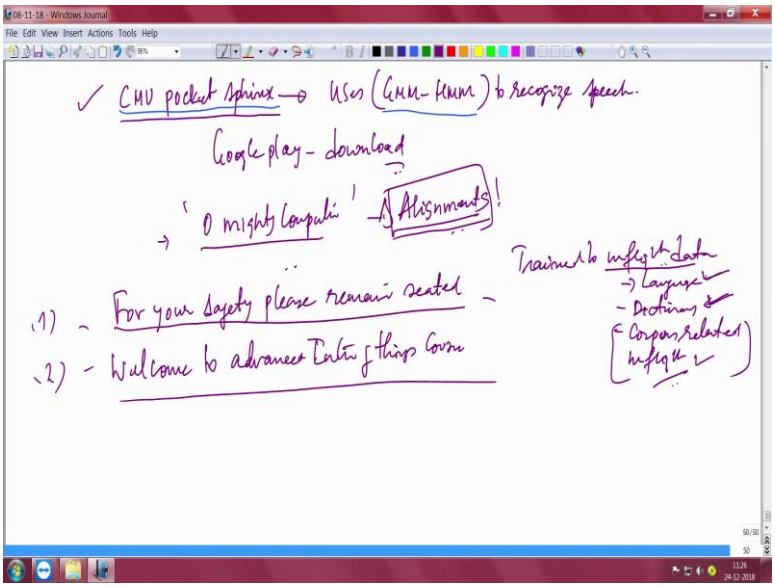
Advanced IOT Applications
Dr. T V Prabhakar
Department of Electronic Systems Engineering
Indian Institute of Science, Bangalore

Lecture - 27
Speech Recognition Part – 4

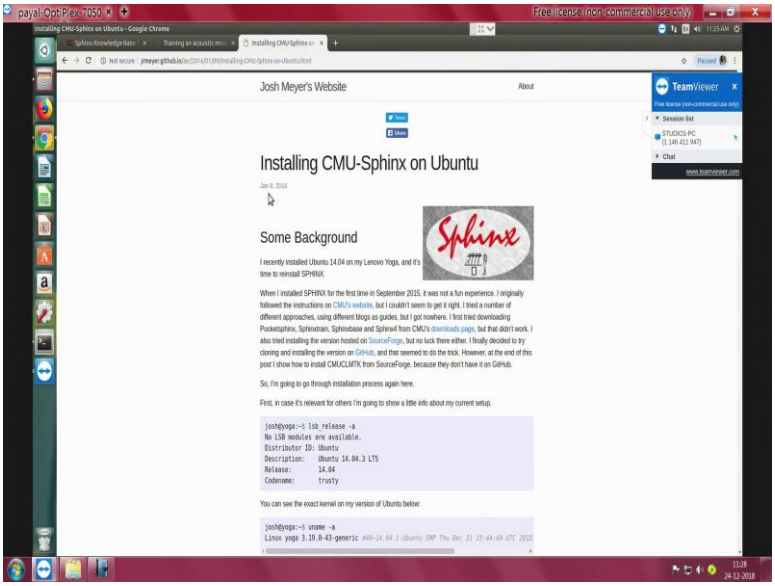
I will show you the list of steps by which you can actually get to train such models right.

(Refer Slide Time: 00:50)

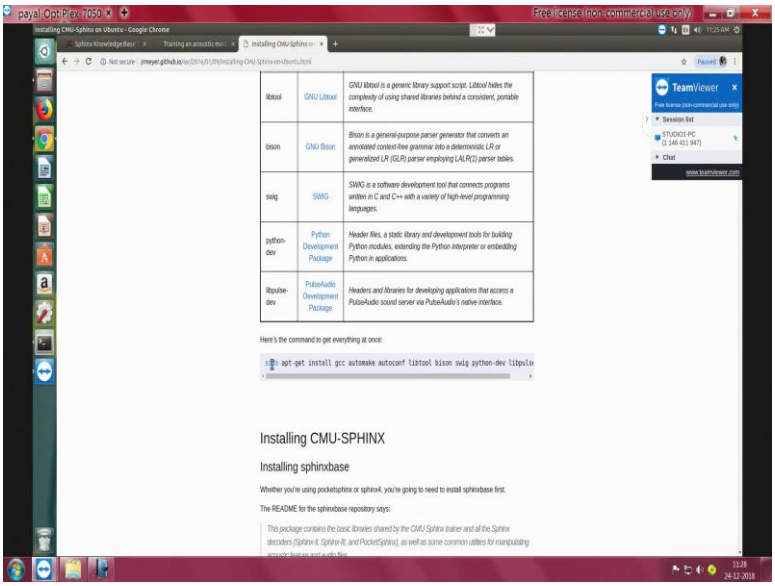


So, let us start with the first step of how you can build applications of this nature by using existing tools that are there particularly with respect to the GMM HMM model which is used by CMUs pocket sphinx.

(Refer Slide Time: 00:55)

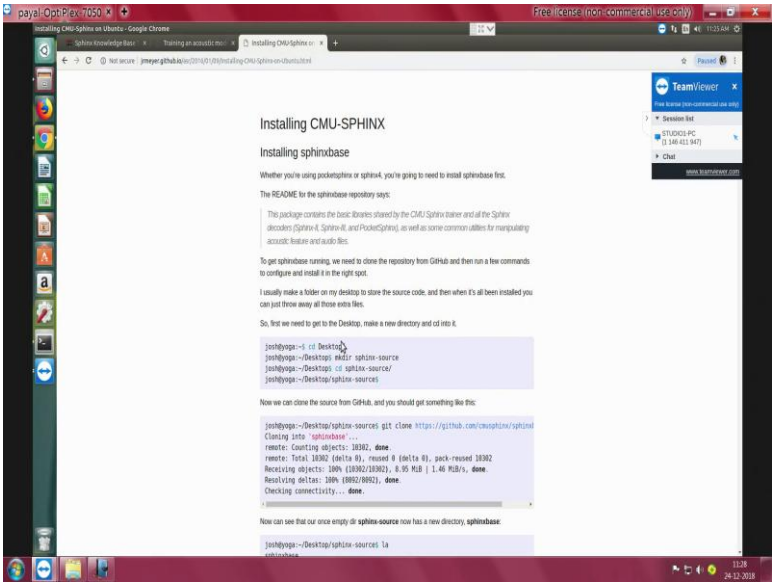


(Refer Slide Time: 01:00)



What are the steps if you go down, the very first step that you may want to do is sudo apt get install gcc auto make you can run this tool.

