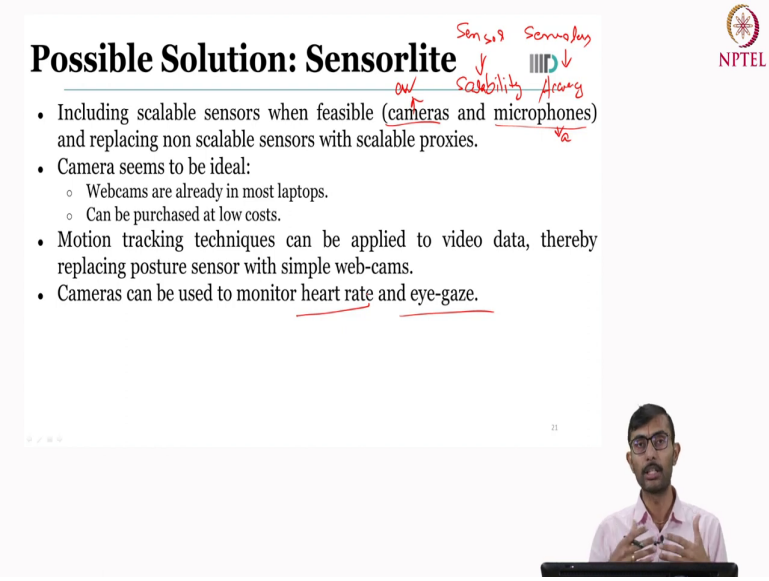
Affective Computing
Prof. Jainendra Shukla
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Week - 10
Lecture - 01
Part - 02
Emotionally Intelligent Machines: Challenges and Opportunities

(Refer Slide Time: 00:22)



So, we saw so far that if we go for a full sensor based approach dedicated sensor based approach then there is a problem of the scalability right. So, let me just write it down for you. So, if you are going ahead sorry if you are going ahead with the sensor, then dedicated sensors they will require you will face a problem of a scalability. And if you are going sensorless, then you will have a problem of accuracy.

So, maybe the accuracy you are somehow making a trade off with the accuracy sensorless right. So, now; so, there could be a possible solution where we are which we are calling it as a sensor light solution. So, sensor light solution means we are going to use the scalable sensors whenever it is feasible; such as for example, you know cameras and the microphones.

They are very very widely available sensors which can allow you to capture the audio visual modalities and only the audio modalities, and that is one approach. And then at the same time what we can do? The non-scalable sensors can be replaced with the scalable proxies. So, for example, if you are trying to make use of camera can be very good option here and for example, because once that it is already available in most of the laptops.

Similarly, for example, it can be purchased at very very low costs and previous research has already shown we have talked about it that you can use the camera as well to for example, record and monitor the heart rate and the heart rate variability and related features. So, this seems like an excellent solution and other thing for example, that you can do you can also simply look at the webcam data and you can apply the motion tracking techniques on it.

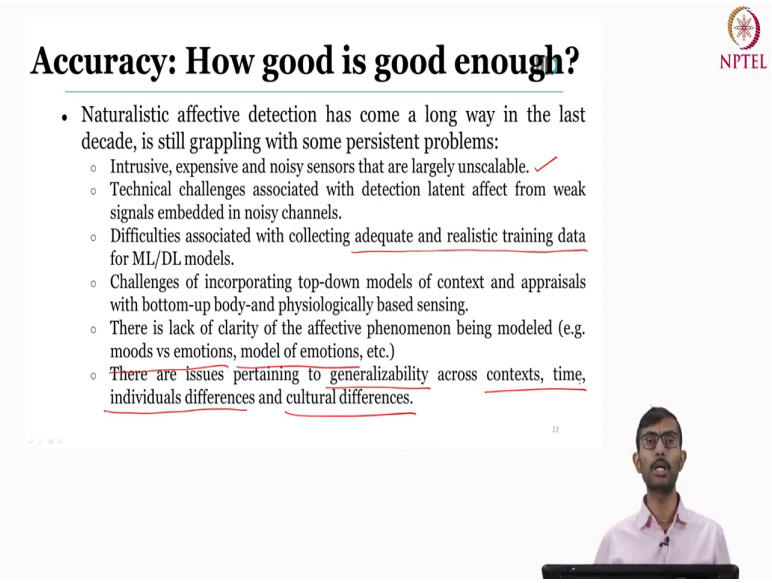
So, for example, you can do the posture analysis, gesture analysis all with the help of the 2D visual data; of course, you will need to work a bit more on the software side. But working on the software side is a bit more easier, it is more accessible than you know working on the hardware and trying to make the hardware scale. So, that for in this way you know like there is one type of possible solutions.

