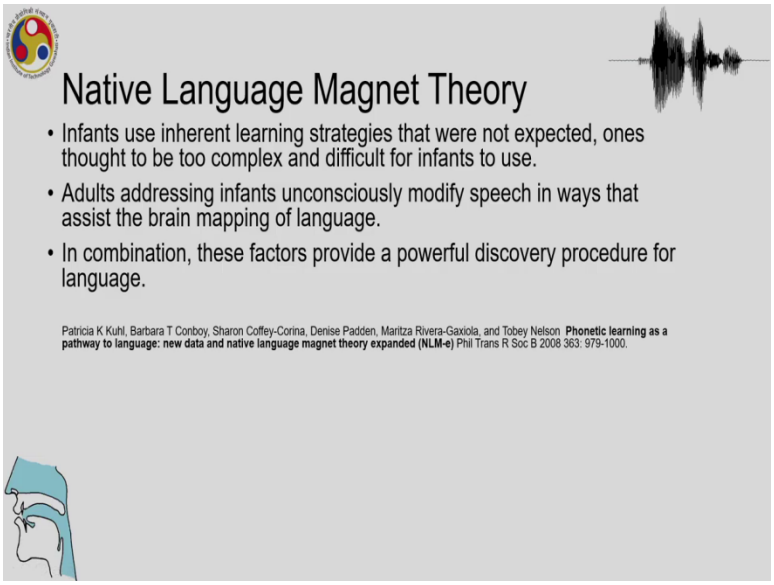
(Refer Slide Time: 46:51)



Data shows that infants perceptually map critical aspects of ambient language in the first year of life before they can speak and statistical properties of speech are picked up through exposure to the ambient language, language in the child is growing up.

And then linguistic experience alters infants perception of speech, warping perception in the service of language. And infants are neither the tabula rasas of Skinner describe nor in innate grammarians that of Chomsky, but they create these perceptual maps infants have inherent perceptual biases that segment phonetic units without providing innate descriptions of them.

(Refer Slide Time: 47:39)



So, infants use inherent learning strategies that were not expected once and they are thought to be too complex and difficult for infants to use, adults addressing infants can unconsciously modify speech. So, in infant directed speech and child directed speech, those there is a lot of unconscious modification in ways that assist the brain to map the language. So, the way child directed speech is spoken is it in such a way that helps the child to map the language.

In combination these factors provide a powerful discovery procedure for language. So, we come to the end of our lecture on speech perception in these lectures, we try to only give you a broad overview of speech perception, we will not able to go into the into the details of for instance, MDS or perceptual confused ability, multi dimension of an analysis to the help of multi dimensional scaling, because this course is gives you a very broad overview of these aspects and, and they have to be.

So, if you are interested, then you are very welcome to follow whatever was discussed very briefly here, you can consult the book by Keith Johnson, which is there in our list of readings,

etc, you can follow those and look into how those things can be actually implemented. Even though we did not go into a lot of detail of multidimensional scaling.

There is always all those things to follow up if you are interested in greater detail, this course, this lecture especially, was giving you a very broad idea about the issues in speech perception and the ways of analyzing problems in speech perception. And also in this lecture, we gave you a very broad overview of different types of theories of speech perception available like motor theory, like direct realism trace theory, native magnet theory or exemplar models, etc.

Telling you how different speech possession theories try to accommodate all these complexities, which are there that that speech is continuous and yet it is perceived in discrete ways, and how to try to understand or analyze or give us an idea of what must be happening in our perceptual spaces, and how we must be achieving the task of perceiving speech all the time, that is speech.

So, and therefore, this lecture was also a very broad overview of the problems in speech perception and the different theories, which are there. And, again, this is an overview, this was not a very specialized lecture on either of these two issues, and you are very welcome to follow up if you are interested in this area.

And thank you for listening. We will continue with phonology from the next module. And this brings us to the end of articulatory phonetics, acoustic phonetics and perception and we will again continue with where we left off with phonemes. In the in the first lecture, we will start again with phonemes as in the history of phonology, etc from the next module. Thank you very much for your attention.