

Digital Speech Processing
Prof. S. K. Das Mandal
Centre for Educational Technology
Indian Institute of Technology, Kharagpur

Lecture – 41
Speech Based Technology Development For E-Learning

So, there is other kind of application. Let us we discuss, we have discussed about that ASR, we have discussed about those synthesis. So, I am not going details of ASR, I have not detail discussing that HMM model better be algorithm because I think that required another courses so, but I give you the glance of that ASR and the research issue in ASR. So, there is lot of research is going on, non convention the beyond statistical model today there are lot of researching find in deep Dural neural network base ASR development, then you find that manner base that at speech attribute base ASR development lot of research is going on because the ASR problem on statistical mode night not yet solved, it can save sub the purpose for certain extent, bit is not solved.

The human speech recognition problem or you can say you cannot solve that extend where we can use it in many cases or you limited application it is. But when the application is large scale up, then there is a some problem even the resource the resource constant problems are there if the ASR; the language which is less resource we cannot develop the statistical base ASR system because there will were error rate will be very high and ASR performance is measured on word error rate.

(Refer Slide Time: 02:05)

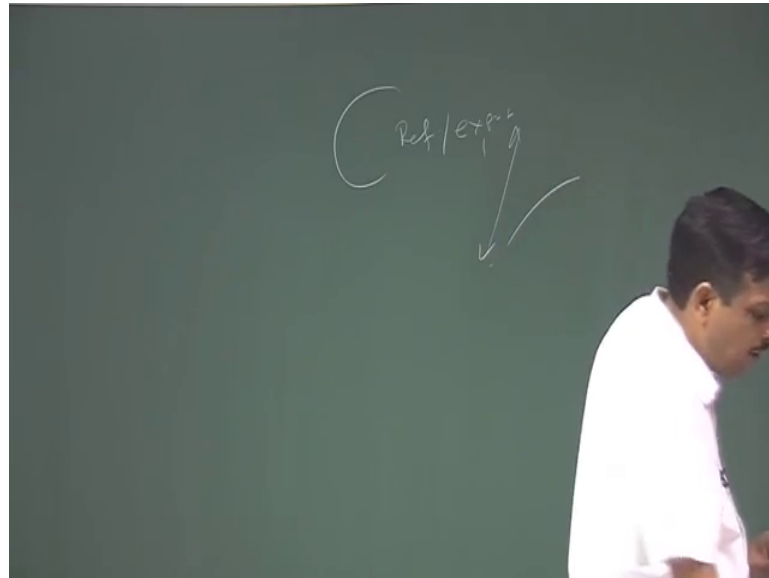
Area of research			
❖ Computer Assisted Spoken Language Learning:			
Objective: To Develop a Computer Assisted Spoken Language Learning tools with correct language pedagogy for faster acquisition of second language speaking			
❖ Accent Conversion:			
Objective: Transform the regional/foreign accent of a source speaker to have similar characteristics to that of another accent			

So, I have not going details on that thing. So, another area of research may be speech is may be spoken language acquisition, suppose I want to teach or I once you that I want to teach the English or I can say second language acquisition or spoken language acquisition I want to teach English. English like that problem, I can describing that the government of West Bengal decided that English will be taught from primary level earlier it was started from class 5 or 6. Now it destroyed that English can we taught from class 2 or 3 level.

Now, if we visited many many school in village area, you see the pronunciation different the teacher's pronunciation is not also not acceptable. So, if I want that we want to spread the spoken English or I want to spread English literacy in village area. So, I want to use the computer because English is a second language that English is not their mother tongue. So, if I want to taught the second language or English spoken language in every school; can we developed E-learning tools or E-learning platform by which we can taught all the school children English pronunciation because if the pronunciation is not correct then you see the spelling mistakes will be happened. So, spelling and pronunciation orality and because the orality is first if you if you read the history that before written script was came the orality is exist.

So, orality was the oldest things for communication. So, first I have do oral then we will go for the written. So, orality; so, if I want to thought the spoken language that spoken English second language acquisition in English.

