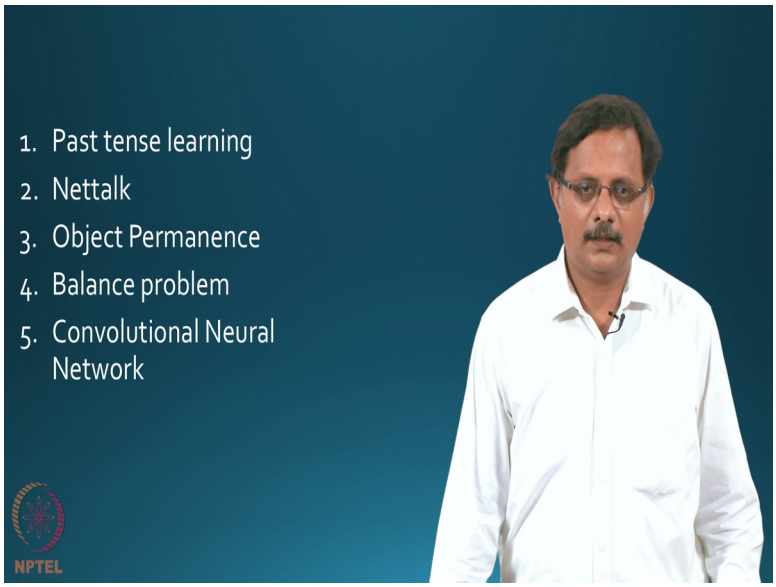
Demystifying the Brain
Prof. V Srinivasa Chakravarthy
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture – 08
Multilayer Perceptrons: Applications in Psychology and Neuroscience

So, in the last segment of this lecture we looked at the Multilayer Perceptrons. We looked at the learning algorithm the back propag propagation algorithm. We saw how perceptrons could only solve linearly separable classes whereas, multilayer perceptrons are free from that kind of a short coming they have this syllabus approximation property they can learn practically any real world function.

So, in this segment we will see how this multilayer perceptrons can be applied to problems in psychology and also neuroscience?

(Refer Slide Time: 00:45)



So, spac specifically we look at 5 phenomena. So, these are like past tense learning that tells you how we how children learn past tense, then net of difference to a how a network can be taught to read test aloud third one refers to the pro phenomena of object permanence and how children learn this concept of object permanence.

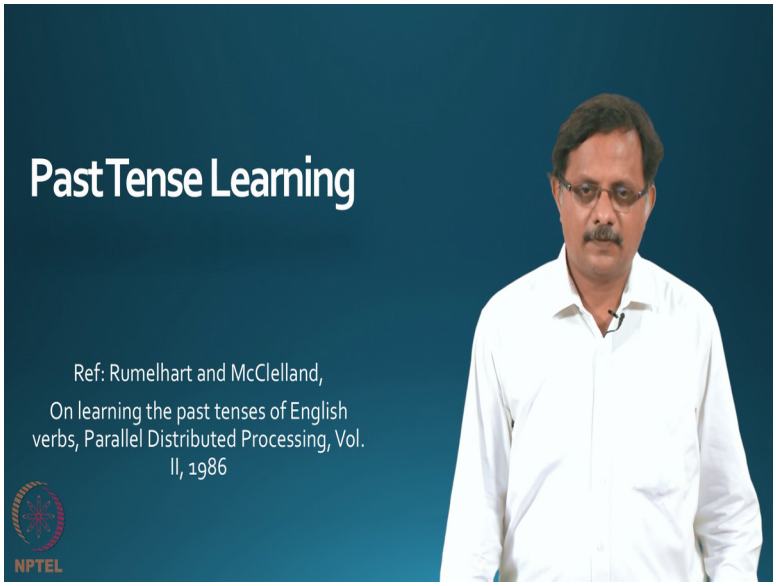
The fourth one is about balance; how the children learn the concept of balance and fifth one is a network model called convolutional neural network. Out of these 5 phenomena

the first four refer to more psychology, a particularly phenomena 1, 3 and 4 are from child psychology. The second is a behavioral task and fifth one can be related to neuroscience. So, now, it can be taken you can compare to certain phenomena that is that are observed in the visual system of the brain. So, first 1, 3 and 4 come from child psychology lot of pioneering work and child psychology was done by Jean Piaget was a Swiss psychologist and Paget main philosophy is that when children learn in various domains they cognitive motor and so on

The learning progresses in distinct stages and this his work was later taken up in the 60s and became very popular in the 60s they it was kind of rediscovered in 60s with many psychologists and who developed his ideas for the and have delineated these theory these stages that are in described by Piaget.

So, in the 80s when connections evolution has taken place people have applied the kind of neural network models that, we are studying to some other phenomena that Piaget and others have studied and have shown how this networks can explain some of the stages that Piaget and others have studied.

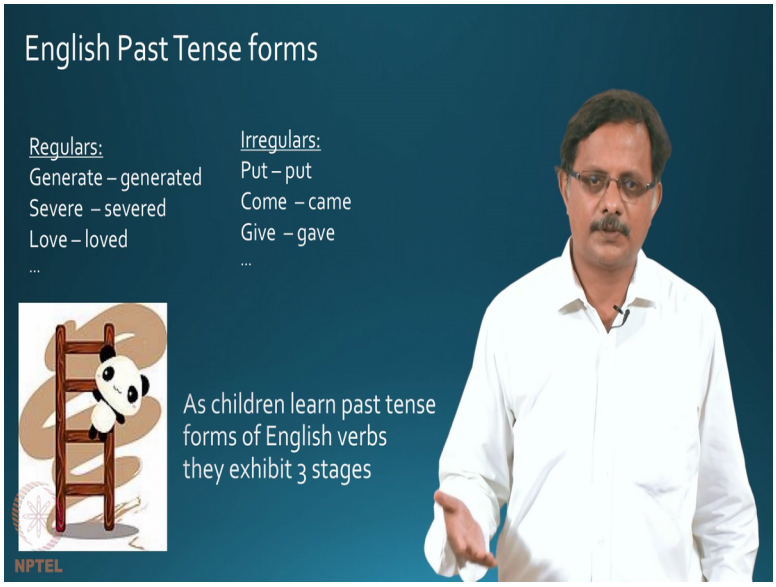
(Refer Slide Time: 02:32)



So, among these examples the first example we will see is past tense learning that is how do children learn past tense? So, learning past tense of English is actually nontrivial if you really have not thought about it much. so, because in past tense there are two kinds

of forms this. So, there are these regular verbs. So, for example, you have generate and generated we have severe and severed or love and loved and so on so forth.

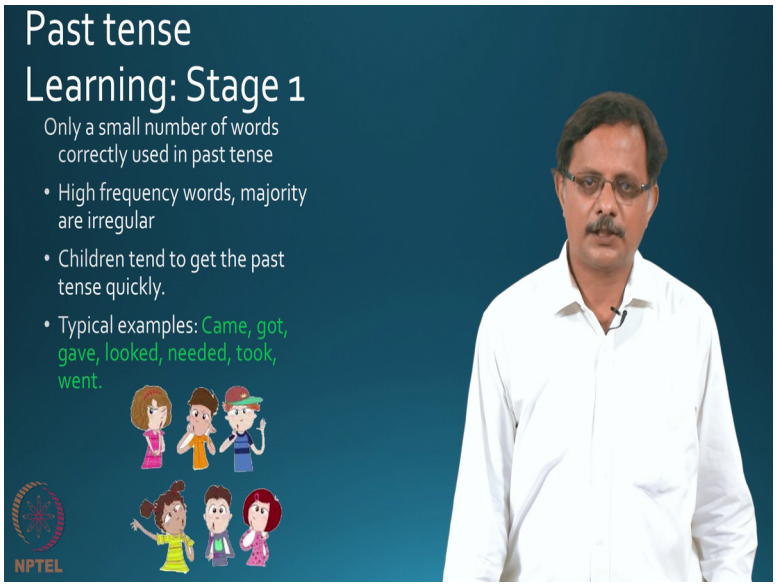
(Refer Slide Time: 02:48)



So, in ka this kind of a regular past tense form they simply add either d or e d to the verb there is lot of other examples of past tense do not fall under this general pattern these are called the irregulars say for example, put and put a come and came and give and gave and so on and so forth. So, past tenses come in two forms a regular and irregular.

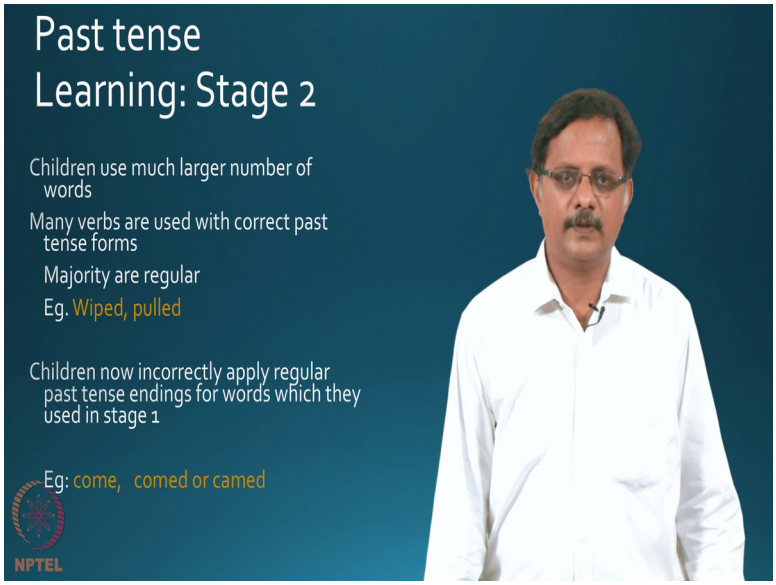
Now, as children learn past tense of English verbs it has been found that they exhibit three stages.

(Refer Slide Time: 03:26)



So, in stage one children expose to only a small number of verbs their vocabulary is very limited. So, they have some high frequency verbs words which are mostly irregular and these they learn very quickly and then there is there is small number of you know words which are regular. So, for example, come and came and get and got and give and gave look need needed take took and go and went. So, children almost memorize; this mappings between the present tense form and a past tense form and they learn these things very quickly.

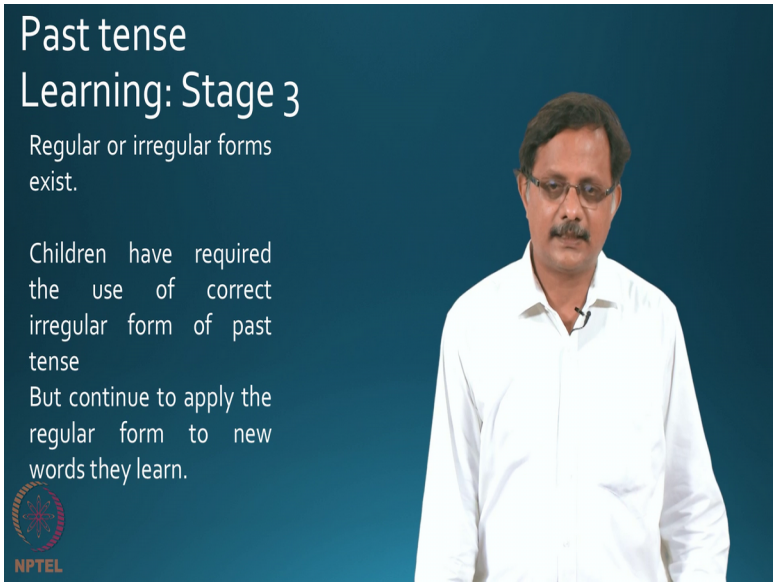
(Refer Slide Time: 03:58)



Then comes stage two where they are exposed to much larger sort of words and it turns out that a majority of the words that they are exposed in second stage are ire regular. So, for example, wipe and wiped and pull and pulled. So, they have this standard d or ed ending. So, children begin to learn this rule this new rule this pattern that is. So, that underlies a lot of past tense forms, but then they also make start making mistakes on verbs that they have learned before in the in the stage one. So, they begin to wrongly apply this past tense rule of d or ed ending to verbs which they have learnt in the previous stage.