Other we already talked about that you know like the camera can only not only be used to monitor the heart rate and the heart rate variability; it can also be used to monitor the gaze patterns of the eyes. And hence it can also replace to certain extent the eye tracking devices and the sensors that we have.

So, that is the conclusion of this thing. If you have the possibility of going making use of the dedicated sensors, go ahead. Many times, your application domain will not allow it. Then you may want to use the existing sensors which are there available in your system to track the

behavioural data. Or for example, you may want to replace the dedicated sensors with the sensor proxies and hopefully it may work wonders for you ok. Sorry.

(Refer Slide Time: 03:39)



So, that was about the sensor and what to do with the sensors that are available to us. Now, let us talk about the accuracy as well. So, in general there is a we know that you know the a naturalistic effective detection has seen a lot of research and it has been improved a lot. But of course, there are lots of problems that are associated still with the affective computing domain and in general the research and the development that is there.

One thing we already talked about it there most of the time the sensors that we require they are intrusive many times it could be expensive as well; many times, it could be noisy as well and more importantly they are not scalable. Other thing that we see that you know technical

challenge if you look at talk about the technical challenges, then the detection itself can suffer from the weak signals that are embedded in the noisy channels.

So, many times for example, you are trying to look at a particular data and then, but that data itself is surrounded by lots of noise around it. For example, maybe you are trying to capture the emotions in the voice, as simple as that you are trying to capture the emotions that are there in the audio modalities of a user.

But now we are no more in a lab setting, we are in a naturalistic setting where there is a lot of noise around it. And hence your target users, voice data is getting embodied is getting surrounded by lots of noises that are around there. And hence it really becomes very very challenging to segregate that data of the user, voice of the user and do some analysis on the top of it.

One thing also that we have seen even though we did not talk about a lot in detail about the machine learning and the deep learning algorithms itself. But most of the time what we want to do whenever we are talking about the fatigue computing it is going to rely heavily on the machine learning and the deep learning algorithms. And whenever we talk about the machine learning and the deep learning algorithms, it is no surprise that the need a lot of adequate and realistic training data in order to make lot of sense.

And many times, what happens that emotions data associated with the emotions may be very very difficult to even capture to annotate. And hence in turn our machine learning and the deep learning models may suffer and their accuracies may suffer. So, other thing for example that may happen that when we are talking about the affective computing most of the time, we are just looking at the transient emotions.

But now, if you want to incorporate the context and the appraisals as well into it, then this becomes a bit troublesome. Because you know if you want to incorporate the context and the appraisals and let us say you know the users beliefs, desires and intentions around it then it

can be become very tricky, because then you will have to be able to you know track lots of different things and which may be very very difficult here.

In general, you know whenever we are talking about the affective computing. Also, there is a common mistake that the researchers do that the developers in the community do and they are doing that many times they are taking replace using one for the another and another for the one. So, for example, whenever they are talking about the model of the emotions, they are not able to discriminate properly between the categorical models or the continuous models.