(Refer Slide Time: 01:07)

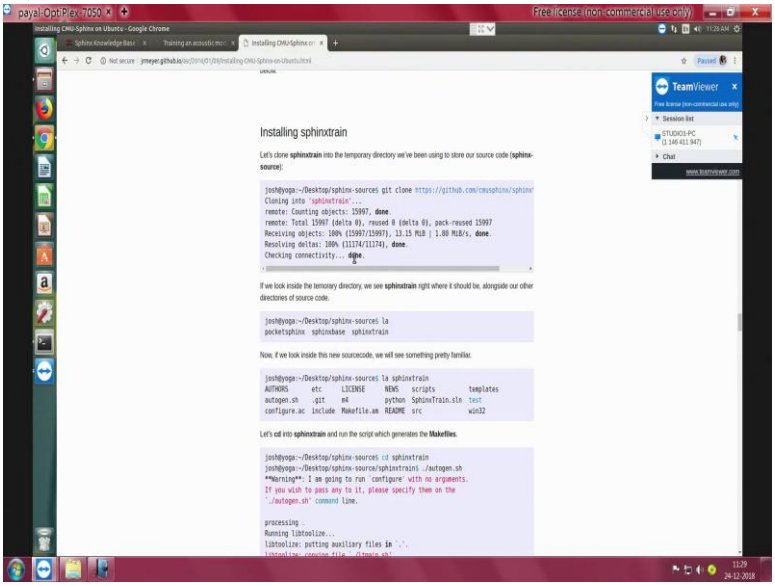


You will be able to install CMUs SPHINX base right. So, you can see installing CMUs sphinx base then say; so, it says cloning into CMUs sphinx base.

(Refer Slide Time: 01:30)

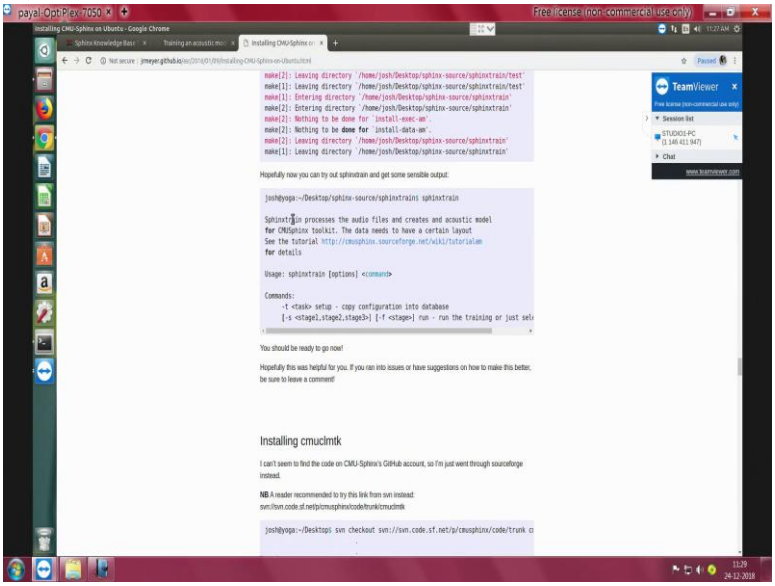
If you go down the next is the next is go down like this, installing pocket sphinx is the next step right. Clearly there are clear indications of what is to be done at step 1 and what is to be done at step 2.

(Refer Slide Time: 01:50)



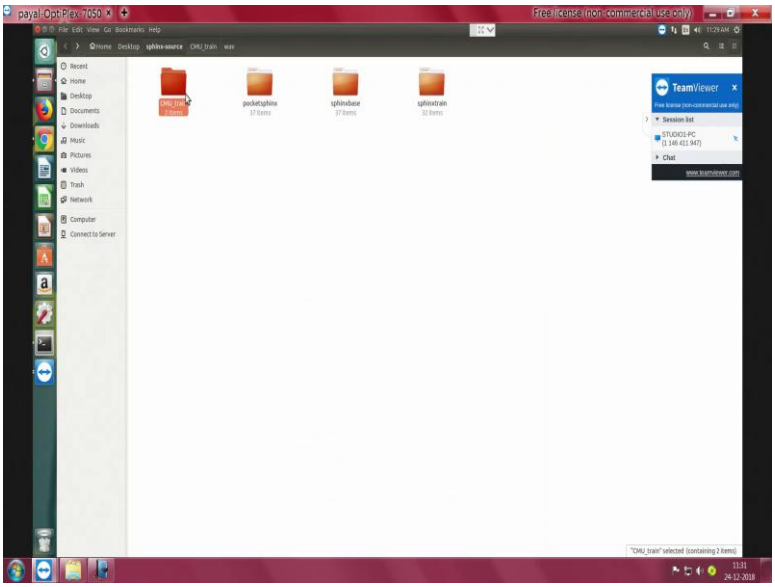
If you still further go down you will see you will see so, this pockets sphinx that you have seen is essentially for Android ok. So, that is the important thing. So, if you still further go down you will install sphinx train ok, you see that sphinx train sings sphinx strain essentially is for training the model ok. So, all the steps for training the model is out here. So, everything is available to you as tools you should just try and get some experience on it.

(Refer Slide Time: 02:28)



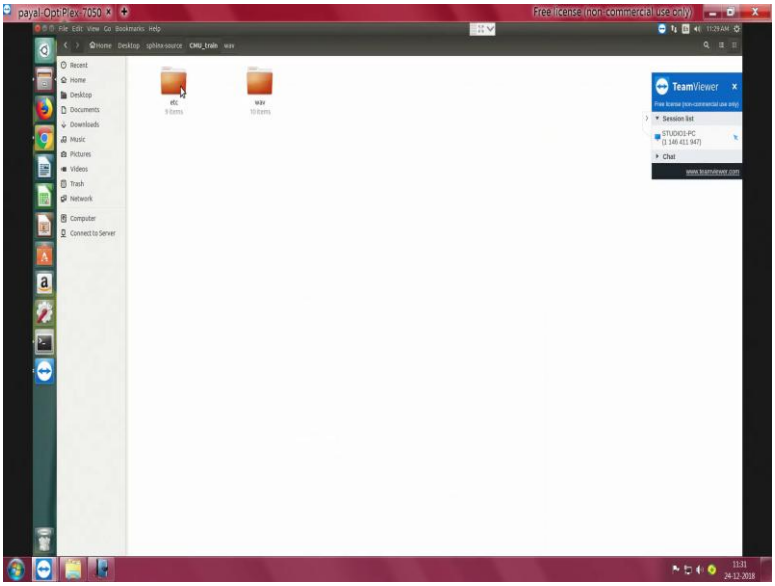
So, like this you could go down and follow all the steps which are down here installing CMU this is not required. I think you should only stop up to this point you should be ready to go now; hopefully this will be this was helpful for you. If you ran into issues or of suggest on how to make this better be sure to leave a comment, you do not have to do anything with respect to cmu clmik because, this is of no concern for us. We are looking at Android part this is step 1 ok.

