Demystifying Artificial Intelligence & Machine Learning For Audio Engineers

An extract from **Sreejesh Nair's** *session titled '***Artificial Intelligence and Machine Learning For Audio Engineers'** *from PALM Conference & Seminar 2023*

Sreejesh Nair *is a celebrated award-winning audio engineer, and a renowned expert on Dolby Atmos. A winner of the 60th National Film Awards and a four-time IRAA Award recipient, Nair has 20 years of audio engineering experience and more than 200 films to his credit. In the recent* **PALM Conference & Seminar**, *held at BEC, Mumbai, Nair, AVID Pro Audio Application Specialist, took the stage to share his expertise and insight on the facets of Artificial Intelligence and Machine Learning for audio engineers. This article is an excerpt of Nair's session.*

Kickstarting a much-awaited session, **Sreejesh Nair**, an **Avid Pro Audio Application Specialist**, combined wisdom and wit on stage. He began, "I am here to talk about something that is a bit controversial – Artificial Intelligence (AI) and Machine Learning (ML) for audio engineers." He added, "We are here to understand what artificial intelligence is, what machine learning is, and if you know me, you know I am not a big fan of maths or technical jargon – so we will be learning this with some *paav bhaji* and other such examples." Talking about how AI and ML is often used interchangeably, Nair notes, "AI and ML are not interchangeable. AI is

anything that can exceed or match human intelligence. This is why, two masters or two mixes will never sound the same. If you use the same plugin, the same EQ, the same rides, two people will always have a different response to these things. This is why, these are certain things that can add a lot of randomness to what is beautiful when it comes to human engineering."

What Is AI?

According to Nair, "AI has the ability to discover, it has the ability to infer, and it

has the ability to reason. But, for all of this, it requires data input. It cannot dream these things out of nowhere. It cannot create these things out of thin air. If you feed it data, it can find patterns in the data or tell you different perspectives of something that you have never thought of before – or reason against you."

So, if that is AI, then what is ML? Nair explains, "As with audio, the beautiful thing about all of these terms is that it literally means what is written on it. ML is actually the capability to make predictions and decisions based on an input data. But if you look at this, it does not tell you if your prediction is going to be right, if that is going to be accurate, or if that is exactly the same thing. But, it will give you a different perspective to things. And that is where the beauty of it lies as a tool."

How Does ML Work?

To illustrate how machine learning operates, Nair employs a relatable example from daily life. "This is exactly how we learn audio. You learn about the technicalities, the integration, the differentiation, and more. We learn with terms that we can understand. This is the most simplified version of the same."

He moves further to illustrate his point, "So, let's say, I am staying at Marriot and I want to have *paav bhaji*. I want to decide if I want a room service. The first thing I would want to look at is – will I save time? Will I want to get up and do things or do I want to pick up the phone and just order *paav bhaji*? So we assign a value – let's call it X1, and to simplify things, let's keep the value of 1 as yes and the value of 0 as no."

Nair adds, "Will I save time? In my case, yes. That's a value of 1. Will I lose weight? Am I health conscious? I am going to order *paav bhaji* – I am not looking for a health-conscious diet. So, will I lose weight? No, I am not concerned. Will I save money? Now, if I am in Marriot and I have Marriot points, or I have coupons, I can obviously put it to my Marriot points. So, yes."

However, there is a caveat. Nair elaborates, "We don't make our decisions based on just these three things. We assign something called a weightage threshold. Let's say we put a number of 5 to that. Now, will I save time? Absolutely, I will save time. So, I will assign a weightage of 5 to that. Will I lose weight? Let's say I am partially health conscious, so I will give it a value of 3. Will I save money? I am not really concerned about money at this point. So, I assign it a value of 2. The calculation looks something like this: (X1*W1 + X2*W2 + X3*W3) – (Threshold) i.e. (1*5 + 0*3 + 1*2) – 5 = 2. In my case, the answer is 2. It's a positive value. Which means, yes, I should order *paav bhaji*. This is how ML works. There are many more complications to this. But, this is the gist of ML."

The Multifaceted Elements Of Machine Learning:

Nair notes that ML can be supervised. He explains, "ML has the ability to create an algorithm of its own. If I am a programmer, then I have to give every condition possible for it to give an output. Now, if any other condition exists, like [for a reverb] if I only have pre delay and a decay, and I need to add reflection, I need to program the reflection into the reverb. I have reflection, but I now need 7.1, so now I program it too. Since its Atmos now, so we need at least 9.1.4 or the client is not going to be happy. So, I program 9.1.4. Now, I may also need to release this in 22.2. So, you can see how this can scale. You have to programme every single thing. Like Batman has no jurisdiction, ML has no jurisdiction."