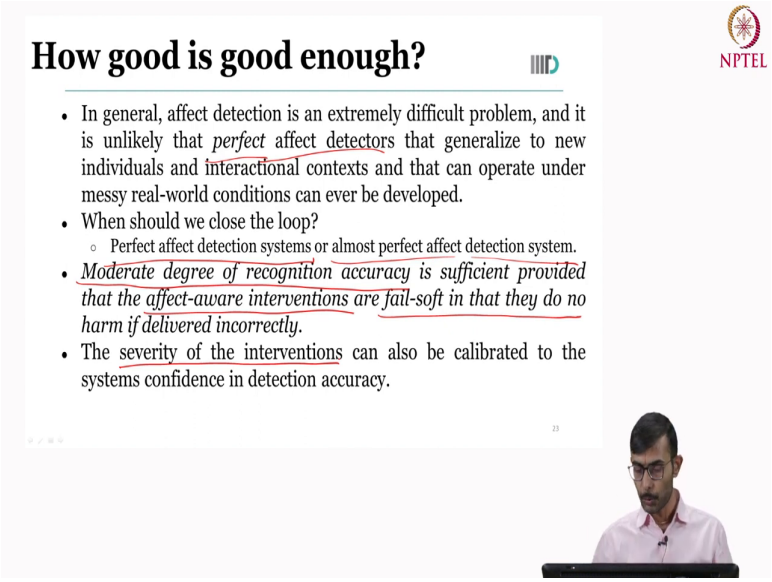
Similarly, for example, many times the mood versus emotions is not being differentiated. So, you may refer to mood which is you know long term emotions or maybe you are referring as a transient emotions. But you are just you know using the one for the another and hence you are notable to differentiate these things and hence there is a lack of clarity on your side.

Nevertheless, even after this if you are able to do some recognition and you are able to make use of some monitoring then what happens that the generalizability itself it becomes a big issue and generalizability in general for the machine learning and deep learning algorithms itself is a problem. But more so with respect to the emotions and then the affective computing it is a bigger problem.

Because, you may want to look into the individual variability, we talked about this; you may want to look at the cultural variability, cultural differences. For example, the death of an individual - at the time of the death of an individual the way it is being expressed in Indian communities is very very different from the way it is being expressed you know the lamentation or the sorrow that is being expressed in the outside community.

So, of course, there are lots of cultural differences that are surrounding it. Of course, then context, time etcetera they also play a lot of role and so, generalizability becomes a big issue on especially when we talk about the affective computing models.

(Refer Slide Time: 08:47)



So, the question that in general we want to ask now is this that ok, what type of accuracy or to what extent an accuracy is a good accuracy with which you know we can go ahead and deploy the system and we can make it work. So, in general we already saw that affect detection itself is a very tricky problem.

And we can very confidently say that ok it is going to be very very unlikely that very soon we are going to have a perfect affect detectors you know that are not only able to you know do 100 percent classification of the emotions of a user in real time. But also they are able to generalize well to the new individuals to the interactional context and in you know very noisy situations that are around us.

So, it is not going to happen anytime very soon. It will require lots of advancements not only on the hardware side, but also on the software and the machine learning and the deep learning

side from us. So, the idea is ok, when should we start taking the information that we are getting from the emotion recognition systems and build on the systems that can adapt to it or in general when we want to close the loop.

So, there could be two possibilities now, whether we can go wait until there is a perfect affect detection system in order to build a adaptive system on the top of it which is going to work perfectly fine. Or we can just go for not so perfect affected affect detection system and we can try to make it work.

So, the thing is here since it is a very very tricky problem it may take a lot of time and resources to arrive to that situation. So, even if we are able to get a moderate degree of recognition accuracy. And what is a moderate degree of recognition accuracy for the emotions that depends on them situation to situation and domain to domain and as per use cases, we will talk a bit more about it.

But the moment we are able to get or you are able to get a moderate degree of recognition accuracy, we believe that should be sufficient to create the affect aware interventions. So, affect aware intervention means, adaptive interventions we just saw about some of the examples some of them.

But of course, we have to take into account the fact that there should be fail soft. Fail soft in the sense that they should not do any harm let us say even if they are doing some adaptation which is based on the incorrect classification. So, that is a very tricky suggestion that you want to look into it and of course, now for example, when we say that the moderate degree of recognition and then accordingly the severity of the interventions.