(Refer Slide Time: 02:53)



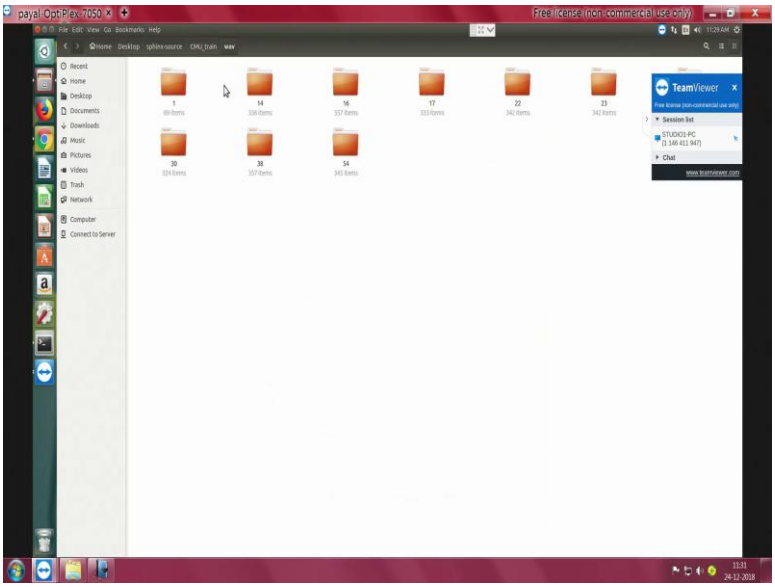
So, next is to train the acoustic model. So, what you should do is you should create a folder CMU train right inside this sphinx source.

(Refer Slide Time: 03:10)



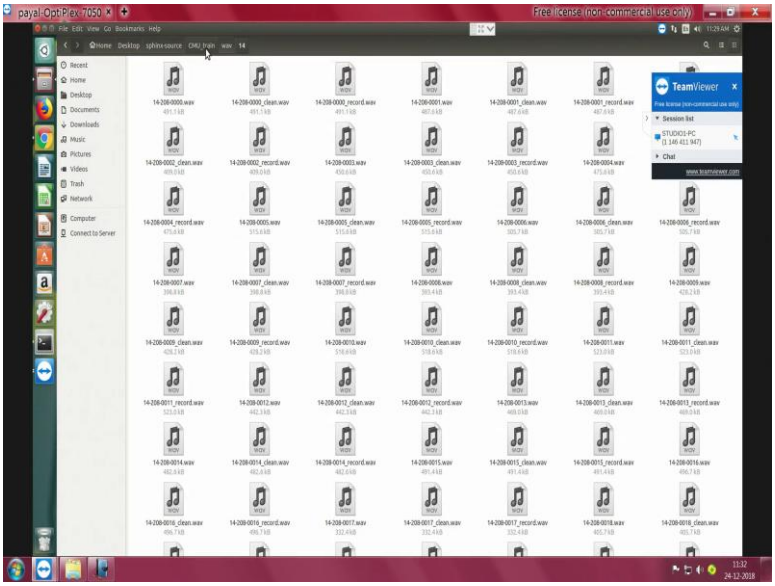
If you double click on that folder you will see etc 9 items and wave wav files 10 items.

(Refer Slide Time: 03:19)



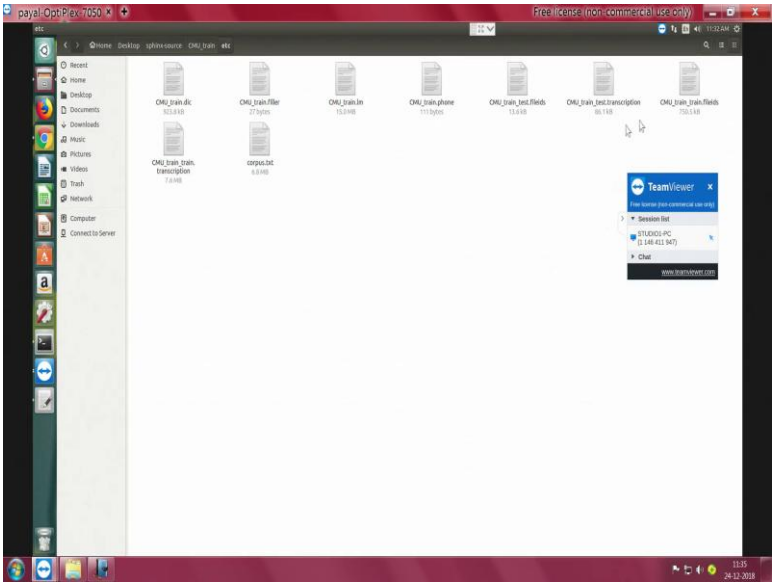
Double click that.

(Refer Slide Time: 03:21)

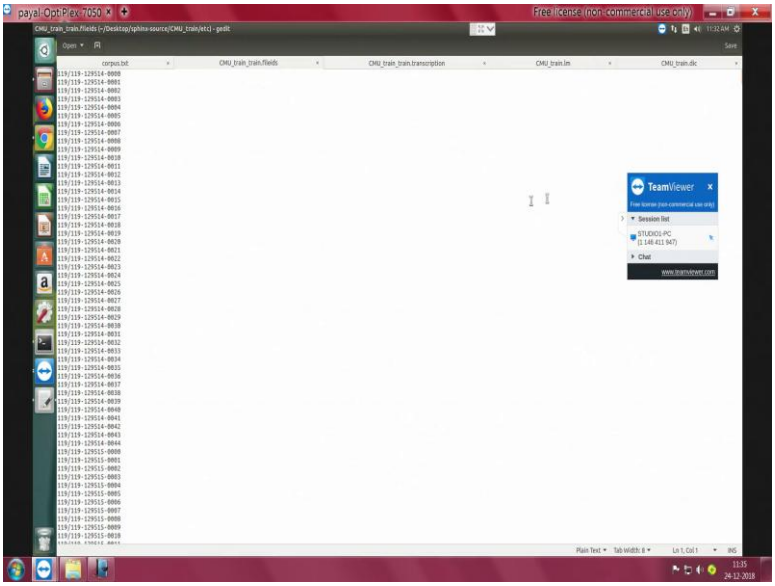


You can see all these recordings are available right. So, essentially these are all the audio files and there is a number right 14 208 0000 dot wav and all that. So, this is nothing, but the name of the file that is nothing, but the file ID as it is called, the name of the file you are given it a name. So, this is actually you will also see this in the etc folder ok.