So, for example, for come they might say they might wrongly use the past of comed or came, but when you when they go to stage.

(Refer Slide Time: 04:47)



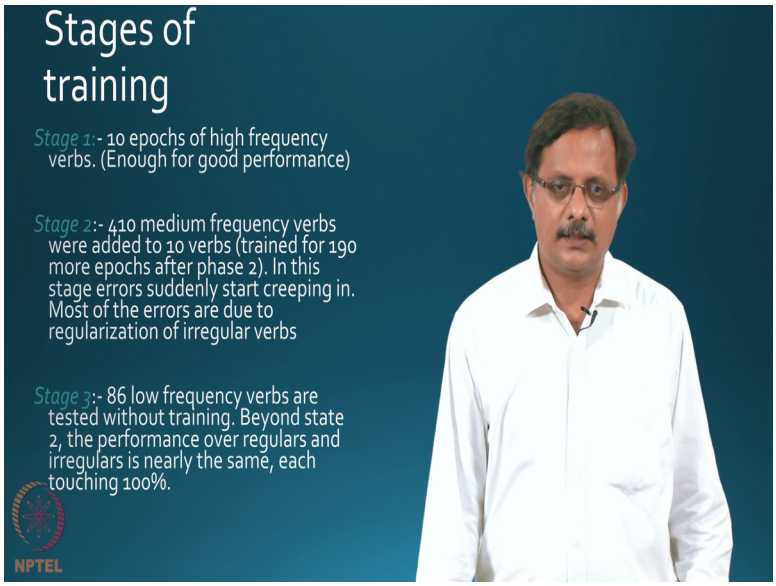
Three then the ca the error of stage 2 is corrected because you have large number of both regular and irregular forms are encountered by children. In stage three and now they have a acquired the use of correct regular form of past tense, but they also can know to apply the correct regular form to the new words that they learn.

(Refer Slide Time: 05:11)

Now, a network can be trained and people have shown that the network also exhibits with all these three stages as the network is strained. So, to do this experiment they have used that as they train the network in three different stages. In stage one as a small number of high frequency verbs are used for training the only 10 of them and among them 8 are irregular forms 2 are regular forms the words are like you know come get give look and so and so forth.

Then in the second stage they used for in 10 medium frequency verbs among them 3 and 34 are regular verb forms and 76 are irregular forms. And the last stage it was 86 low frequency words among them 72 are regular and 14 are irregular. So, how you train the network?

(Refer Slide Time: 06:02)

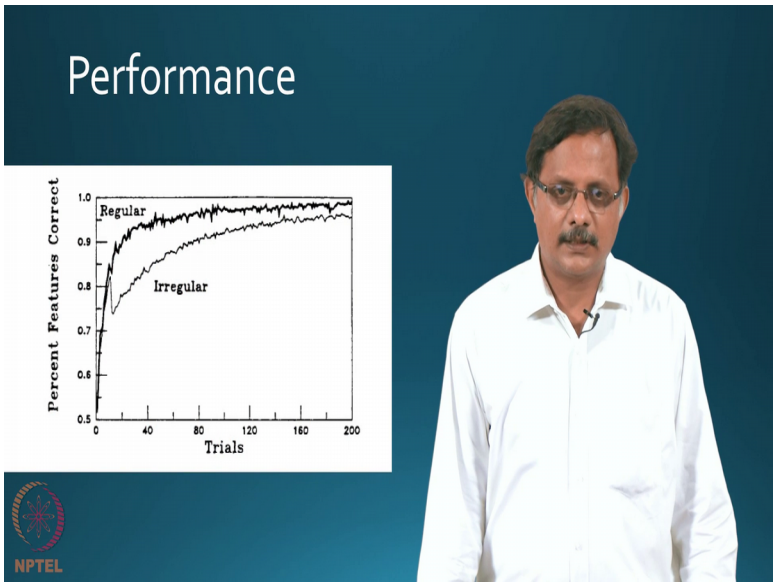


In stage 1 at 10 epochs of high frequency verbs were used. So, epoch consist of 1 present tense or the entire set of patterns or verbs.

So, with 10 epochs the network is able to learn they perform the correct past tense form for all these words. In stage two for in 10 medium frequency verbs were added to the first set of 10 verbs and this was trained this stage the went on for 1 and 90 at more epochs after face 2 that is up to 200 epochs. So, in this stage certainly the error started creeping in.

So, the network started making mistakes on the irregular ver irregular verbs and start regularizing them that is it started giving supplying wrongly the ed or d ending to some of the irregular verbs in this stage in stage 386 low frequency verbs were tested without training and beyond stage two. The performance of regular and irregular stage nearly the same each touching. So, 100 percent that is once you go beyond 200 right, the performance on both regulars irregulars tend to be the same.

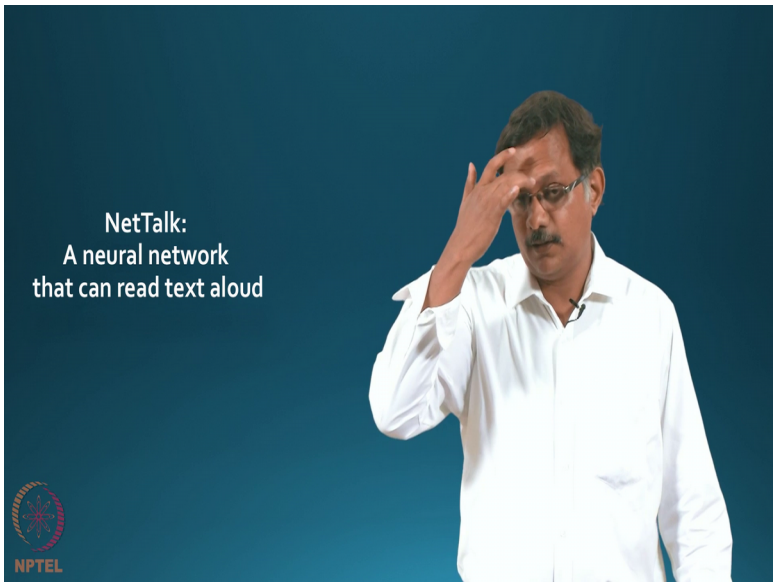
(Refer Slide Time: 07:12)



So, in this graph you can see the networks performance the two graphs within this plot one graph shows the performance on regular verbs and second on the irregular verbs. So, up to 10 epochs. So, network showed high same performance you know on both regular and irregulars right and beyond that when the new set of words are added in stage two, it suddenly started making lot of errors on the regular verbs and that is because it was wrongly supplying the regular ending d or ed to all the irregular verbs, but as training progressed and up to 200 by this time the performance on regular and irregular became almost similar.

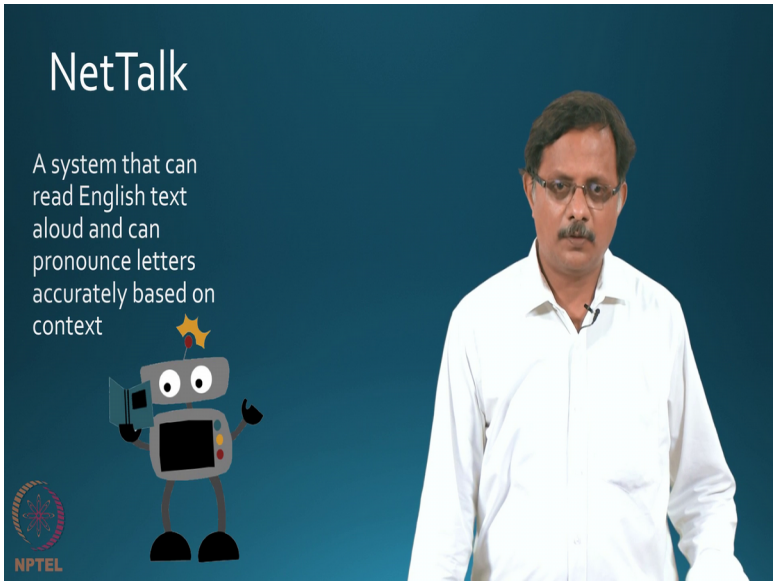
So, you see this general pattern of you know of learning past tense that is spread by the network and which is very similar to how children learn past tense and how the kind of mistakes they make at various stages as they learn past tense.

(Refer Slide Time: 08:08)



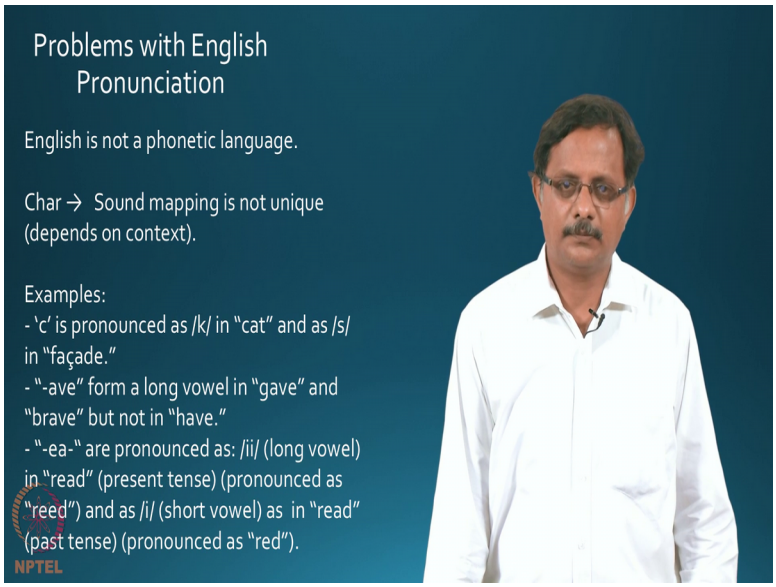
Now, interestingly this network that was used for this study is not even a multilayer perceptron it is only a perceptron that is there are no hidden layers. So, what is interesting here is? Even a simple network like that can be used to used to capture some of the patterns of learning; that are displayed by real humans. The second study we will talk about is what is called network this is a neural network that can read text aloud and although it is a very simple network model somehow the its learning patterns can be matched some of the learning patterns that you see in real neurons.

(Refer Slide Time: 08:41)



This study is (Refer Time: 08:41) performed by Sejnowski and Rosenberg in a paper published in 1986. So, the net talk is basically system that can read English text aloud. So, it can pronounce letters accurately based on context ok.

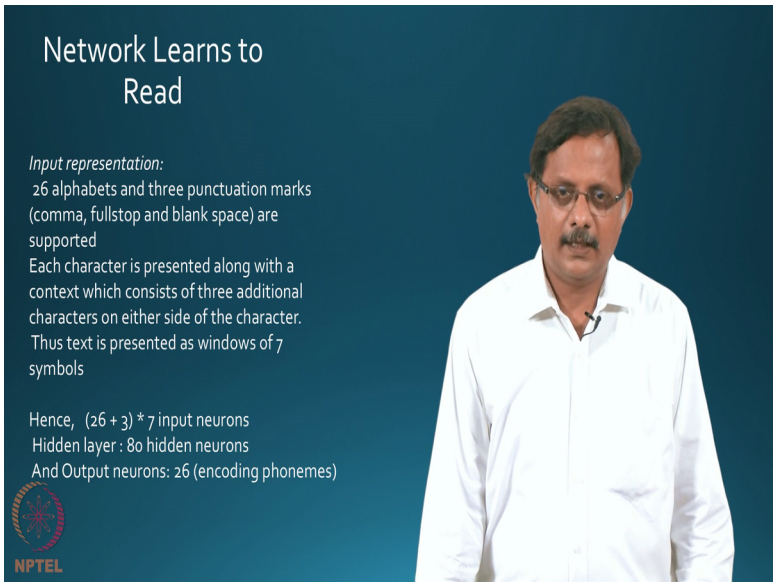
(Refer Slide Time: 08:56)



So, pronunciation part particularly in English is bit of a challenge, because pronunciation of as a mapping between the character and sound is non unique it depends on the context because English is not a phonetic language it is a alphabetical language and not an alpha celebic language for example, Indian languages are alpha celebic there is a much more unique mapping between character and the its pronunciation whereas, an English pronunciation of a character depends a lot on the contrast.