So, of course, you want to make an adaptive system, but since you are making an adaptive system which is not entirely, which is based on not so perfect affect detection systems. So, you may not want to put a lot of hard confidence in the severity of the interventions. So, you may want to take it up in the pinch of salt that ok I got a particular classification.

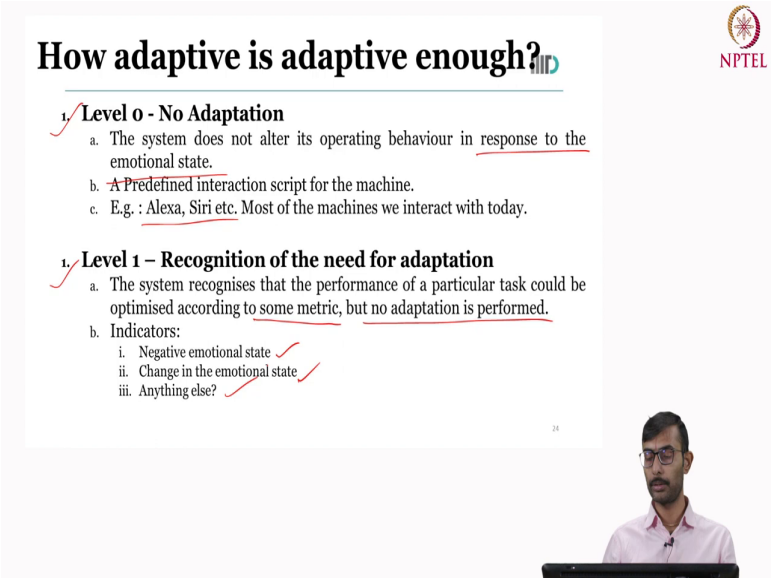
I got a particular emotional state of a user based on whatever sensors or the models or the software that I have deployed. And there is a possibility that it may not be so correct. Hence, the adaptation that I am going to do is going to be accordingly very not so severe and could be a moderate. And then it can be calibrated to the extent you know that I am able to put the confidence in my detection accuracy.

So, for example, imagine I am I can come in my particular use case, I am able to use make use of the dedicated sensors, dedicated hardwares. And I am able to use instead of the art algorithms and I know the detection accuracy is good or is almost perfect in my case.

So, then maybe you can put more confidence in the system and accordingly the adaptations that you are making maybe you can put more confidence in the adaptations as well and accordingly so on so forth right. So, this is a very good important thing to understand that you need to take into account the system that you are building.

And then the use case that you are building and accordingly you need to define that ok what could be a moderate degree of recognition for you, what could be the severity of the interventions that you want to go ahead. But bottom line is you may never want to wait for a perfect affect detection system in order to start building a adaptive system ok; so, that I hope is a bit clear.

(Refer Slide Time: 13:27)



Now, having talked about ok the sensors, having talked about that what is a good accuracy, what enough what accuracy is good enough; now, let us talk about a bit on the adaptiveness right. So, the question that we want to answer ok what should be the adaptiveness of the system and this is a very tricky question again. And the severity of the adaptive interventions as we said already is determined on the basis of the confidence that you are able to put into the system.

But let us see what are the different levels of the adaptation that we can have in the system. To begin with for no surprise, we can have a level zero adaptation which we also call it as a no adaptation at all. So, in the no adaptation at all what it means that we do not expect in the system to alter its behaviour in response to the emotional state as simple as that you know.

Whatever is going to be the emotional state we are going to monitor it, but we are not going to do anything about it, we are simply going to take that information. You know for some other analysis purposes, but we are not going to let the systems behaviour altered by the response in response to the emotional state and that is what is happening for most of the systems that we have.

And for this what happens you know a predefined interaction scripts mostly you know is used for the machines, for the services that we use we just talked about the gaming examples. So, you know when you are sad the non-playing characters also around, they show a sadness.

But all this is very very pre defined script and it does not really know first thing it does not really know that what your emotion is. And even if you knows, it does not it decides not to do anything about it; so, this type of system is where there is no adaptation as all happening.