(Refer Slide Time: 03:48)

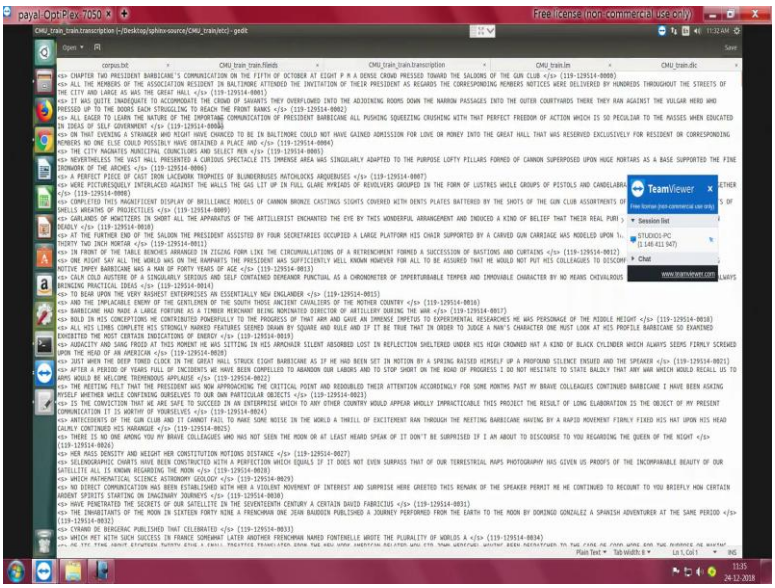


(Refer Slide Time: 03:50)



If you go there you will see that you see this the same numbers are showing up there and essentially the etc will contain the mapping of all the waves and wav files. And the corresponding file IDs are the file names of those wav files, there are also transcripts right.

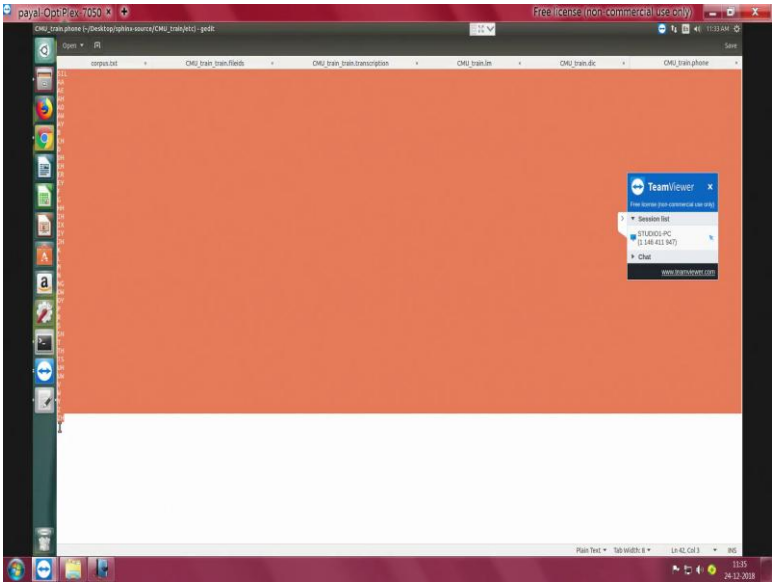
(Refer Slide Time: 04:11)



The transcripts corresponding to that audio clip is here, you can see all that essentially you need the wav file which is the sound file which has all the phonemes and all that

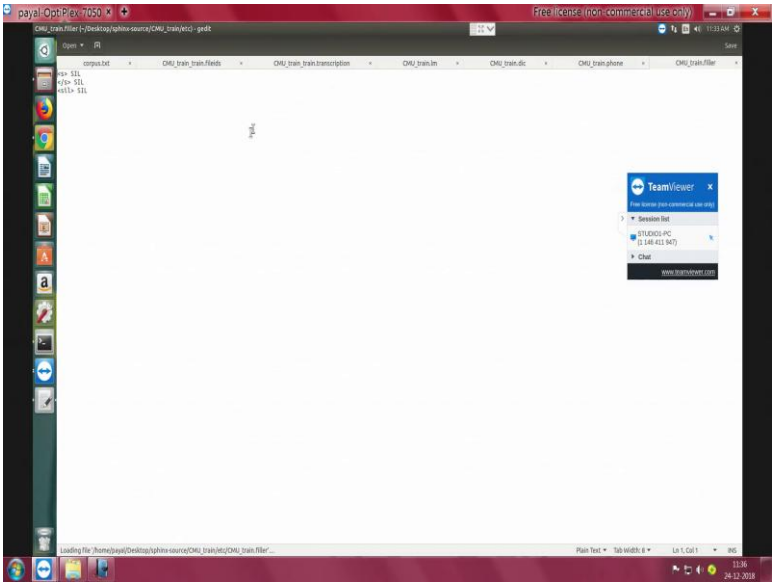
which we will not get into. And, the corresponding transcript of that each wav file is out here.