(Refer Slide Time: 03:49)



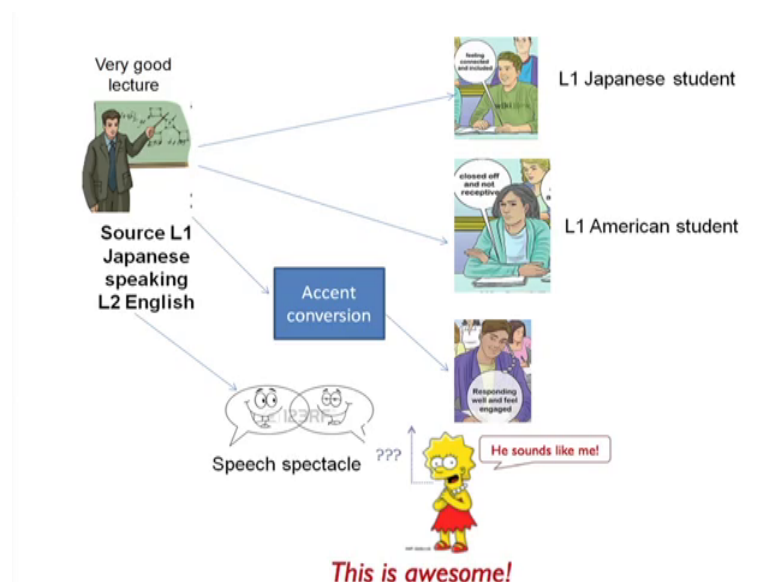
So, can we developed a systems where I can say that there is a lets reference speech reference or expert speech; expert speech is according in computer means in front of the computer the child speech child spoke there is a microphone as pronounced some English word when it compared with the expert page. So, 2 parameters are there segmental and super segmental is compare with the expert speech and told me which area I have to stress which area I have to improved you he may say ok.

This word lets the pronunciation is zoo then say you are not pronouncing zoo either pronouncing joo. So, zoo make it joo. So, I can many times, I can listen and try to pronounce once its match congress to a yes your achieving it. So, that way I can trend the students that match their pronunciation with experts speech and computer will told me where it is different; no, no, no, no, it will be look like this. So, like that music listening a or an you can say the acquiring of music from a guru that you use you are saying a song then guru said no, no, no, no, no, you upper no it is not touching that is gurus perceptual judgment.

Now, I replacing that guru by a computer judging, I have pronounced zoo if I initially pronounced zoo computer is no, no, no, no, it is a zoo, then I say joo, no, no, no, it still it is zoo then I can say zoo is achieve.

So, that is one of that finest application you can developed you can take a project that let us start with spoken language I want to teach spoken language through computer. There may be you can take the help of the language pedagogy which kind of words has to be taught in this level. So, language pedagogy is there pedagogy will be there in the language background language pedagogy. The second problem is that I am saying accent conversion.

(Refer Slide Time: 05:55)



If you see this slides, suppose a Japanese speaker; Japanese professor taking a lecture lets Japanese professor taking a lecture; I am not saying it that I am the Japanese professor taking lectures.

Many of us; since we are not a custom with the Japanese accented English, if I sit in that class we may not understand that accented English or clarity of that accented English to me will be recognize that clarity will be very difficult. So, I am not clear what professor want to communicate with me. So, what will happen I will not understands his lectures completely means I will understand, but not completely I want to a; I customs with that lectures. So, what will happen after few times I will switched off my; so, I do not want

try to understand that things, but same lecture if it is listen by a Japanese students he a custom to the Japanese accent he enjoy the lecture.

Similarly, if I taking a lectures and if it is listen by a Japanese students. So, what will happen he is not a custom with this my English accent. So, he will not enjoy my teaching because he is not 100 percent understand he is not 100 percent you can say attentive to my lectures. Now similarly, suppose I put a device in middle of here whatever I say it is my accented English I am not converting the language I am saying my accented English and when the listen by a Japanese student; he listen Japanese accented English understand. So, I am speaking in Bengali because my first language is Bengali. So, I have Bengali accented English I was speaking.

The Japanese students sitting in my class listening it Japanese accented English similarly it is very common if I South Indian professor. So, suppose a Malayalam professor is taking in my class in will first language is Malayalam, he is taking a class his English will be Malayalam accented, when I am listening that English I say lot completely it is not that clear, but I want, but suppose I invented device which will take the same English and converted into what I want to listen on air to one pollution is accent correlation ultimate dream is that if you think about the spec spectacle think about the spectacle.