And most of the machines that we are happening interacting with the today or the services including for example, you know the voice agents. Such as Alexa, Siri, most of the machines are like that. Their adaptation is not based on our emotional state and many times our emotional state is also not being monitored; so, that is the level 1, sorry level 0.

Then comes the level 1. So, the level 1 basis basically you know what it does, it tries to monitor your emotional state number 1 and then it also tries to recognize that ok there is a need for the adaptation at a particular time. So, it tries to identified the time of the intervention or the need for the intervention. And, but of course, it just does that. It simply tries to identify, identify it is that ok there is a need for the adaptation, but it does not perform any adaptations at all.

And for example, you know the of course, all the when how will we identify that ok, now there is a need for the adaptation, it could be on the basis of many different metrics. And some metrics of indicators could be like this. For example, the system is able to system or the service that you are interacting with is able to understand that ok you are experiencing a negative emotional state because of your voice or through any different modality.

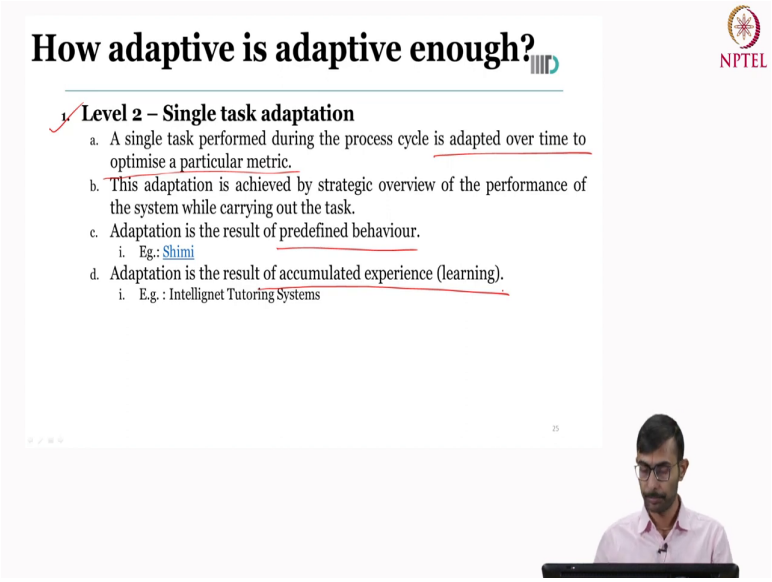
And now it knows that ok, since you are experiencing a negative emotional state you are feeling low, you are feeling sad, you are feeling bored and so on so forth. So, what it means, that ok there is a need for the adaptation here. Similarly, whenever there is a change in the emotional state maybe you can just look at this metric, the systems can look at this metric that whenever there is a change in the emotional state of a user maybe the system also needs to adapt to that.

And that is the metric that you are; of course, you are not doing the adaptation as we already talked about it, that no adaptation is being performed at this stage, but you are able to identify the need for the adaptation. Negative emotional state, we already talked about it. Changes in the emotional state - whenever you change your emotional let us say from one to another, maybe you are feeling happy and suddenly you started feeling sad.

So, then maybe the system will get an alert, the individual was feeling happy and now the individual is feeling sad. There is a change in the emotional state and maybe this is a right movement to do some adaptation, but in the level 1 of course, no adaptation is going to do.

And then you can think about like lots of other different metrics, I will let you explore and think that what could be the other metrics on the basis of which you may want to do some interventions or adaptations right and so, that is the level 1 system. So, level 0, no adaptation, level 1, it just recognizes the need for the adaptation, but in general there is no adaptation as of now yet we are not doing any adaptation ok.

(Refer Slide Time: 18:09)



Ok. So, now we go to the level 2 which is a bit more fascinating which is the single task adaptation. So, single task adaptation means what? What it does? A single task that is being performed during the entire cycle is adapted over time to optimize the particular metric. So, we already have some metric using which we are saying that, ok, we have, we first we are tracking the emotional state of the user.

We also have a metric or indicator which is telling us that we are recognizing the need for the adaptation. And then we are also in the level 2, now on the top of it, we are also doing the adaptation in a single task. And this adaptation can be you know, we may want to look into that, ok, what is the metric that we really want to improve in terms of the performance metric. And how can be it improved by doing a particular type of adaptation just to make example very clear.