(Refer Slide Time: 04:35)



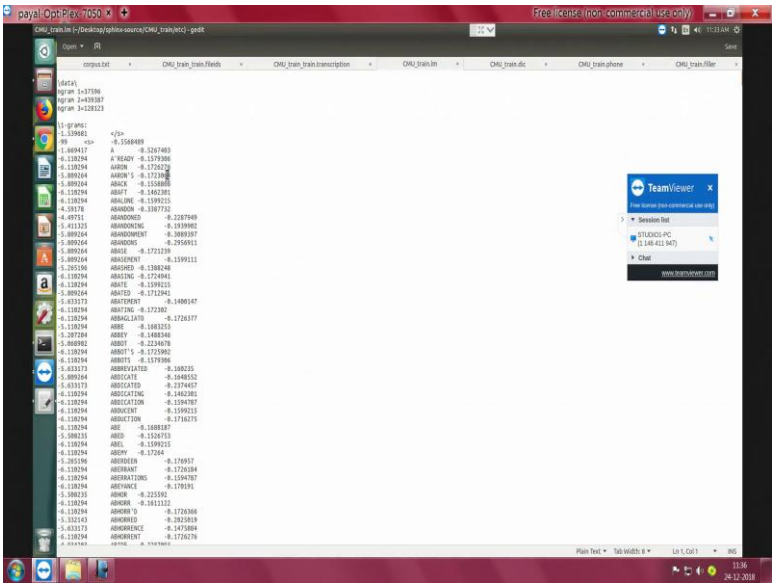
So, these two things are required essentially and then of course, phonemes right. So, look at the phonemes you see the phonemes here, you can see that 44 phonemes that we have you will have a listing of all of them out there.

(Refer Slide Time: 04:49)



Then there are also between what you talk there are also silence periods also called the fillers and that filler is also shown here and these are the fillers that you can see. So, these are the main things.

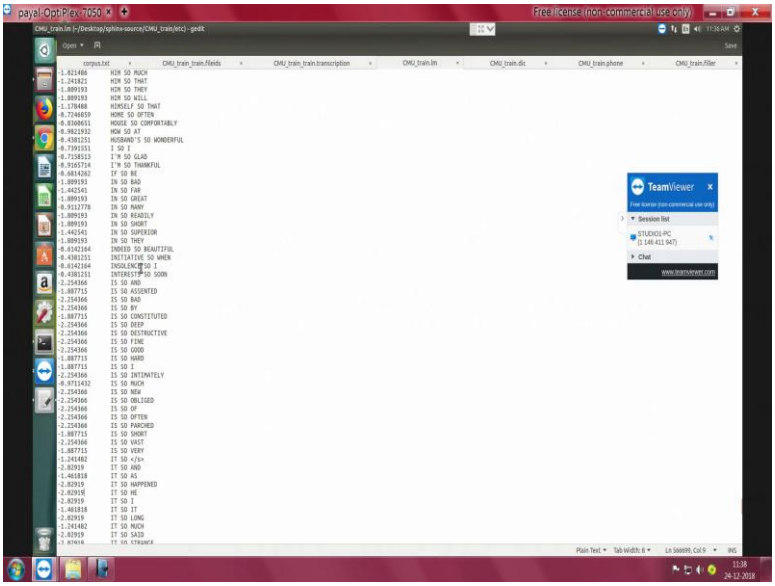
(Refer Slide Time: 05:06)



The next thing is you have to create the language model. So, you have words from the words you have you have; so, you have words and from words you have sentences. Essentially you need to have a dictionary which essentially talks about the complete corpus and from there you should create the language model which essentially is nothing, but the list of probabilities its nothing, but the probabilities that are associated.

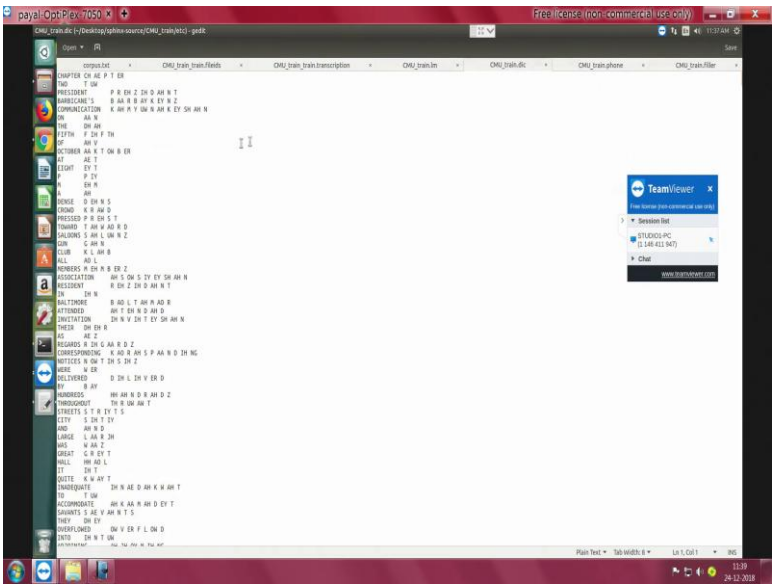
So, our starting point is the corpus which was generated from the wav file from speech you got this corpus. The corpus is essentially used to create the language model and also that dictionary ok, now the language model when you talk about it.

(Refer Slide Time: 05:52)



Let us see the language model; this essentially is a trigram weights of these which have been generated from the corpus ok. So, that is the first part you can see there is a weight on the left side and on the right side is the words essentially which is typically 3 words that are there here you can see. How so comfortably, how so at, husbands so wonderful. So, it is all nothing, but the trigram model that is the first part.

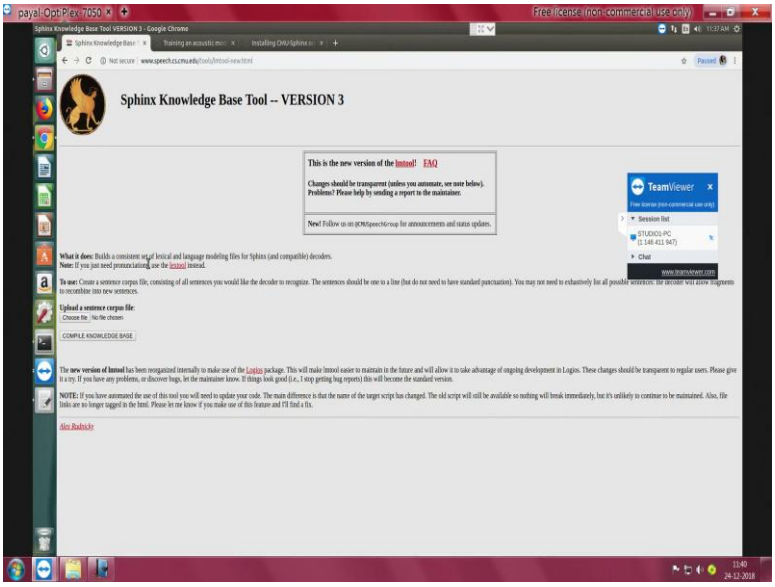
(Refer Slide Time: 06:34)