So, for example, c is pronounced as ka in cat and a sa in facade or if you look at the letters a, b, e that letter strink is pronounced as ave as an gave or brave, but as ave has an have like similarly the letters e pronounced as e or as an read or p as an read. So,. So, there is lot of difference from context to context.

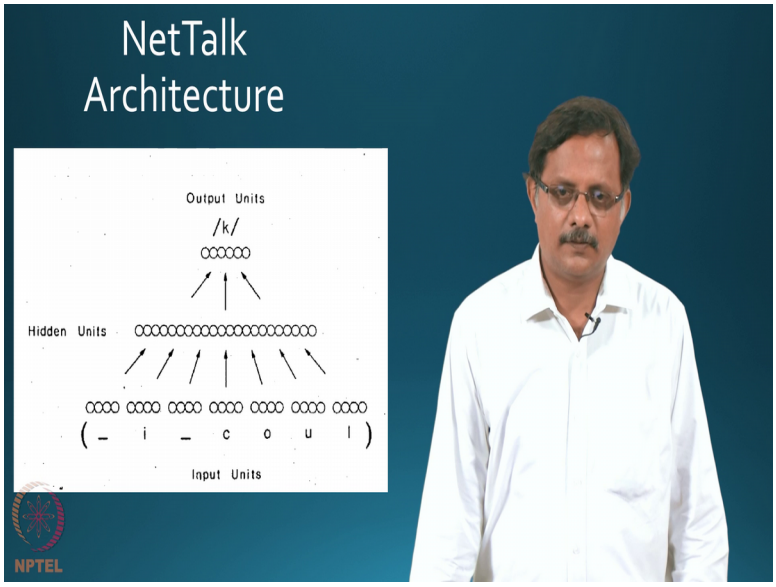
(Refer Slide Time: 09:51)



So, therefore, if you want to train a network to read you cannot give single letters as input and expect the network to reproduce the correct pronunciation of the phoneme you have to give a certain context of that letter.

So, how do you give a context? So, the simplest way is in addition to giving the letter you should also give certain neighboring letters. So, that is at the context for that letter.

(Refer Slide Time: 10:14)

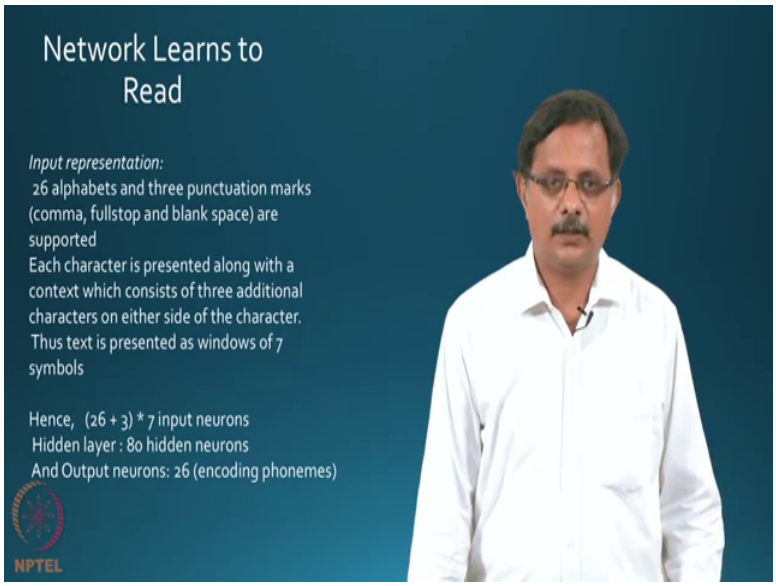


So, this picture shows the architecture of the network. So, you have the input layer a hidden layer and output layer.

So, the in the input layer you gave at any given time 6 characters. So, you have a running string of the text. So, you take a window of 7 characters and then there, there is central character is what you are interested you want you are trying to predict the pronunciation or the corresponding phoneme for the central character. So, in addition to central character you gave three characters on the right they had some characters on the left. So, this 6 characters form the context of the central character.

So, this window of input of text is given as input to the input layer and which is presented to hidden layer and then you get a presented to the output layer where the phoneme is recognized. So,.

(Refer Slide Time: 10:55)



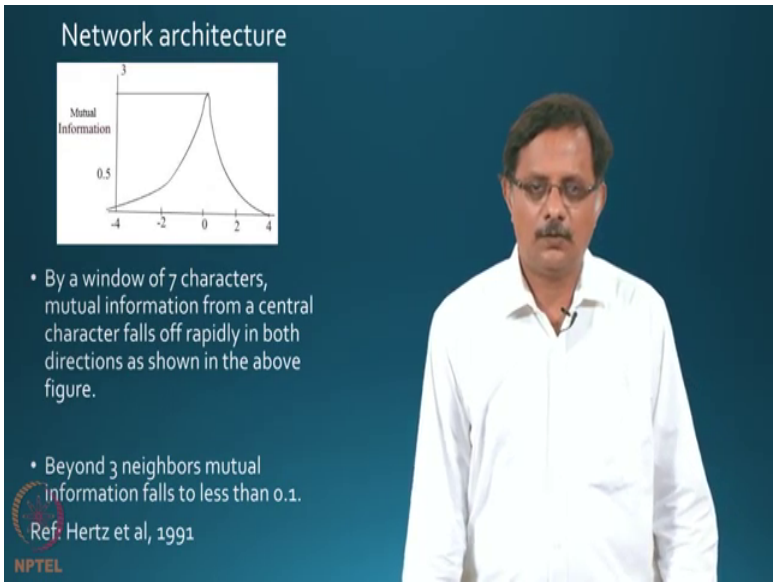
So, there are and each input character is represented by a set of 29 bits, because the network can understand all the 26 in English letters and three punctuation that is full stop comma and the blank space.

So, at any given point there are 29 characters of symbols are possible. So, therefore, the character is represented as a 29 dimensional binary vector with a corresponding bit set to one and all other bit set to 0 right; and then there are 7 input characters. So, the inputs consist of totally 29 times 7 right a matrix of 29 by 7 there are 80 hidden neurons and 26 output neurons and this 26 outputs code for a phoneme.

So, for example, each of the output neurons code for some property of a phoneme for example, is it applausive or is it a fricative of is it a labials kind of sound for example, pa pha bha all these are labials right or this dentals you know and so on and so forth.

So, these properties of the phonemes are called in this 26 dimensional vector as output and the network is strained to map the pronunciation of the corresponding phoneme of the central character onto the output phoneme.

(Refer Slide Time: 12:07)

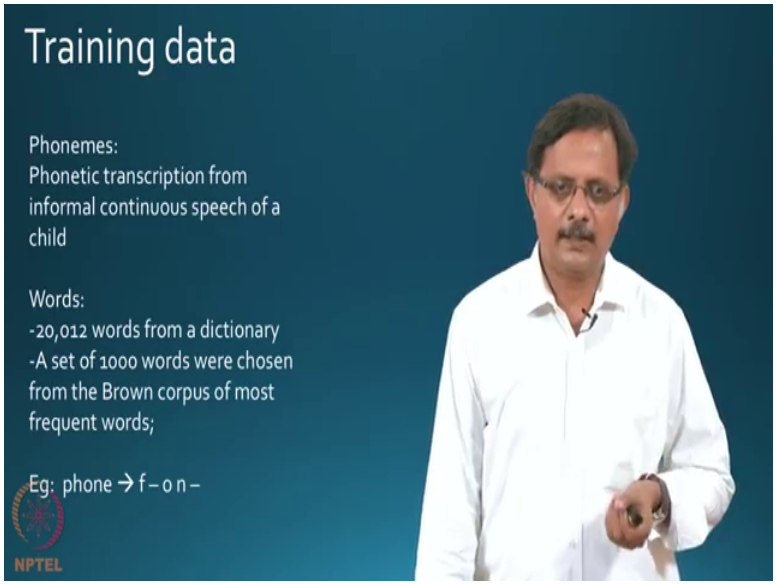


Now, one question is why is the context of 7 chosen y naught 17 or y naught 3 and so, on ok. So, the how do you define a context right?. So, if you take more you know bunch of characters beyond the central character what we should be considering is? Do the other characters convey some information about the sound of the central character; that is how far can you go before the, this far off character does not provide any information about the pronunciation of the central character. So, so here the orders of the study used a concept called mutual information. So, mutual information tells you with what probability you can predict a certain variable x right if you had knowledge of another variable y ok.

So, when they calculated the mutual information between the central character and any other character in the window. So, they found that when you go to a window size of more than 3 right on one side right the mutual information falls to a very low values like about

0.1. So, beyond that the characters do not convey the information about the pronunciation of the central character.

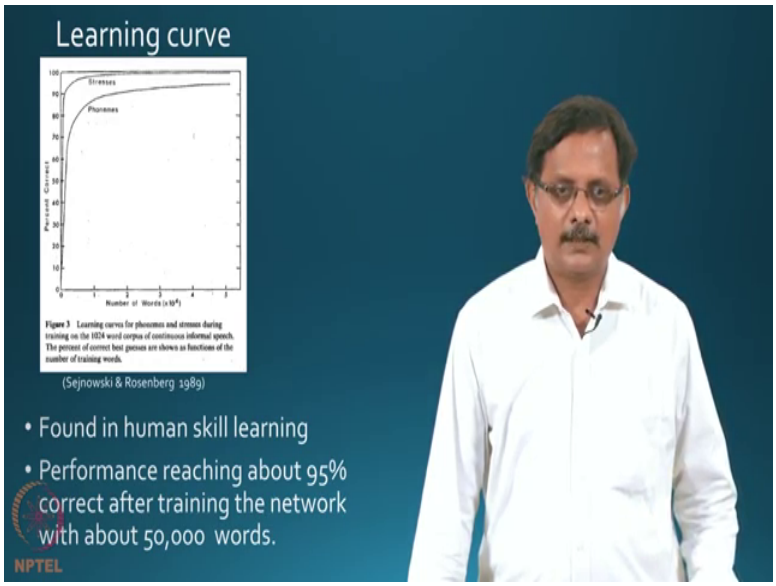
(Refer Slide Time: 13:18)



So, therefore, they have taken a window size of 7. Now the data that they have taken for tending network they have taken from a dictionary having about 20,000 words and among them 1,000 words were chosen these are very high freq umm high frequency words as were specified in this brown corpus of most frequent English words and then these words were mapped on to continuous speech which was annotated this was speech gene know taken from speech of a child and the way characters are not known the phonemes are like this. Suppose you have the word phone right the first syllable is pha right then both p and p and h together are mapped on to the you know the phoneme pha and then there is space.

Then o the letter o is mapped on to phoneme o and letter n the letter sequence n e are mapped on to the phoneme n, because of wherever there is a missing character you know the in the output there is no phoneme there is just a missing space.