For example, maybe you know, you are looking at let us say, let us say you your aim that ok you are building a conversational AI agent and then your aim is to make the user happy. So, you know, at the end of this interaction, the user should be feeling happy, user should remain happy, user should feel happy about the interaction of once the interaction is over.

So, whenever the user, you are monitoring the emotional state of the user and whenever you see that ok, whenever you experience that ok, user is feeling a bit low then what you want to do? You may want to do certain type of adaptation which will make the user happy because you your aim was to make the user happy. And; so, you will do a particular type of adaptation that is going to make the user happy.

Similarly, for example, if you are talk about the effective learning systems, so, where your aim is to improve the learning, then you may, maybe what you want to do, you are tracking the emotional state of the user and your continuous aim is to keep the user engaged or keep the boredom of the user as low as possible and so on so forth.

So, then accordingly, the type of the adaptation that you would like to do will be to address that particular goal. Maybe you want to make the user happy, maybe you want to make the user engaged, maybe you want to make the user's boredom low and so on so forth. So, I hope that this is what this is what it is clear that you are able to adapt with respect to the performance metric that you have kept as a goal for yourself.

And many times, this adaptation, you know it itself, what kind with what kind of adaptations could be there? It could be the result of some predefined behaviour. So, you are saying that ok, my user is feeling sad, I am going to make the user happy, but what will I do to make my user happy; now, this is a question. And this particular type of behaviour in this particular system, in the level 2 system is the result of a predefined behaviour.

So, for example, we already saw, we already looked at the example of the Shimi Robot. So, in the case of the Shimi robot, for example, what happened that the Shimi robot was able to you know play some music as per the emotional state of the user. But this particular type of music

or the behaviour of the gesture of the robot was predefined by the script that the developers have already built in.

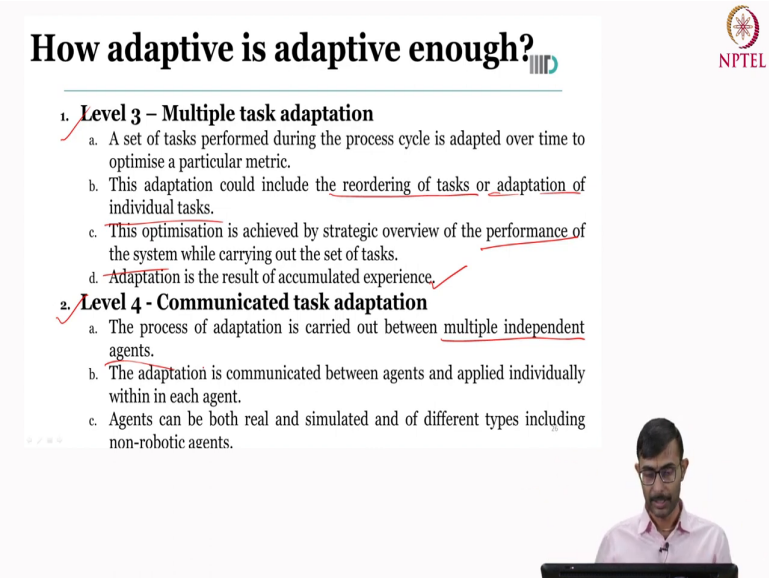
Similarly, adaptation can also be the result of the accumulated experience or the learning that happens over the period of time. So, for example, while it is not happening to the extent that we want it to be done, but in the intelligent tutoring systems, what can happen that you may, you know that the entire experience of the user say interaction with the system over a period of time. And then you can learn from that particular experience and you are able to adapt on the basis of that.

And, you know, for example, what type of adaptation will work for this particular guy. So, very simple example is, you know, when for example, as a teacher, when the teacher is taking a class in the of the for certain students, then teacher knows by the time. You know, in the beginning maybe the teacher may not have a very good idea that you know, what, how should I teach a particular topic to a particular student in order to make him or her understand.