Then you have to somehow get to the phonemes from here right, from the word you have to get to the phoneme. So, you say president take the word president yes P R EH Z IH

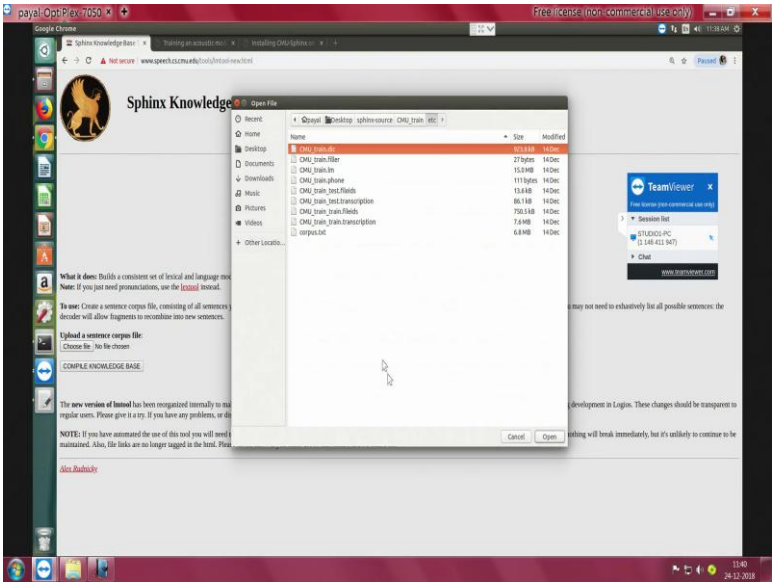
DN T that is what is president. So, you are trying to get for each word you are trying to get the phoneme which is essentially nothing, but the dictionary.

(Refer Slide Time: 07:00)



So, how do you get to how do you get to the language model and the dictionary; for that what you can do is you can feed the corpus on this icon.

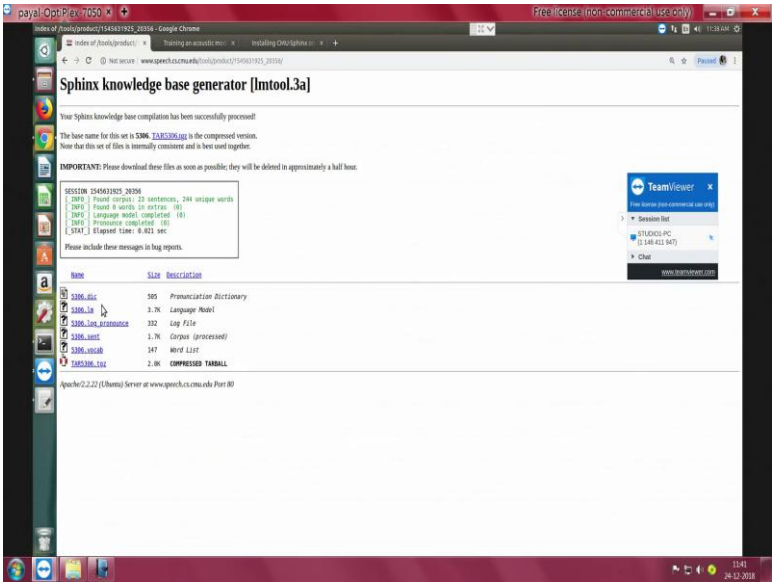
(Refer Slide Time: 07:17)



Just say choose file and if you choose the file you should be able to feed it into this website you have www dot speech dot cs dot you can look up that URL; you can you can do choose file. So, when you do choose file you will see a set of files and then you

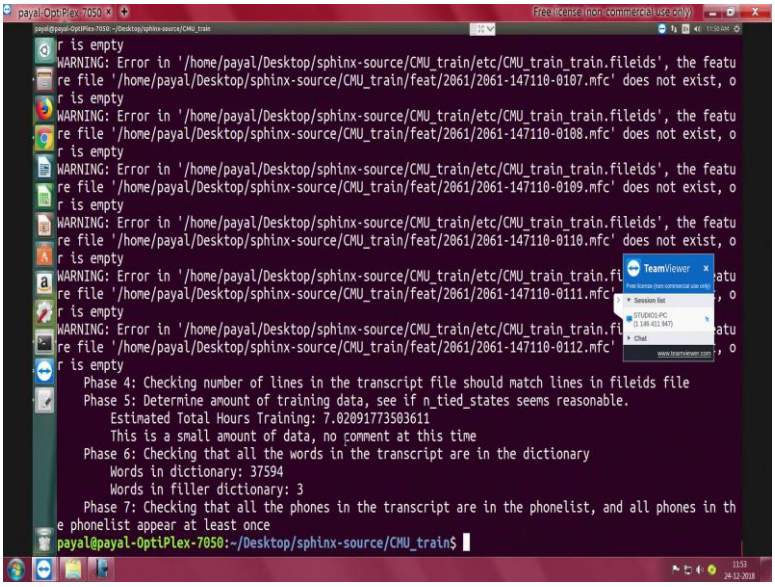
choose the file of interest which is NPTEL Corpus and then you say compiled knowledgebase.

(Refer Slide Time: 07:46)



It will generate you can see the dictionary, the language model that this is the dictionary, the language model which is also trigram right. And what else does it create? It creates the word list ok. So, look at what where we are at this stage? We are at the, we are inside the etc folder, if you look carefully you have 9 files right; you have 1 2 3 4 5 let me show it to you ok. You have 1 2 3 4 5 6 7 8 9 files.