Everybody spectacle is personalized; that means, as per my spec power spec condition of the eyes spec is described. So, I want to see what I want; that means, as per as per my either spectacle is designed. Now think about speech spectacles you are speaking to me, but my accent is different. So, I cannot understand lets I put my device. So, is convert that your accented English to my accent oh I understand fully. So, ultimate aim is to develop a speech spectacle and if I; it is awesome that suppose somebody is giving an American professor giving a lecture and I am second language is in English and I cannot understand what is saying then I where my speech spectacle I put my device there and convert in my accented oh I understand.

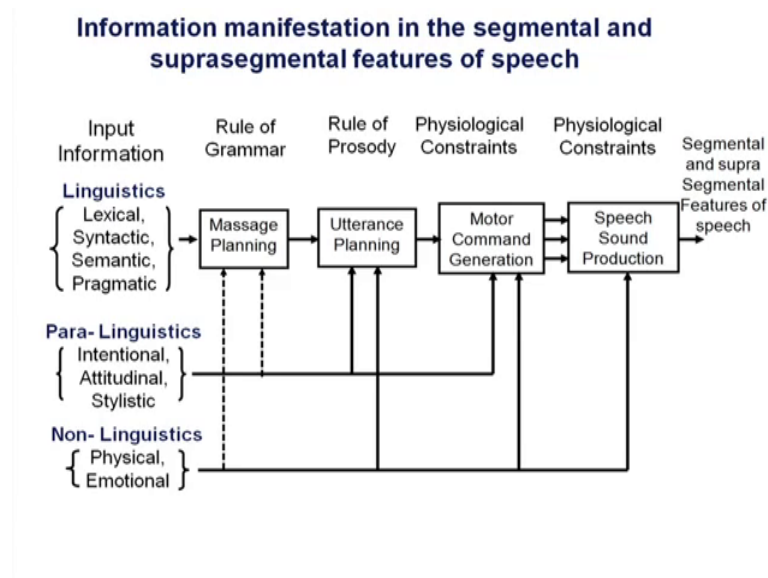
So, this is ultimate dream then can I develop this speech spectacle can I develop this. So, this is a open ended research problem people lot of peoples are working in this see the area you can take up any students can take up in teachers can take up this project and try to develop because it is possible speech has 2 parameter segmental and supra segmental. That means, array and speaking that you can say that prosodic modification accent and

contextual the segmental modification if I am. So, there is a lot of study required from segmental study and supra segmental study. So, there is a sees in this area that the difference between American; native American English means whose L 1 is American English and a spoken a Bengalis people whose L 1 is Bengali.

So, there is a there is a thesis on that and what are the segmental and supra segmental difference is exist is it acquirable or it is not how it can be convert all thing are discussing those thesis. So, if you are interested you can search the thesis on this. So, here there some thesis is done by my guidance one students as done that. So, he is compare or you can find out the segmental and supra segmental difference between L 1 American English and L 2 Bengali English. So, whose L 1 is Bengali, but L 2 is English. So, there is a there is Ishka project if you see there is a lot of study is going on non English is a word language you can say that. So, there is lot of study on English of def different countries English.

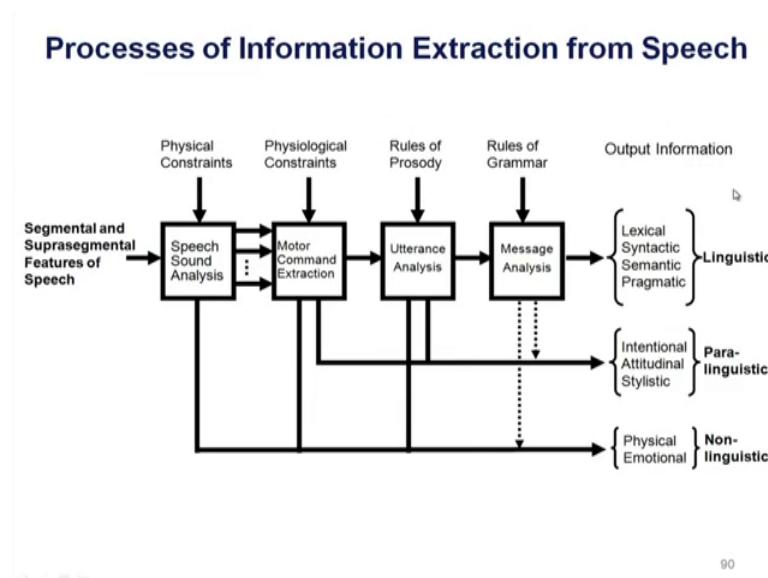
Malayalam speaking English; so, there is a single you can say the Thailand people are speaking in English. So, all kind of things are which are going on, but this is my ultimate dream in one day I want to develop the systems which will convert that lets English it can be used for any pair of language suppose I have suppose somebody is speaking in Bengali which is a second language suppose Hindi people are speaking Bengali, oh, it is strange. So, what I will do I put a converter there I want, but way I wanted to listen. So, that can be that can be apply for any pair of language that is open ended challenge. So, I can say the speech spectacle I can develop a speech spectacle personalize my speaking what I want to listen.