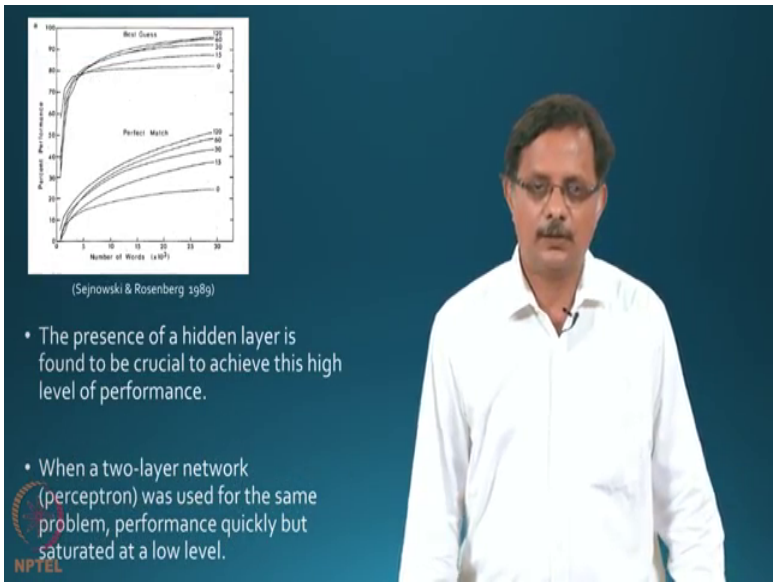
(Refer Slide Time: 14:12)



So, as a as a trained network on this data what they found is ah? So, the performance of the network gradually increases and it shows a exponential kind of behavior that is it rises quickly in the initially and then flattens out after sometimes.

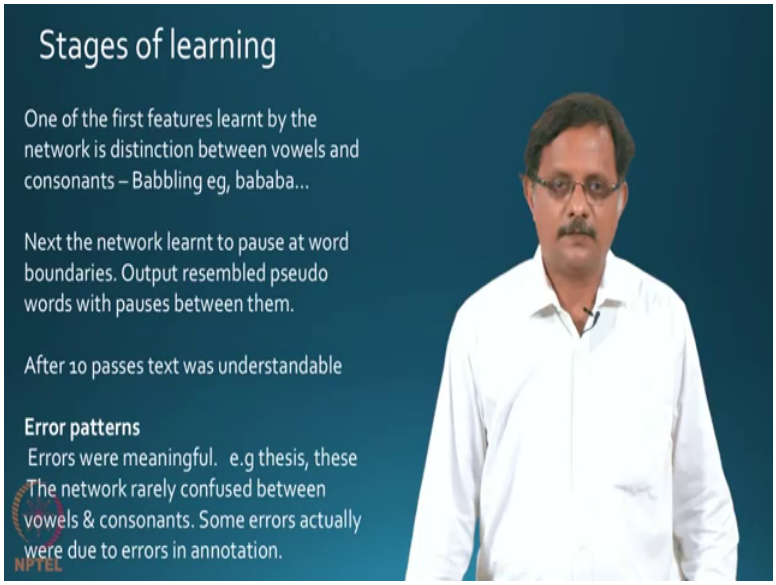
So, here the kind of the learning behavior of the network is stronger as similar to human skill. When humans learn a new skill right they learn very quickly in the early stages and after that, so in; so the performance saturates or flattens out ok. And in addition to the phoneme accuracy they also looked at stresses, I mean they also told the network also try to predict stresses and network is able to do predict stresses more accurately than it predicted phonemes.

(Refer Slide Time: 14:54)



Now, they also compared the performance of a perceptron with a multilayer perceptron. So, that is when there are when the network had no hidden neurons. So, that is a perceptron network learnt very fast initially, but flattened out very quickly and. So, they saturated a very low level of performance as you can see in a in a couple of these curves. So, in some of this curves 0 means it is a perceptron. So, number of neurons is 0.

(Refer Slide Time: 15:24)



And as they kept increasing the number of hidden neurons the performance kept on increasing then stages of learning. So, the what is interesting is? As a network kept on

learning; so, the output of the network which is a phoneme sequence was given to a speech synthesiser the standard you know commercial speech synthesiser. So, that you can listen to what the network is producing as output and the pattern of speech that is generated is very similar to how children learn to speak which is what know very interesting because the network is very simple.

We really did not incorporate anything about the speech production systems in the brain or anything like that. So, another first thing that the network was has learnt is distinction between ovals and consonants. So, first of all the network was never told that certain characters are vowels and certain characters are consonants it was given just given a continues you know string of letters on the input side and a continues sting of phonemes on the output side.

But when they act in the earlier stages. So, when the network produces character in these words it look as if it is babbling. So, for example, it would it would map all the consonants on to one consonant again all the vowels on to the one vowel. So, it would produce something like ba ba ba ba pa and so on

And that is interesting because these are the kinds of sounds that in children at a very early stage when they just learn to speak right a most comfortable with. So, therefore, the even the family members are often given names which are like product results of babbling or productions of babbling like an mama and papa and kaka and so on. So, network also makes this kinds of sound which is interesting.

Next the network learn to pause at word boundary. So, initially it was producing this continuous you know strings of sounds whereas, after some learning after some training network learn to pause at word boundaries and word boundaries are denoted by blank spaces right. So, it looked as if the network is produced is producing random words or pseudo words right when a word is given, but it had learned to pause between those pseudo words.

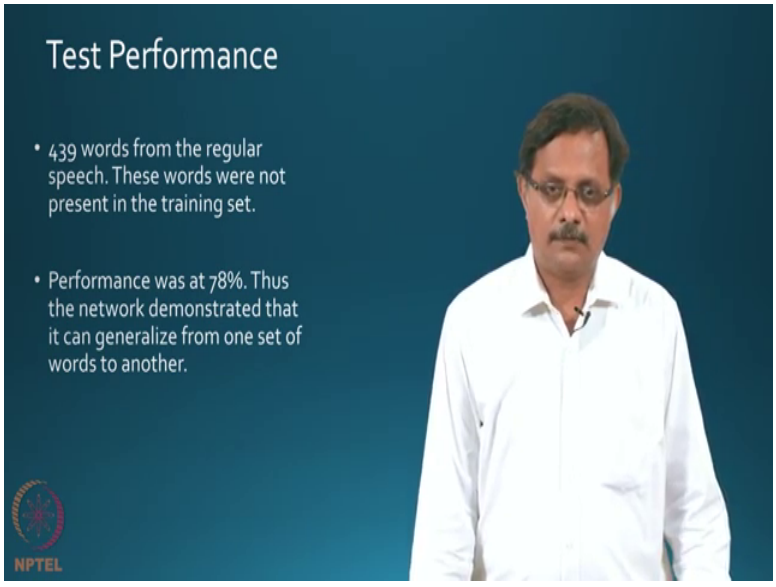
After about 10 passes the networks speech was understandable. It would it could make sense out of what the network is saying right and this looks a lot like; how children would learn to speak? And the error patterns if you look they are executed by the network are also quite meaningful for example, if the word thesis is given network

output might go something like this which is very close to cases and it also rarely confuse between vowels and consonants.

So, it is very clear that it is the distinguishing vowels and consonants in a separate study it was also verified; that some neurons in this network in the hidden in the hidden layer some neurons respond specifically to vowels and some neurons specifically to consonants.

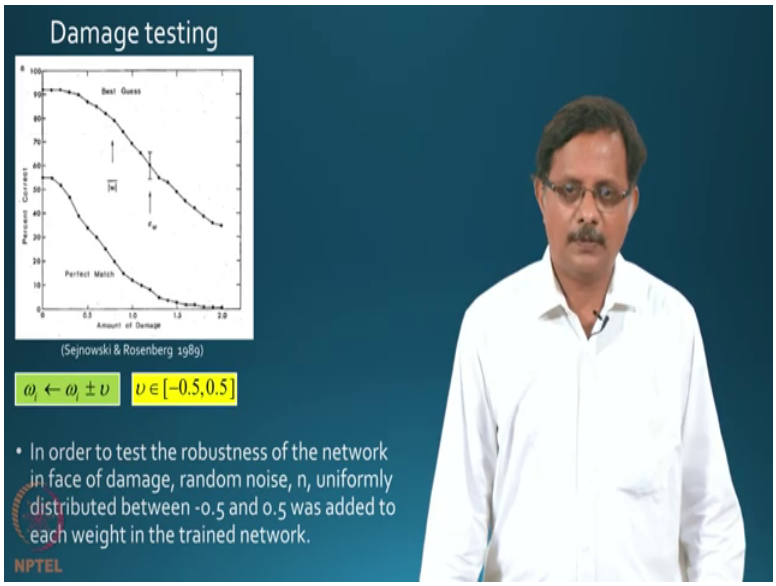
So, network automatically had learned that consonants and vowels are two different classes of sounds and its learn to distinguish is to classes very clearly also; what is interesting is? Some of the errors at the network made are not really errors by the network, but these were actually errors in the annotation. So, in the training side there were mistakes and that actually seem to make it look that network and made the mistake, but actually network was correct and the networks output had pointed out the errors in the annotation.

(Refer Slide Time: 18:28)



So, test for test performance the network was tested on 4139 words from regular speech, and these are not present in training set and performance was 78 percent which is which is pretty good. So, what that showed is that network and generalize for once at a words to another.

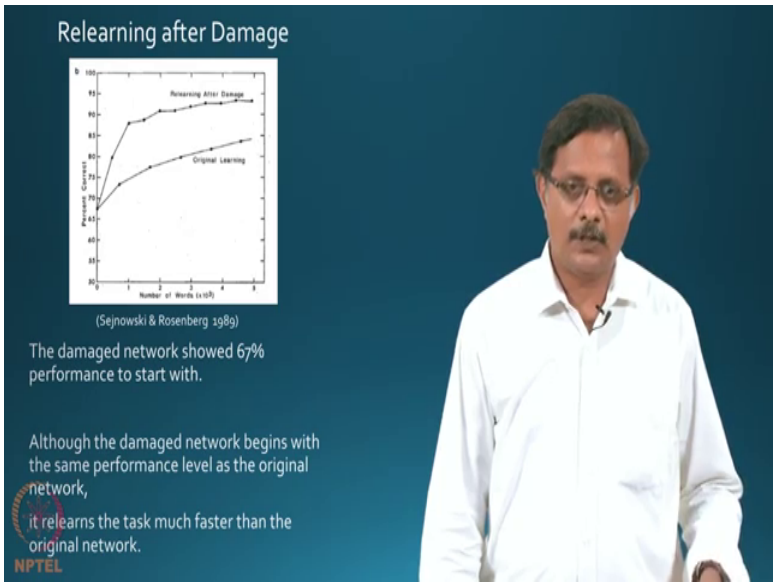
(Refer Slide Time: 18:45)



Now, people so this is the receptuals also studied; what happens when you damage the network. So, when you damage the network so this was done by adding noise to the weights right and. So, if you can with the two curves in this graph. So, then the top one is the best guess what is called the best guess and the bottom one is called the perfect match.

So, basically when you when the network produce output there is a 26 dimensional output which corresponds to the properties of this of the phoneme. So, so when the output exactly matches the correct properties. So, that is what is called the perfect match? Right in other cases; what they have done is? This to see which is the phoneme vector which matches the actual root of the network best and that is what is called the best guess? So, best guess performance is; obviously, better than the perfect match performance and also the best guess performance turns out to be more robust to damage than the perfect match, because in the best guess performance when you added a noise up to point five performance did not change that is the network is very robust to damage.