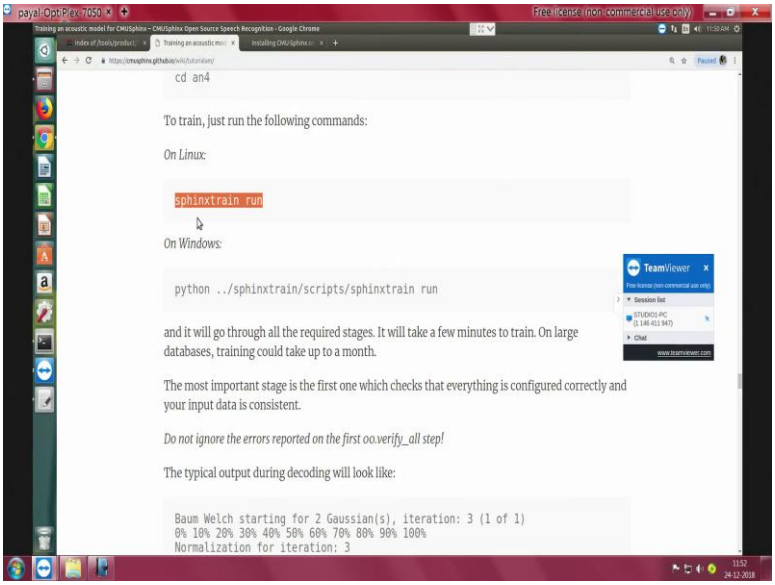
(Refer Slide Time: 08:42)



In order to set up our database we need to run that command line which we will do now and if I run this sphinxtrain underscore t CMU train setup it does it does something setting up the database for training. And, now you see 2 files which is the cfg and the params file alright. So, this is the 2 files that were created were the feature parameters which is under shown here as feet dot params.

And, the other is the sphinx underscore train dot configuration file; this needs a small modification. The language models that we have essentially is dot lm ah, but it would come with a dot lm dot DMP you have to make changes in 2 places. So, you remove it here you can do a simple search in order to replace dot l m dot DMP with just the dot l m and then you store it back. So, in the 2 places you modify and then store it back.

(Refer Slide Time: 09:54)

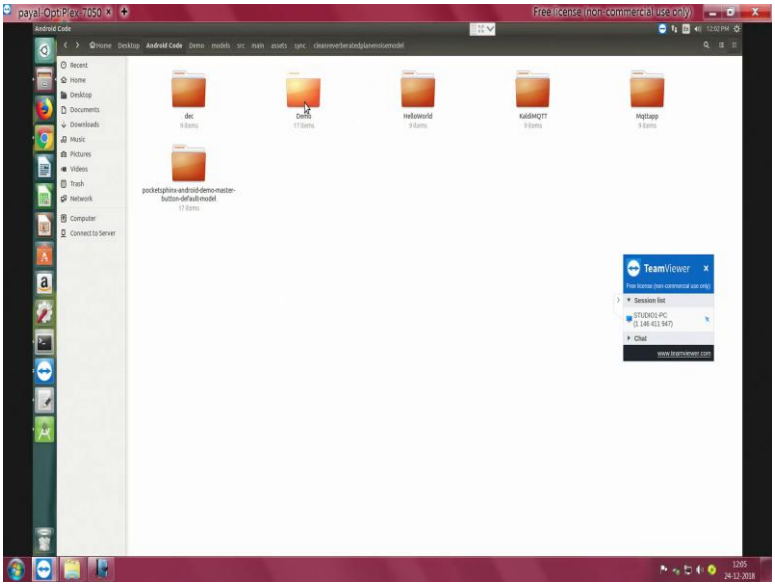


Then the next step is to start the process of training, there is a simple single command for you to do and that you can do with sphinx train run. If you do that let us see you do that you paste the command there and then you do a run. You can see that extracting features from segment starting at part 1 of 1 it is doing a few things and then ultimately you will get a acoustic model. When we talk about speech to text I mentioned to you that there are two aspects to it; one is there is something called training and there is something called testing that we have to do. So, when we come to training that I have shown you how the process of training a model is completed ok.

Now, you have to do the testing part; step by step how do you want should go about doing the testing part. Before, we get into the detail I must tell you a few things upfront. Normally, when you do training you do training with clean speech you can download from Libre speech for instance, you can download Libre speech create the language model and the dictionary from for clean speech and do the training part. The testing part can be essentially what you are looking at, the training part included sections like clean speech which is from Libre and for testing the in flight data was taken. So, the lm the language model and the acoustic model specific for the testing part is specific to the application of interests ok.

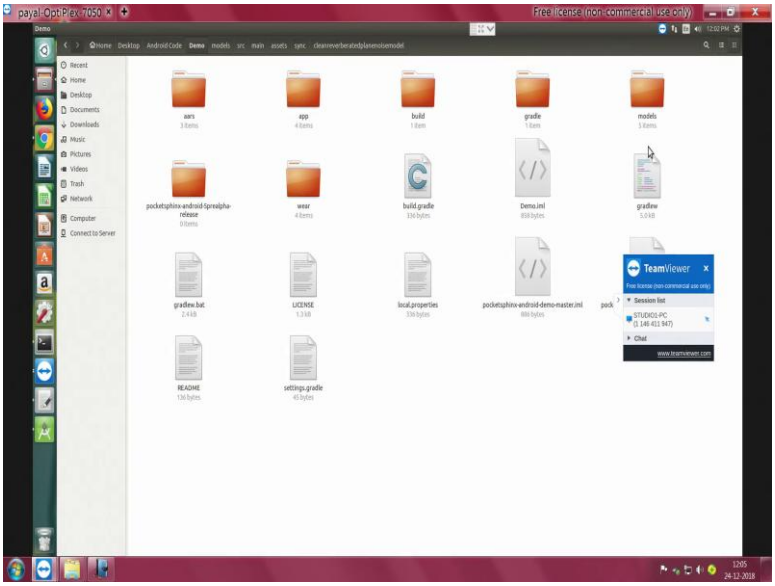
So, this you have to differentiate clearly when you do when you do your training and testing. So, how good the acoustic model you create actually weighs heavily on the fact that it has the ability to decode correctly during your testing phase. So now, let us look at after the models train what are the set of steps that you have to do, you have to follow before you can deploy the app onto the phone.

(Refer Slide Time: 12:22)



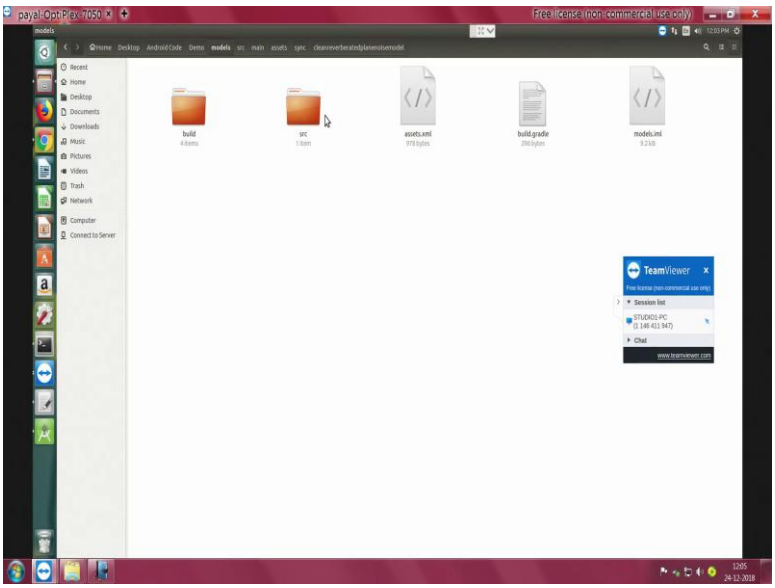
So, the first step is so, let us go to Android Code there you will see demo.