But over the period of time you know, the user know the teacher knows, a good teacher at least knows that ok, I need to address this particular problem of a this particular student in such a way so that, you know, like it helps that particular student in a specific way. So, the type of the adaptation that the teacher does for a specific student is different that it does for the other student.

And this is the result of the learning that teacher has done over the period of time. And this is what we are envisioning here, that if the system can learn from the accumulated experience of interaction with the users, then it can improve the it can create the it can does the adaptation not on the basis of some predefined behaviour, but on the basis of the learning that it is doing and it is going to be very very adaptive and personalized.

(Refer Slide Time: 23:19)



So, that is the level 2 adaptation, which is a single task adaptation and of course, keeping a particular performance metric in the mind. Now, level 3 is a next level adaptation. In the level 3, what happens that rather than targeting a rather than doing the adaptation in a single task where for example, you know, maybe in the case of the Shimi robot, maybe the robot was only able to do the adaptation in its voice or maybe was only able to do the adaptation in its gesture.

Now, we are talking about that set of different tasks that are happening during the process cycle can be adapted over the time in response for to optimize a particular performance metric. And this adaptation can be of many different types; so, for example, adaptations could be that there are multiple tasks that the agents are doing, that your machines are doing, that your services are doing, you can simply do the reordering of the task.

Or you can simply do the adaptation of the individual task that are happening, but in parallel. So, for example, a good example of this thing would be that if you want to say that, ok, you want to do the reordering of the tasks based on the adaptation. Maybe, you know for example, you are trying to create an intelligent tutoring system here.

And the tutoring system, you know the movement; for example, the user logged in to learn a particular concept, maybe the system felt ok, the individual looks a bit, you know, like low on the energy today. So, if the individual looks a bit low on the energy, I am going to, you know teach, maybe topics that are easy to follow first, rather than, you know, start with the hard topics. And maybe I am going to teach in a way that is, you know at a very very basic level.

For example, multiple ways that this type of adaptations are being done and so, your reordering also and the multiple tasks are being, you know taken into account. So, you are looking at, you are not only adapting your teaching style, you are also changing the content that you want to teach for example, or in the other way, you are simply adapting multiple tasks, but you are not doing any reordering of that.

For example, imagine that you are interacting with a chatbot, a conversational bot and the conversational agent is not doing the reordering of anything. It is following the order that it is supposed to follow, but maybe you know, it is not only adapting its gestures in response to your emotional state. But also, it is adapting its voice as well in response to your emotional state.

So, there are two tasks at least you know, and maybe on the top of it, maybe, you know the task that it is supposed to perform actually, maybe you have a problem with the bank or something like that, it is also able to do it in a way that really pleases you. So, there are lots of different task that are happening, lots of different processes that are happening and the adaptation of all these processes or multiple processes are happening in at the same time in this level 3 type of adaptation.

And of course, nevertheless, no matter whether you are adapting a single task, multiple task, whether you are reordering the tasks, whatever the type of adaptation, this all the adaptation has to keep in mind the performance of the system. So, basically, you have a particular goal in mind, you want to make the user happy, you want to make the user feel fulfilled, you want to make the improve the learning of these or whatever.

You have already all this goal predefined and all this adaptations are happening as per the goal, broader goal that you have set for your system right. And in this case of course, you know, rather than having a predefined script, what you simply have? You simply have adaptation that is the result of the accumulated experience.

So, it is very very personalized and very very customized for each different user; so, the adaptations are going to be different for each different user depending upon their likes and dislikes; so, that is quite interesting. Now, in the level 4 adaptation, it is very much like the level 3, but there is a critical difference that the process of the adaptation is carried out between multiple independent agents.

So, what happens that till level 3, we are assuming that we have a system or a service or a machine where there is only one agent with which we are interacting with. And that agent is adapting to our task or is not adapting or is adapting multiple task that the agent itself is performing.

Here, we are saying that we are interacting in the multi-agent setting or we are interacting with the multiple services at the same time. And all these multiple services, multiple agents, they are talking to each other and they are saying that ok, you know like you adopt this, you adopt this and I will adopt this and let us collectively make the user feel good about the entire thing.