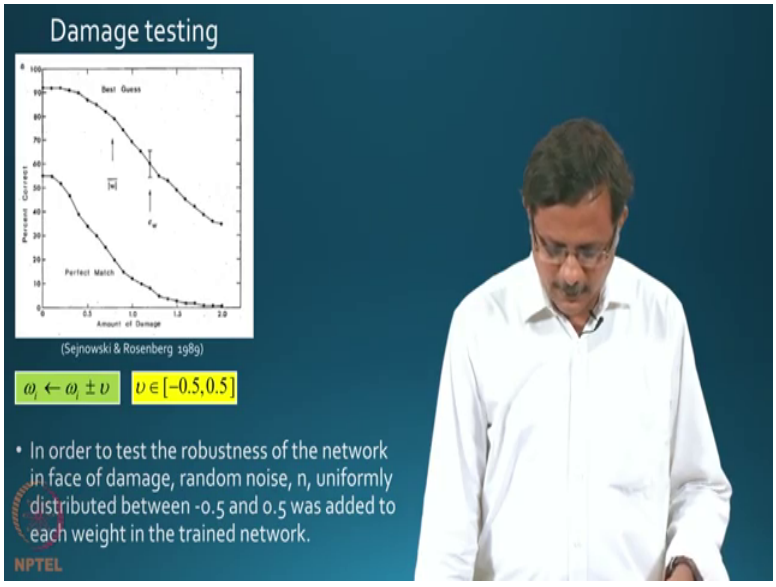
(Refer Slide Time: 19:54)



So, one more aspect of damage which is very interesting is that is; how does s network relearn after damage? How does it recover from damage? So, in this study they have taken a network they have added some noise and sort of damage the network by adding noise.

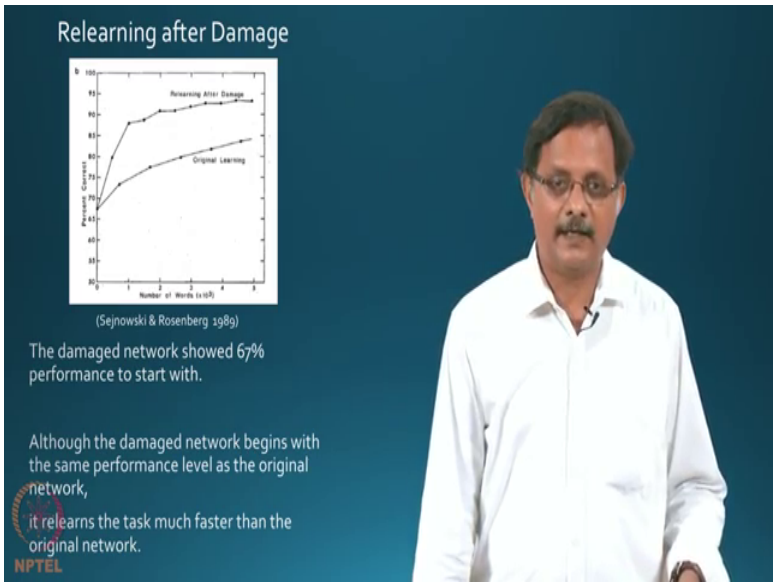
And they are at the damaged network now has a performance of about 67 percent.

(Refer Slide Time: 20:15)



Now, you can see in the previous graphs that network performance goes all the way to 95, 96,000 on the unfading set. So, they have added noise. So, that the performance degrades up to 67 percent.

(Refer Slide Time: 20:23)

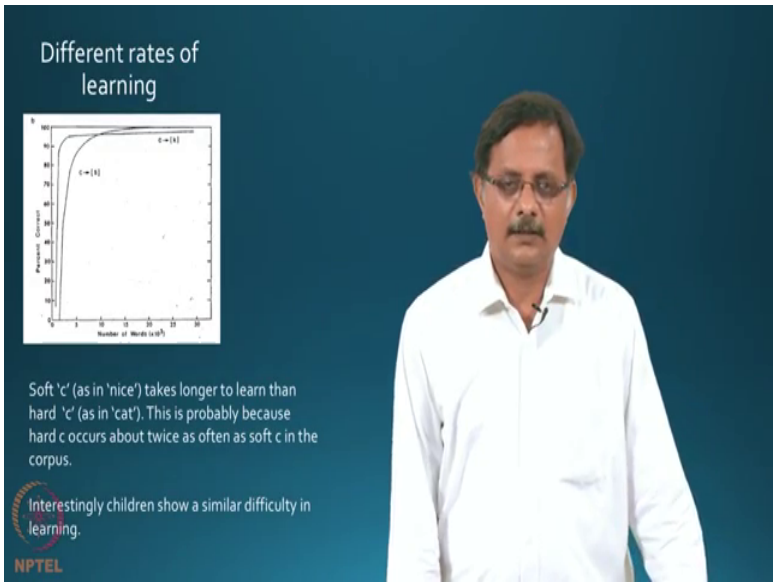


Now, so there are two curves here one corresponds to original learning and the other corresponds to relearning after damage. So, original curve learning curve basically ta tells you how the network has learnt beyond 67 percent original the other curve relearning after damage tells you how the learning proceeded once you have tra started retraining the damaged network.

So, you can clea clearly see that network has lea has learned much faster than the original network when you when you train the damage network. So, this kind of a study is interesting and you know throw shed some light on how people learn after say a kind of a brain disorder or a brain accident like this like cerebral stroke.

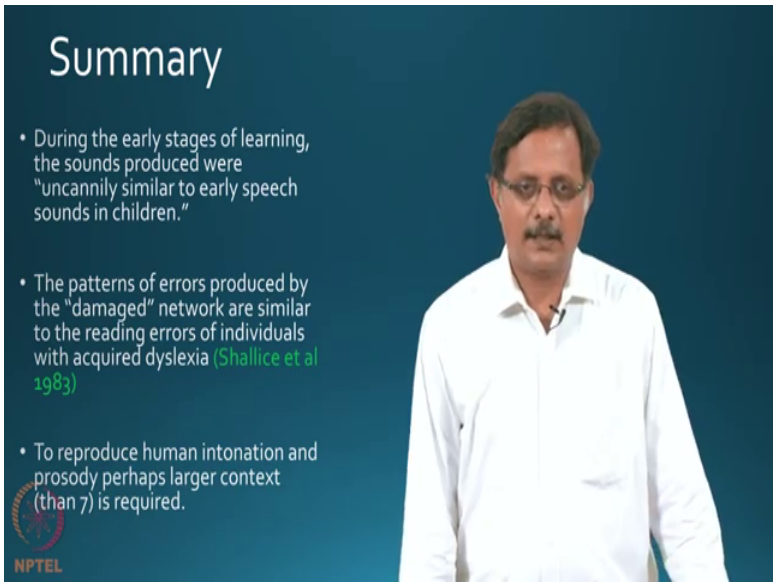
So, stroke is a disorder of you know it is a neurovascular accident where, because of loss of blood supplies and part of the brain can get damaged. Now once the stroke occurs the patient is put through lot of you know physiotherapy cor cognitive therapy and so on the kind of therapy that is appropriate for the kind of stroke that they have and it turns out that when stroke is not too extensive the patients recover much faster right a learn relearn that task or that that function must faster than what they would have taken right in the beginning as they when they have learned that task as a child.

(Refer Slide Time: 21:56)



. So, so the networks behave also reflects some of these phenomena from critical neuroscience then it is also when observed from this studies that the network learns different sounds at different rates. So, for example c has no the character c can be mapped on to different phonemes c is pronounced as. So, as in nice, this is called a soft c and c is pronounced as ka as in cat this is called hard c and it turns out that soft c takes longer than hard c and that is what the network has shown and that is; because you know hard c occurs what about price has often as soft c in the data set interestingly children also show the similar difficulty in learning similar pattern of learning when it comes to pronouncing c.

(Refer Slide Time: 22:36)



So, in summary we have seen that the restages of learning the pattern of speech that the sounds produce is very similar to that of children. In fact, in the original paper it was observed that the sounds produced for uncannily similar to at least speech sounds in children.

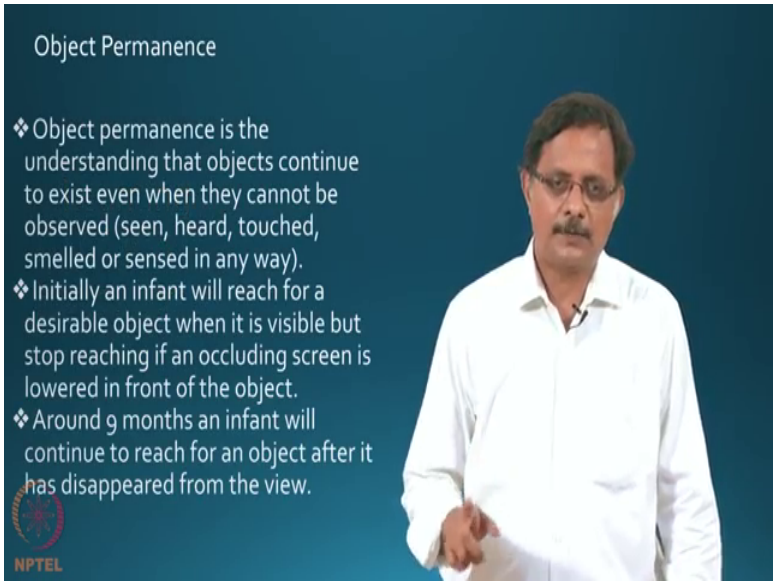
The patterns of errors produced by the damage network and also similar to getting errors of individuals with acquired dyslexia refers to difficulty in learning and reading. So, the kind of an arise that individuals dyslexia showed is very similar to the performance of damage network, but to produce show an intonation and prosody that is kind of the variation pitch that you see as you know matter a whole sentence right as for that you need much lin longer contacts than the seven characters is to have a idea of that the structure of the entire sentence to be able to produce prosody.

So, that needs you know a different kind of a network architecture. So, the next phenomena will be studying is learning object permanence ok. So, before we talk what object permanence let us discuss what are objects? So in fact, let us ask the question right; what do we do with the sensory input. So, with the eyes we look at the visual world with the ears we look at the; we receive the sounds for you know from the surrounding world and we can touch and feel the world you know through touch sensor through skin.

Now, when we receive all this information; what we are doing with the all the sensory input that is; streaming in is to extract knowledge of extract concepts. So, we do not look at the world as you know extracts you know blanches of color which do not have any labels right very quickly we map this blanches of color and form on to certain distinct objects. So, I do not think that you know this kind of a red blanch that is seen on the right is just a patch of color, but it is actually a person who is wearing may be a red shirt.

So, similarly this yellow blanch that I see on the on my left is not just a just a yellow patch of color, but it is a yellow wall an object. So, so by analyzing the sensory input right we are carving out objects and we give them certain na labels. So, the need to a carved objects is that objects are more permanent they have they are more lasting they are they represent a certain variants right in all the variation and uncertainty that you see in the world outside.

(Refer Slide Time: 25:01)



So, one aspect of objects is there is object permanence that is the understanding that objects exists even they cannot see them you know directly. So, for example, if you are watching somebody walking and somebody just walks behind a tree; obviously, you do not think that object has the person has suddenly disappeared. So, the person is there, but just behind the tree and you do not think that person has gone somewhere and for any moment the person might emerge back from behind the tree.

So, this notion of that objects exist even though they are not observed directly right is called object for permanence and infants do not have it at a very early stage and only around 9 months right in the infants show this proof of object permanence.