(Refer Slide Time: 12:25)



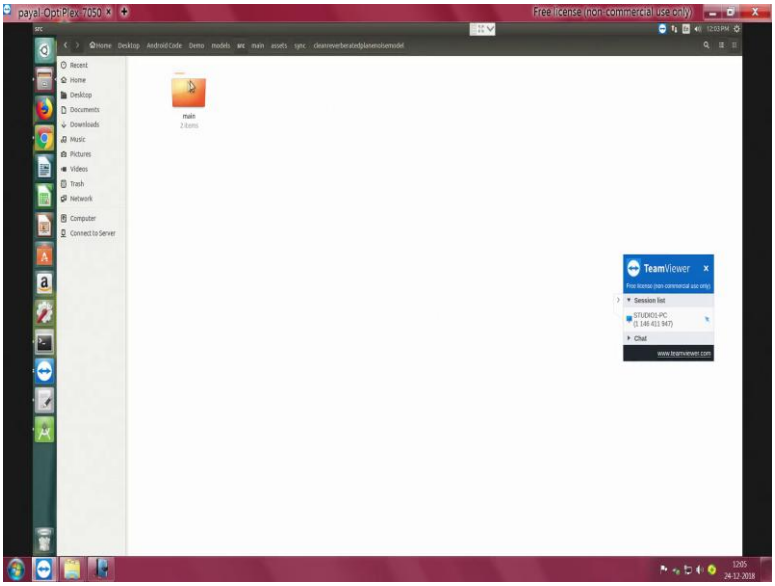
Inside the demo folder you go at to the app and the models, you will see two folders one is called app the other is the models.

(Refer Slide Time: 12:38)



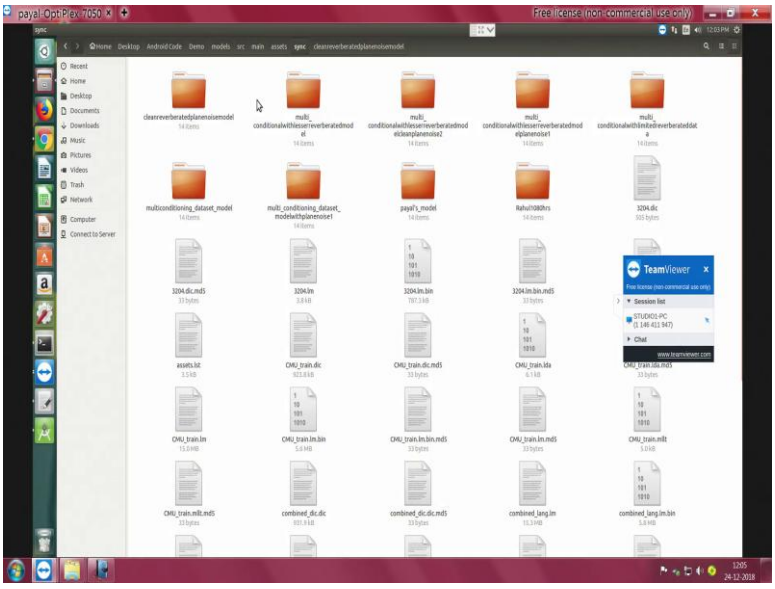
So, if you click on models.

(Refer Slide Time: 12:40)



You will see that the resource.

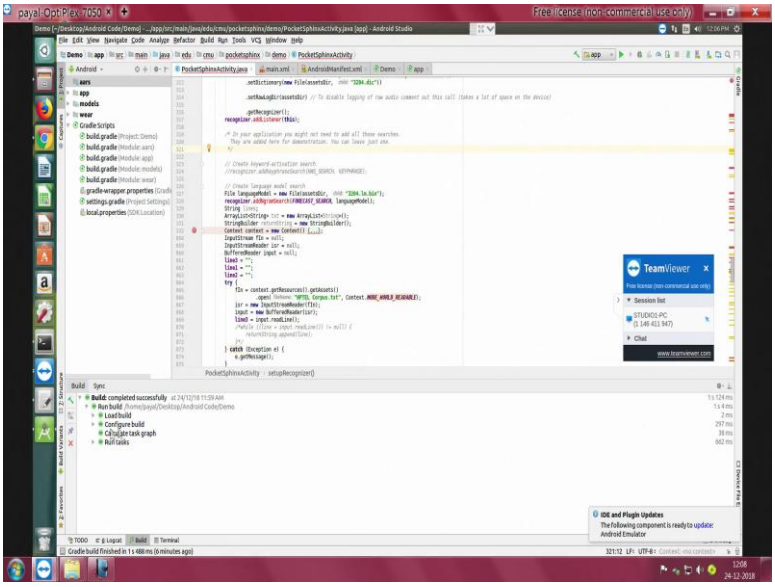
(Refer Slide Time: 12:44)



Then essentially you have the clean reverberated plain noise model. So, here what you see is the clean reverberated plain noise model that you see here is nothing, but the trained model that is the model that got trained. The CMU train underscore dic which is the dictionary and the language model for testing application right. So, these three together form the part of testing ok. Now, you have the trained model, you have the lm

and dictionary ready which has to be deployed onto a phone for that you need an app correct. So, you have to open android studio.

(Refer Slide Time: 13:34)



So, let us open android studio and then link these three which is the trained model and the lm training lm the sorry the testing lm and the testing dictionary also has to be part of the app. So, you can see here the place where he has highlighted the clean reverberated plane noise model is shown and same way the dictionary is also shown and also the language model is also shown. You feed all these three, you put all of them link them together and then click on build you should you should be able to rebuild the project. And, that essentially creates the app which you can choose to download on to the phone and then try your testing.