In supervised learning, Nair states that ML can iterate on the data being fed to it and create an output on its own. He explains, "You give ML a set of inputs and label it, and it will listen to what you feed and try and create a pattern from that. Later, you give it an input without telling it what it is. It will try and figure out an output and you can check if your output is correct. If it is not, then you go and tweak the algorithm. You give it more data. For this to work, ML needs an extensively huge amount of data. Especially for audio."

ML also functions on something known as 'unsupervised learning'. Nair enlightens his audience, "Unsupervised learning is where ML makes its own decisions. This is where Deep Learning comes into play. And that is actually where you will see most of the future developments in the field of audio. It's not about taking away jobs, but about seeing all of the possibilities available to us."

Last, but not the least, Nair remarks that for ML to work, users also use a model known as 'reinforcement'. "Reinforcement is like a reward model", Nair shares, "You give it an input. It gives you an output. Is the output right? If yes, then you reward it. If it's not, you punish it. It's amazing how close ML is to us. But, it's essentially trying to emulate how we think."

Understanding AI:

After fleshing out the basic concept of Machine Learning, Nair informs his audience that within ML, there exists a subsection, which is known as Deep Learning. When ML and DL are clubbed together, what we arrive at is essentially a part of AI.

Nair adds, "AI has other subtexts as well. NLP, which stands for 'Natural Language Processing', which is exactly what GPT is. There is an input. There is a model. There is an output."

How do we create this model? Nair answers, that while it seems like a complicated task, it is actually quite simple in practice. "If I am a programmer and if I have to give answers to everything, I can't figure out what combinations I am going to have. This is where data and learning system enters. The learning system is what actually creates this model, and it will iterate again and again. The more iterations you have, the better output you will get. The way these things work is they create an output, then match it to see if it is believable enough for a model. If not, it will go back, re-iterate, create. Go back, reiterate, create - until it gets one that is believable enough for the model to match."

He continues, "This is how the model is generated. So, if I was a company trying to make a plugin platform, that was going to do a De-Rustle on my track, I cannot create an algorithm to remove rustle. Because the way it works is that it has to understand what it is that I am trying to remove first and also fill in the blanks. If you remove rustle and you are left with a hole, then you might just as well call a sound editor."

Detailing why ML is the key to generating a better audio output by using AI, Nair expounds, "Imagine all of these things in an audio. You cannot communicate the terms of clarity, the terms of a feel, the terms of an openness to that particular system like you would to a human being. These are not technical terms you can actually put into a system. That is why, you need somebody like us. That is why it cannot be fixed by a plug in. That is why you need ML."

The Role Of AI and ML In Audio Engineering:

Nair moves on to describing a list of functions that AI and ML can perform to general a more sophisticated audio output, and how, through this, AI and ML are poised to transform the field of audio engineering.

Nair first talks about automatic mixing, "When I say automatic mixing, that doesn't mean that you take your tracks and tell the system, "I have three bass guitars, four guitars, three keyboards, one drum, two loops, and a voice – now, give me something that sounds like Honey Singh". No, that is not how it works. There are things that it can analyse and position, and can also create a well-balanced mix that it thinks is a good place to start. It is like, if you are a mix engineer, you are getting a well-produced track. Who doesn't like a well-produced start and mix?"

AI and ML can also assist with speech-to-text conversion and vice-versa. Nair states, "When we look at AI in general, there are some things that we take for granted. For example, the very fact that I am standing here, walking across, moving my hands, talking to you, creating a filter with my mouth, changing my tongue, moving my lips, looking at each one of you – it's a coordination of a lot of things. My central gravity has to be kept stable. The fluid between my ears has to control my balance. This is a very, very, very complex thing. Right now, we are still in the primordial soup of AI. So, if you think about AI, these types of functions being available, especially for blind people, physically challenged people, is a HUGE thing. Imagine the kind of musical instruments you can create for them to perform. Imagine the kind of DAW you can create for them to actually make it work."

Sound classification is another area where AI and ML can make the task smooth for audio engineers. Nair opens the conversation with a question, "How many of you have actually used ChatGPT to do sound editing? Explain your scene to ChatGPT and see what comes from it. It will give you suggestions that you have probably never thought of. Imagine you are trying to figure out a sound you want – a very emotional, sad blow of wind. The rain has to feel that agony and pain. AI can help classify the sound. Not just that, it can also suggest genres of songs for you. You will probably end up hearing songs you have never heard before."

AI can create and enhance an adaptive audio system. "This is excellent", affirms Nair, "since it is able to take elements in your mix and automatically place them in a space, depending on whether it is doing that automatically or using ML to find out the different components that make up an instrument and split it."

Nair adds, "Moreover, for sound synthesis, you can actually create sounds or instruments that never existed before. You can have different layers of different things, one following the other on an envelope, all of that creating a completely new experience."

Help with physically challenged musicians. You can actually have motion-controlled instruments that are triggering other things and have constantly changing music based on the performance happening on the stage. You will never see two shows in the same way.