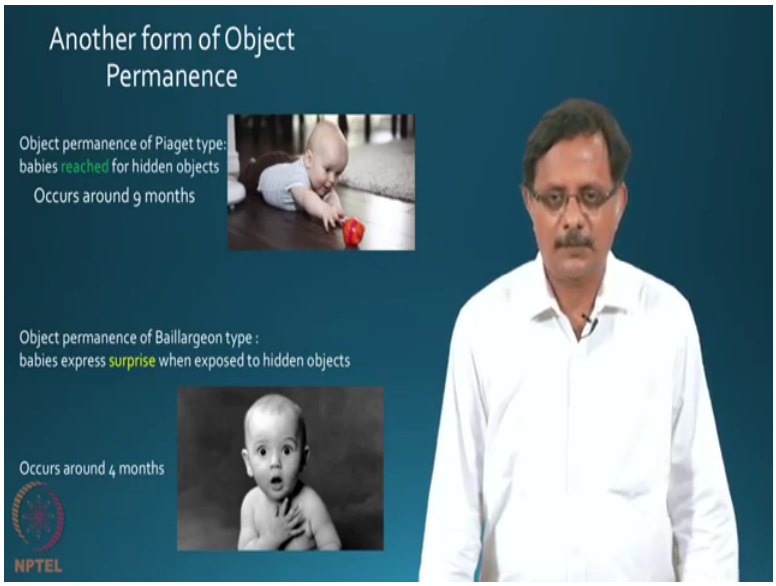
So, for example, if you take an infant who is you know lesser than 9 months right the infant might ge you know might reach out to an object which is visible, but if you hide the object under a carpet or something like that the infant would not go for it while after 9 months right even if you hide the object under carpet right the infant might try to struggle to go under the carpet and pickup the object.

(Refer Slide Time: 26:06)

So in fact, mothers played is simple games with their babies; what they are testing or what they are may be triggering in the minds of the infants is this? You know perception of object permanence. So, very young infants may be less than 4 months they do not know understand this game. So, in this game the pickup of game the mother has to cover her face very moment early and then you know uncover her face again.

So, when the when the child the baby understands that the face exist even though the face is covered right. The child the face shows you know feelings of happiness or surprise right and the that. So, in this in this game basically the infant is revealing the you know the ability to recognize object permanence.

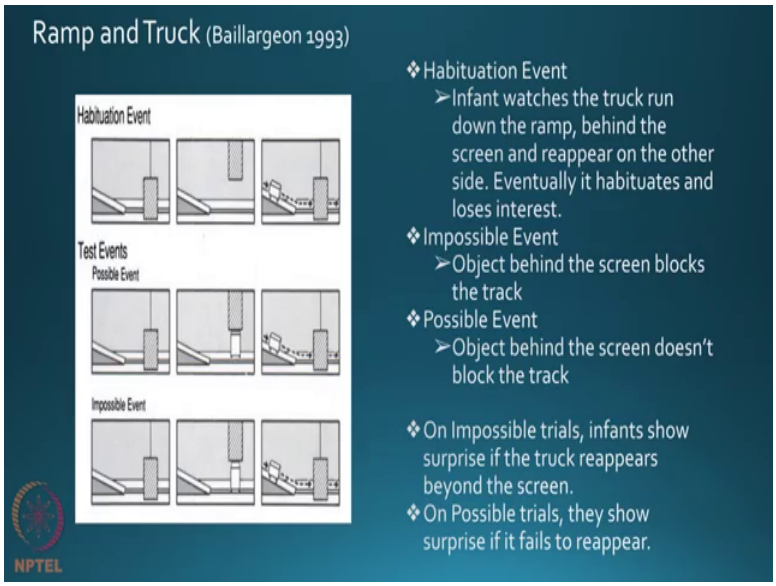
(Refer Slide Time: 26:51)



But it has been found that the two kinds of object permanence. So, say the earliest work on object permanence was done by Piaget like, I mentioned he is a child psychologist a Swiss child psychologist a lot of his early work was done on his own children. When they were infants and he noticed that; so babies reach for hidden objects right when they are around 9 months, but before that they reach for only visible objects were not for hidden objects.

But later on another scientist named Baillargeon right found a different kind of object permanence he has found that babies express surprise when ex when exposed to hidden objects and this kind of a object permanence can be observed even around 4 months ok. So, certain early stage they simply show evidence of the that capability to rely object permanence right ah, but by simply expressing surprise, but in the later stage they show the same proof of object permanence by actually reaching out to hidden object.

(Refer Slide Time: 27:53)



So, Baillargeon experiment you know from his 1993 paper goes something like this. So, there is a little ramp right and there is a screen in front the. So, the rectangle with dash lines says is a screen and there is a trunk that rolls on the ramp line the ramp and rolls down and along the track and as it rolls down the track it goes behind the screen momentarily and ca emerges from the other side of the screen.

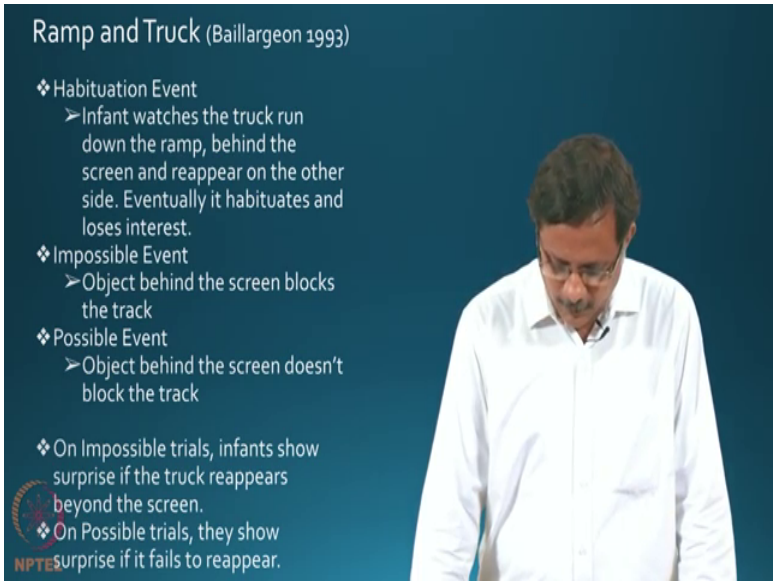
Now, sometimes the screen is lifted revealing the track. So, in the first set of trials which is called which are called the habituation panels or habituation events like the truck rolls down the ramp and rolls along the track goes behind the screen emerges from the other side of the screen ok. So, now, in some cases the experimental places in objects or an obstacle right behind the screen. And here again there are two subcases.

So, in one case which is called the possible event the obstacle is placed behind the screen, but the obstacle does not block the truck. So, as the truck rolls down the ramp and rolls along the track this the; observe the truck and roll forward behind the screen and emerge from the other side of the screen whereas, in the in impossible event right the obstacle is placed right on the track. So, that when the truck rolls down the ramp and rolls down the track it gets blocked by the obstacle and cannot emerge from the other side of the screen.

So, now what we are looking for is the understanding that in the possible events the truck can emerge from the other side of the screen whereas, in the impossible events the truck

cannot emerge from the other side other side of the screen because there is an obstacle a blocking its path.

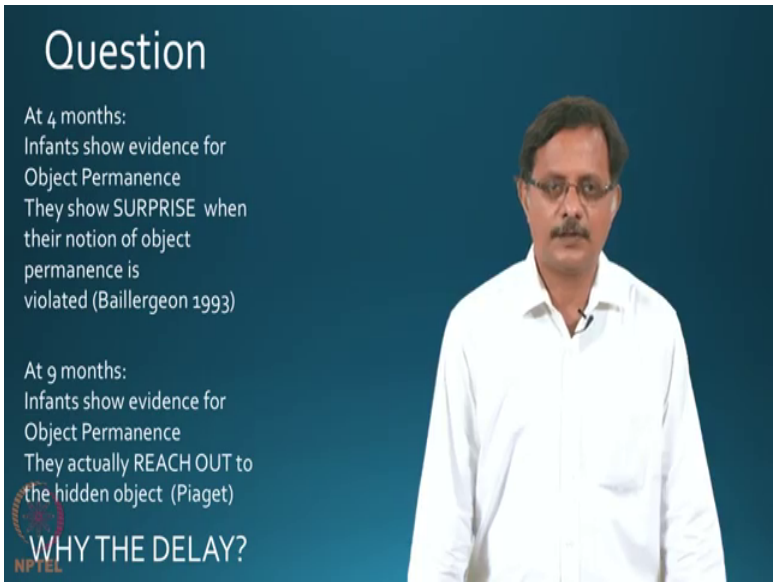
(Refer Slide Time: 29:44)



So, when this experiment is performed with babies what was is what was observed is? On impossible trials the infant show surprise if the sa if the truck reappears beyond the screen, because there is no support to appear because suppose to be blocked by the obstacle.

Whereas on possible trials they show surprise if it fails to reappear because on possible trials this; the truck should be able to come out from behind the screen because the obstacle is does not blocking the path ok. So, this is what has been observed.

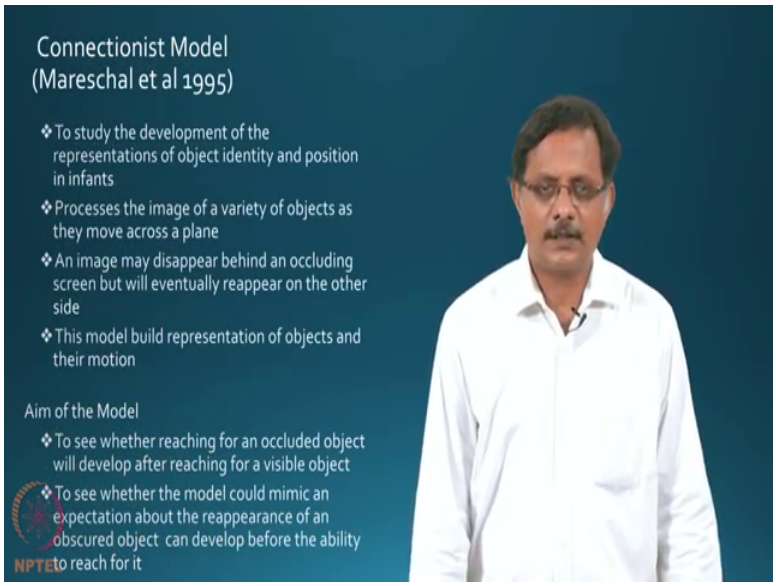
(Refer Slide Time: 30:14)



So, the question now is. So, we have seen that there are two kinds of object permanence or the children or infant show the knowledge of object permanence in two ways: one is simply by the surprise and the second is back to reaching out which is; what Piaget respond and the second kind of which is by surprise which is what Baillargeon has found.

Now, at 4 months infants show object permanence they show surprise when there are notion of object permanence is violated whereas, at 9 month infact show infant show evidence of object permanence right they actually reach out to hidden objects and that is the kind of na object permanence that Piaget has discovered, but why is there this kind of a delay and can be reproduce this kind of an effect using a computation model using a neural network model.

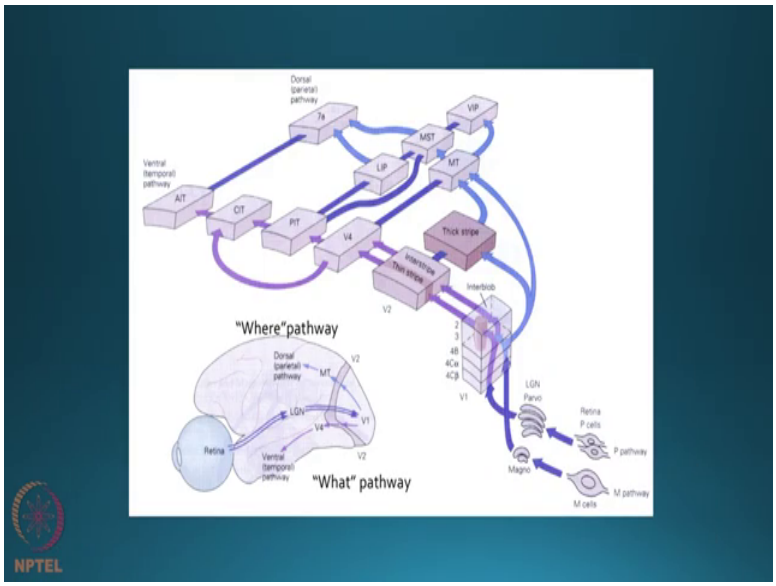
(Refer Slide Time: 31:02)



So, connections model or a m l p kind of model of this phenomena was developed by Mareschal in a paper published in 1995 right. So, in the aim of this model is to see whether reaching for an occluded object will develop after reaching for a visible object because that is what Piaget has found right children or infants reach out for a visible object first and only much later they are all the concept of object permanence and reach out for an occluded object.