(Refer Slide Time: 13:03)



So, that is there. So, I am not going details orality, there is a History. So, now, the how is it possible yes it is possible if you see the when you produce the speech there is a segmental and supra segmental speech also this speech. So, there is a linguistic, paralinguistic and non linguistic driven on this information segmental and supra segmental is developed if you say processing of information extraction.

(Refer Slide Time: 13:22)

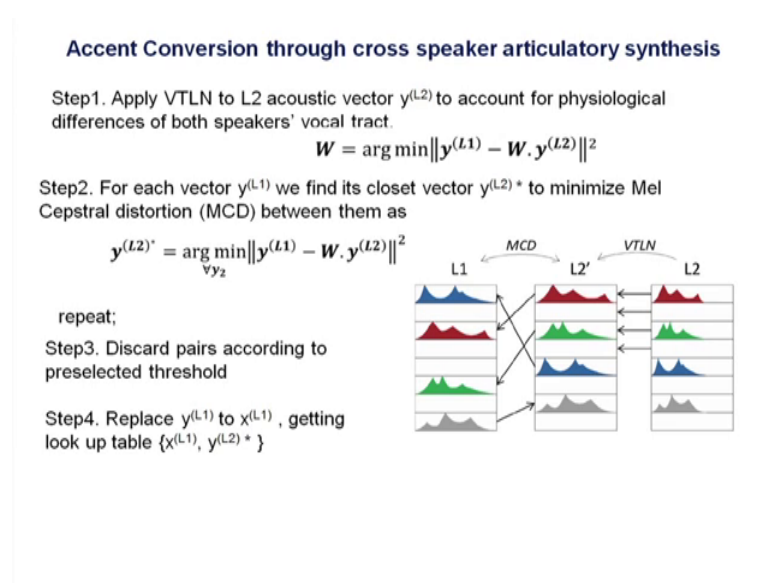


So, here all speech which is segmental and supra segmental acoustic signal is provided here. So, physical constant speech sound analysis then physiological constant motor

common excitation physiological prosody analysis utterance analysis rule of grammar message analysis then output is linguistic paralinguistic and nonlinguistic for.

So, all I can do from the speech parameter also. So, that is open ended research we are still doing on that things. So, there is a some study who is taken from the net there is a some accent conversion works are there research trend there that there is a accent conversion through dia-concatenative synthesis technique then accent conversion through articulatory synthesis technique, then there is a accent conversion through cross speaker articulatory synthesis technique and here we have we are proposing some technique which still in research end that we are proposing that since which is a combination of segmental and supra segmental features; we try to change the segmental supra segmental features using deep lib network or deep neural network kind of application can we do that still it is in research stage we are doing it.

(Refer Slide Time: 14:15)



So, this is end of that speech application what I have discussed; I can discuss more application in next week like this way prosody modeling and speaker identification and speaker verification I can discuss the preliminary. So, you have a rough idea what is speaker identification as speaker verification. So, that that idea is there and I will give you some idea about the prosody modeling which we are doing in our lab that speech prosody modeling for Bengali language long time who are working in this area. So, there is a HMM base things, I will discuss there will a lot of papers also there. So, I will

discuss the some research issue on those things. So, I am not details discuss about that that side that side of that sub computing side. So, you can read the sub compute sizing and try to co relate with this speech application.

Thank you.