So, for example, you know maybe you are playing a game and the game there are multiple characters. So, when there are multiple characters, rather than one character adapting to you, all the different characters they are adapting to you, but they are doing in sync by

communicating with each other. So, this is really fascinating because; now, the your user experience is being looked at holistically and comprehensively and maybe it can provide a better experience overall.

And while doing so, of course, what the agents can be do? They can communicate the different adaptations and they can be applied individually within each agent. So, as I said agent A can say to agent B, service A can say to service B. So, to you know we have to do this, this and service A will do this type of adaptation, service B will do this type of adaptation.

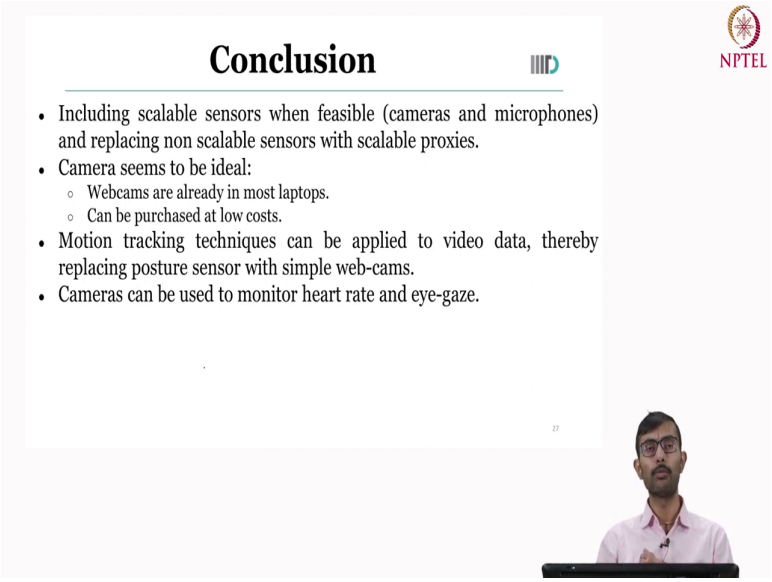
Similarly, you know agent A, agent B, machine A, machine B or whatever different types of things. So, for example, you know there are different, let us say in the gaming itself, I will there are two different characters that are around you, they are supposed to be helping you in I do not know finding some treasure.

And then; so, agent A is going to you know maybe say that ok, maybe the user is not able to find it, let us help the user find it. So, agent will say, ok maybe I am going to you know clear the path for the user while let us say you know you take care of the, I do not know, enemies that are there on the path; so, different things, right.

So, different, of course, depending upon the capabilities of the agents or the services and as I said again the creativity is the only limit here again for you, on the type of the adaptations that you can do for your machine ok. So, again, when we are talking about this multiple agents, this multiple agents can be both real and simulated and of different types as well.

So, you we are talking about the agents in the games, we can also have one agent in the embodied agent, in the animated agent, one robotic agent and then so on so forth. So, basically all the different types of agents or the services that we can envision, they can work together. Imagine you have a robot at your home, you have a Alexa also, you have a Siri also and they all are talking to each other in order to make you feel happy, I mean that would be really nice for example.

(Refer Slide Time: 30:14)



So, that is what is known as the communicated task adaptation. Now, in conclusion, when we looked at the; so, this is the, let us look at the conclusion now. So, when we looked at the open issues here, we already saw that we can use then scalable sensors, whenever there is a feasibility of it, such as in the case of the cameras and the microphones. And we can also replace the non-scalable sensors with the scalable proxies which can look at the latent behaviour, such as behavioural data.

The one that we, for example, the way we capture it from the keyboard, typing and so on so forth. Camera is a very very ideal choice, because it is already available in all the laptops systems and if not, then it can be also purchased at very very low cost. Motion tracking techniques can be applied to the video data, hence, you know, like it can also enable do this

kind of tracking. And so, this is, these are some of the things that you want to look into the, when you are looking at the conclusion.