So, the second question is to see whether the model could be weak an expectation about the reappearance of an obscured object which can develop before they will try to reach for it. So, that is. So, the expectation here is the surprise that they have found in Baillargeons experiment. So, the surprise or experi expectation right about an object about an obscured object can be converged even before the ability to reach the object. So, to look at the network before you look at the network architecture let us quickly look at sim some aspects of the visual system how brain passes visual information.

(Refer Slide Time: 32:08)

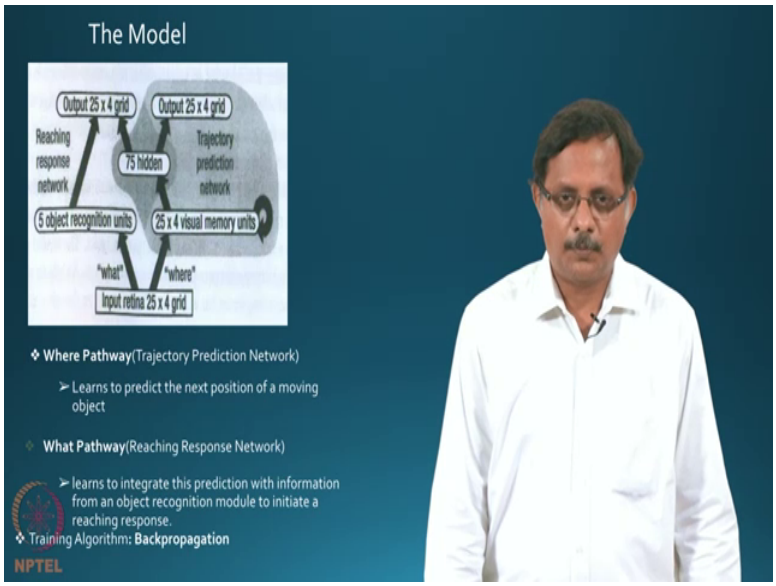


So, in this picture we see a kind of a simplified picture of the brain side view of the brain and you can see the eyes and. So, when you look at something the image of that scene falls on the retina as you all know and retina, then sends the image information in the form of electrical signals by a bundle of fibers called the optic nerve. And this information goes to one stage in the brain called the l g n which you can see in this figure and from l g n then it proceeds fur further and reaches a part of the occipital lobe which is called the V 1 this is also called the primary visual cortex

Then from V 1 it they have fibers which project to you know in two different directions. So, the. So, it goes to V 2 which is the secondary visual cortex and from V 2 it goes to it emerges along two pathways one pathway goes to another area called empty and so on into the parietal areas or in the superior parietal areas this is called the parietal pathway or the dorsal pathway, because it is dorsal to the brain and it is also called a where pathway because these parts of the brain which crosses where is the object that you are looking at. Then another stream start begins from V 1 and proceeds downwards by the mental pathway, because it is a mental path of the brain and it goes into the temporal lobe into the inferior temporal areas this is called the what pathway.

Because at the end of it in these parts of the brain the brain recognizes; what is that object ok? So, and this is relevant to understanding the kind of architecture that is; used to model object governance.

(Refer Slide Time: 33:46)



So, in this proposed model by Mareschal at all the network architecture is shown in this figure. So, there is a input layer which is a 2 dimensional array which represents a retina. So, input is called the retina which is an array of size 25 by four which look something like this.

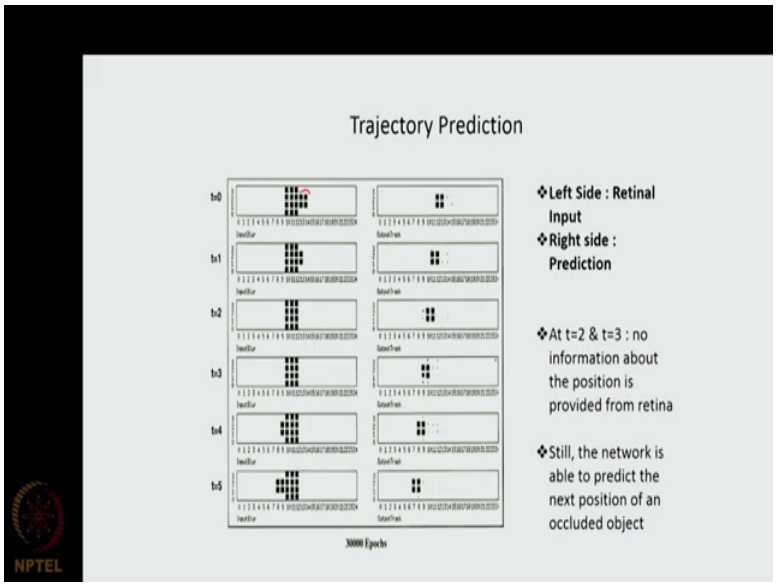
So, what you see on the left is the input retina, then and there is the two hidden layers that you can see and so, the from the input you proceed in one direction towards the right along the so, called where pathway. And there you have again another array of size 25 by 4 and this is a visual memory a layer, because here there is a feedback from layer to itself that is this layer memorizes the state of the layer from the previous step and that is used this also projected to the next layer.

Now, on the left side you have a layer which has 5 neurons and these are object recognition neurons ok. These neurons recognize; what is the object? And the neurons on the other layer on the right side recognize where is the object?

Now, the 25 by 4 visual memory units layer projects to one. Another hidden layer of size 25, 75 neurons and this in turn projects to two output layers. There is one on the left which is 25 by 4 grid and the 1 on the right is also 25 by 4 grid, but both these outputs represent different things they ca the outputs are different they train to produce different outputs ok.

So, what are they how is the network trained and let us look at that. So, the network on the right. So, the output on the right is trained to predict the next position of the input object. So, here in this image.

(Refer Slide Time: 35:31)

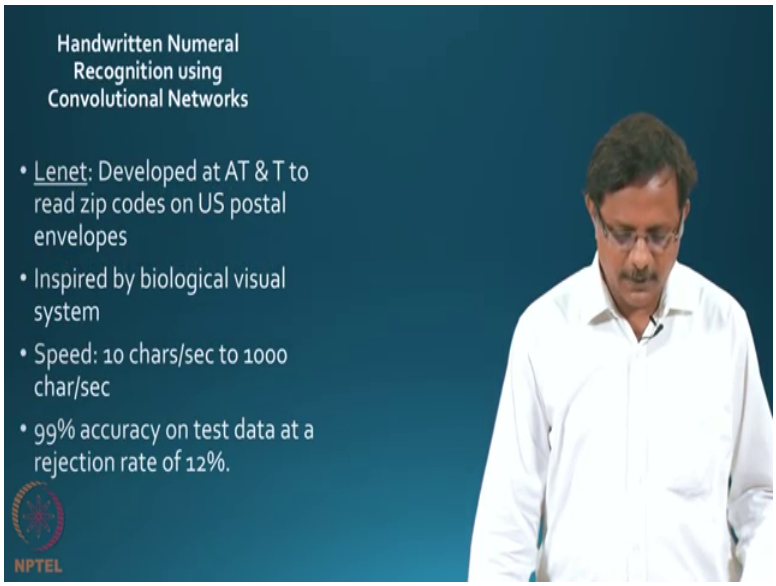


So, this little square right of size 2 by 2 is the object and this bigger rectangle of size 3 by 4 is the occluding screen. So, what you see in the left set of column of images is you see the object coming from the right and it just enters and goes behind the screen and completely it was behind the screen and it stays like that for two steps and then a t equal to 4 it just begins to emerge out of behind the screen and d equal to 5 it comes completely outside out of the screen.

So, this is the input that is shown to the network and the right output layer that you see here. So, the when the output on the right side is; suppose to as predict, the next position of the object so this is called a trajectory prediction, next we look at a class of neural networks called the convolutional neural networks.

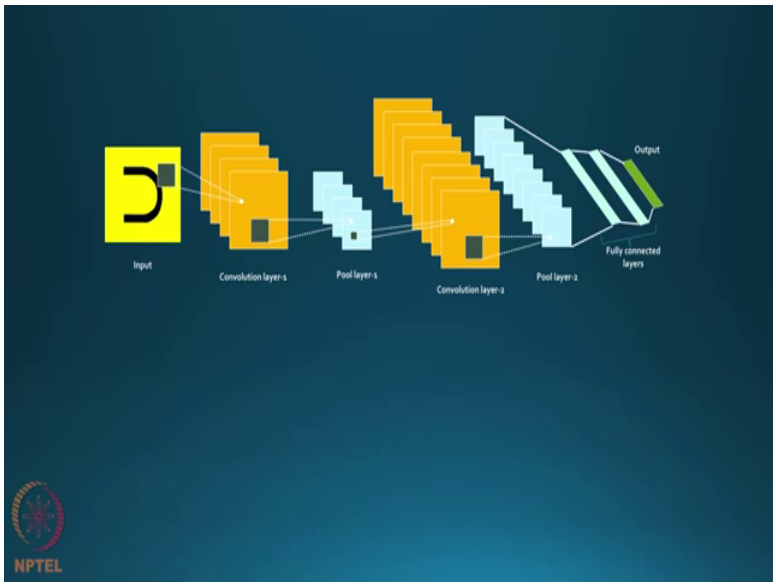
So, these networks are actually you can think of them as special case of multilayer perceptrons. And the only thing is multilayer perceptrons are three layered networks that we have most of them that you have seen a three layered networks. So, although the original theory can be extended to any number of layers more specifically; the convolutional network or a special exam case of multilayer networks because in this case in this networks all the layers are two dimensional layers ok.

(Refer Slide Time: 36:55)



And the original convolutional networks were developed as an application to read hand written numbers on postal envelops this was done by a sponsored project by u s postal service it was done by a company called AT and T and.

(Refer Slide Time: 37:15)



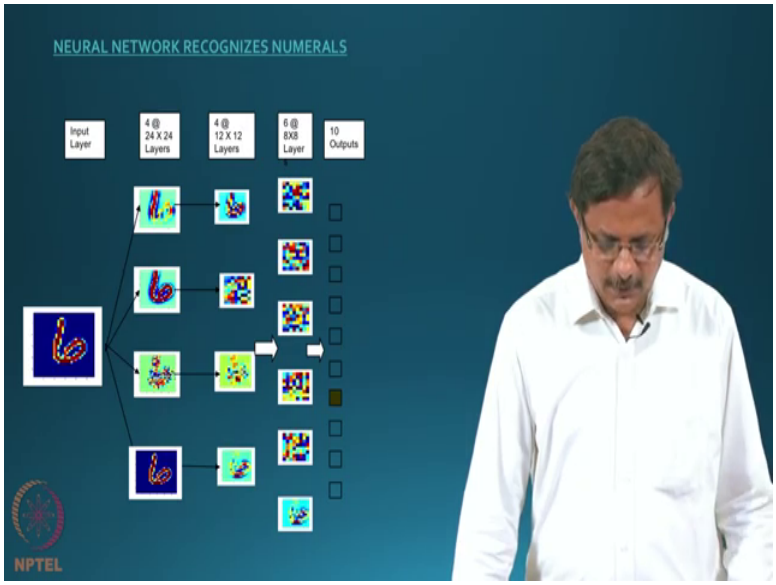
So, in the convolutional network, the architecture of the network looks like this. So, I give an input which can be a character or any image of a visual pattern which is shown here. So, each of the hidden layers consist of multiple 2 dimensional sub layers. So, this is the orange stack that you see here is; the first hidden layer it consists of 4 sub layers

and that is each of this sub layers is a two dimensional area of neurons. So, neuron in any of this sub layers are does not look at the entire input image, but looks at only a small window of that image.

So, similarly if you go to the second hidden layer which where you see this blue stack of layer of sub layers and the each of the sub layers is again a two dimensional area of neurons a given neuron in this hidden layer again looks at only a small window of neurons in the previous layer and so on and so forth. So, then again the next layer you have another start of two dimensional layers and each neuron in this sub layers that also look at only a small window and not the complete array in the in the previous hidden layer.

So, at the end of it you have one array of single one dimensional array of neurons. So, in this case since we are trying to classify numerals from 0 to 1 right the output will have only 10 neurons correspond to the 10 classes ok.

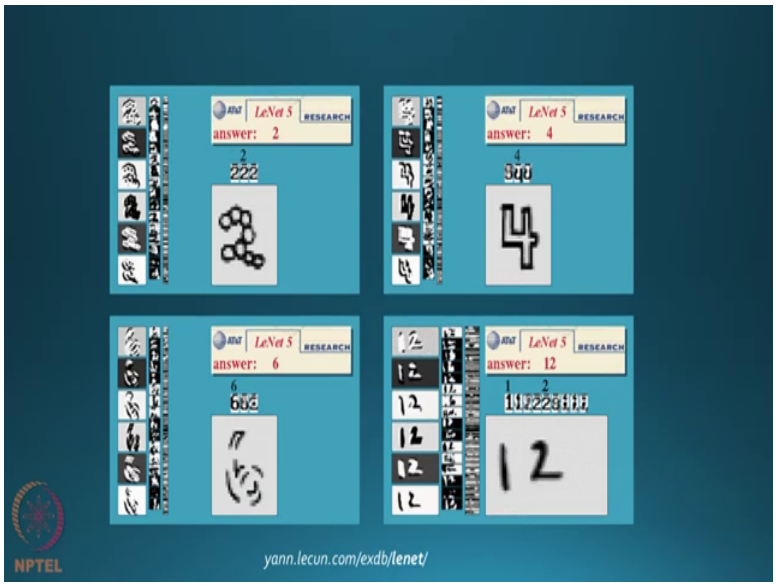
(Refer Slide Time: 38:39)



So, this shows a schematic of what happens when you give a sample image? Let us say for example, 6 right the each of this arrays show the response of the sub layer right at various hidden layers.

So, once the network is trained they for testing the network they have given very complicated you know cases like for example, when the network is trained.

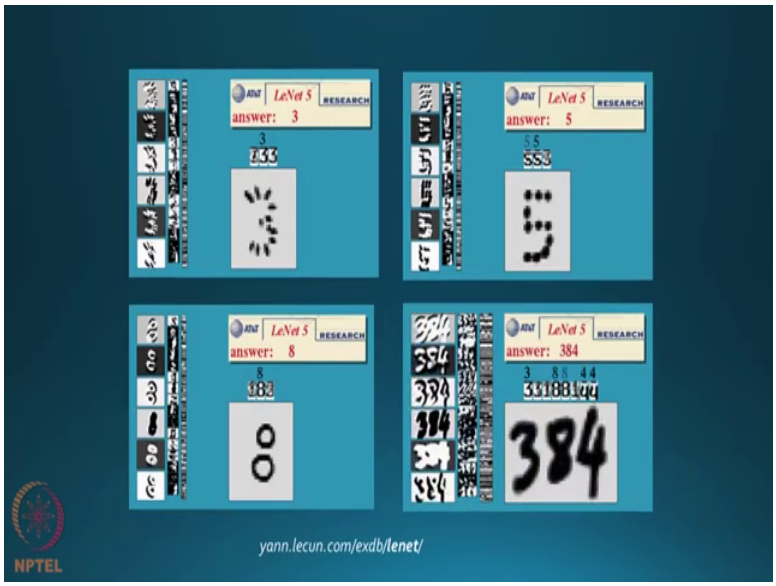
(Refer Slide Time: 39:03)



So, the kind of patterns that they were given are standard numbers like you know 2 is written like this or 4 is written like this and 6 is written like this, but for testing they have given all these complicated a contrived cases; where 2 is written as a ring as a chain of lots of small rings or the 4 is a kind of a hollow, 4 there is a there is a white space in the middle of the 4 and 6 is a hollow 6 ah, but also it has lots of breaks in the outer contour and so, on and so, forth.

But in spite of all this gross distortion of the input the network is able to no identify the input accurately.

(Refer Slide Time: 39:42)



So, for example, there is a 3 which is ex written using small flashes small strokes and 5 which is written using lots of dots and 8 is written as a combination of 2 circles right in all these cases the network is able to recognize the character correctly.

So, this shows the network seems to generalize very well to complicated cases just as humans would generalize.

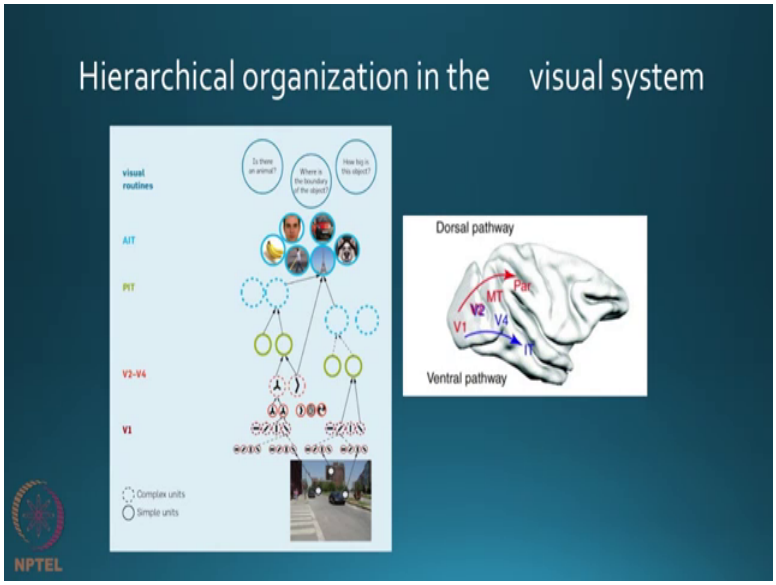
(Refer Slide Time: 40:05)

So, let us look at one more example, the convolutional network in this case was trained to recognize faces the human faces. So, after training they went into the network and

probe the response of neurons to find out what it is responding to? So, it turns out that neurons in the lower layers right, since they can look at only a small part of the input image right. They are actually responding to edges right and whereas, if you go to neurons in the next hidden layer neurons here responding not to the entire face, but only to parts of face like for example, an eye no or a nose right or a lip and so on.

Whereas, if you go to a next level right neurons here seem to be responding to entire faces, but not specific faces they seem to be responding to whole class of faces right now. So, this is very interesting because the network seems to break up the problem in very special ways. Now is that what happens in the real visual system. So, let us look at some data from real visual system.

(Refer Slide Time: 41:02)



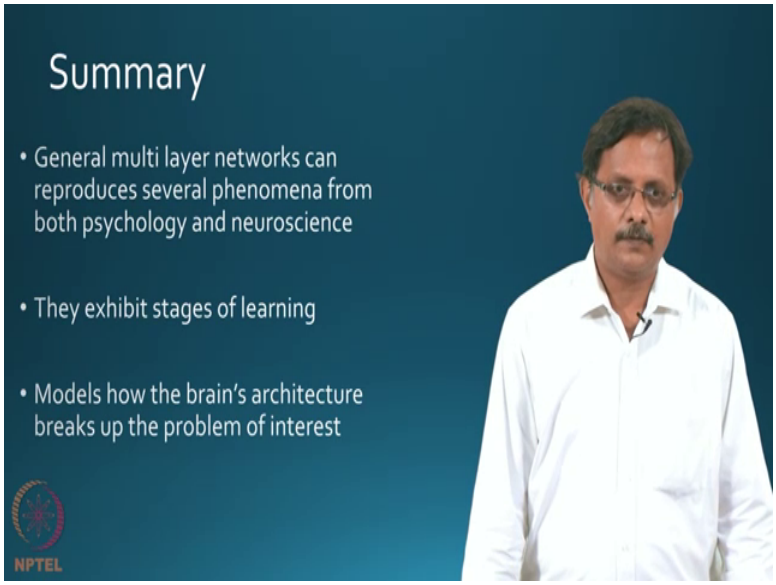
So, you have seen before when you receive input from the eyes though the information goes to the first cortical stop over which is the primary visual cortex they called V 1 right. And then it goes to V 2 and then from there in the ventral pathway, it goes to a region called I t or inferotemporal area and this is where the complex objects are recognized right.

So, if you look at what neurons in different layers of the of the visual system represent? right. So, in the figure on the left you can see that V 1 neurons represent or respond to only edges. So, these edges could be of any orientation. So, they respond to edges with orientation.

Then if you go to next level say V 2 right you have neurons which respond not just to edges, but combination of edges like angles or you know or even pa you know features, where three lines could come together forming a trijunction and as you go higher and higher right in ah; and in different parts of different temporal cortex like anterior inferotemporal or posterior inferotemporal P I T or a I T you have neurons which respond to whole complex object.

So, you see that more primitive features are represented in lower layers of the visual system whereas, if we go to higher and higher stages a more and more complex objects are represented just like what we have seen right in a in an abstract network we just trained on faces.

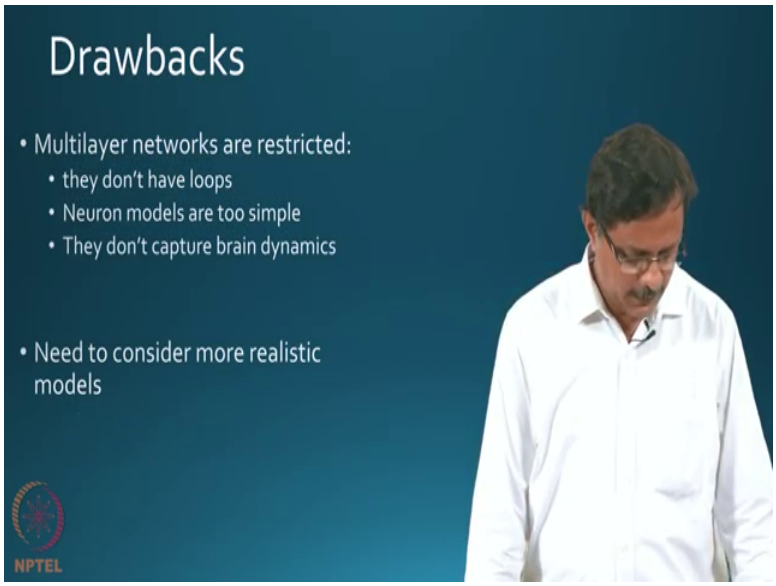
(Refer Slide Time: 42:29)



So, what we can learn from all this examples is that general multilayer networks which are not now which are not very faithful to real world real brain or an real neurons can reproduce a many phenomena from psychology and neuroscience and they exhibit several stages of learning right as; what can Piaget and others have observed in stages of learning in children. And these kinds of networks also modeled brains architecture and it mo the model, how brains architecture can break up the problem of interest?

So, in this case we have seen how the visual system breaks up a problem the for recognizing complex objects and we have also seen; how networks with multi multiple layers like also break up the problem of visual pattern recognition in very similar ways.

(Refer Slide Time: 43:17)



But the network models that we have seen now are somewhat restricted because here when it is; obviously, lot more complicated the models that we have seen. So, far are somewhat simplest for example, they do not have loops whereas, in real brain and the connectivity has lots of loops and so, we need to incorporate that and the neuron models that we have seen also in this in this examples are too simple right and they do not capture real brains dynamics. So, we need to construct consider more realistic models which is what we will do in some of the future lectures.

Thank you.