

Teaching a robot how to speak

Contributor names:

Baris Serhan
Paul Piwek
Angelo Cangelosi

BARIS SERHAN: What is this?

ROBOT: Just a second. It may be a cup. As you can see, I am a humanoid robot created by SoftBank Robotics.

PAUL PIWEK: So the history of computers and using language actually goes back to Alan Turing. During the war, actually, people started developing the first digital computers. After the war, they were then also developed for more commercial purposes.

Turing worked, specifically, on this. And at the time, these computers were often referred to, not just by researchers but also in the popular media, as electronic brains. And so, there was a whole discussion about can these machines, can they actually think or not? Or will they ever be able to think? What does it actually mean for a computer to think?

OLD MAN: Hello.

ROBOT: Hello.

OLD MAN: Hello.

ROBOT: Hello.

OLD MAN: Hello.

ROBOT: Hello.

[LAUGH]

OLD MAN: Oh dear.

ROBOT: We could do this forever.

OLD MAN: I know we could.

ANGELO CANGELOSI: We are trying to teach machine to speak. In particular, we are working with humanoid robots. And the reason we are doing this is because, in a context where you have a humanoid robot living in your environment with you, helping you. For example, to prepare food you want to be able to communicate via language, which is our natural form of interaction.

ROBOT: Would you like to talk about my impressive hardware?

ANGELO CANGELOSI: For example, the latest one is the Pepper robot, developed by a company in France, SoftBank Aldevron Robotics.

ROBOT: I'm a sophisticated combination of hardware and software designed to interact with humans and bring them joy. I attach sensors on my head, each of my hands, and inertial sensors in my chest and legs to help me keep my balance. I can detect and avoid obstacles using the Sonar, lasers, and sensors built into my legs.

ANGELO CANGELOSI: Communication between people, it's not only speech-based-approach. But it's a nonverbal approach. So for example, emotional communication. We communicate expressing and understanding emotions through body movements, or through face or emotional expressions.

ROBOT: I'm still learning how to understand human emotions. But I can analyze your face and notice when you smile or frown.

WOMAN: OK.

ANGELO CANGELOSI: The off the shelf version of the Pepper, actually, is very similar to at home assistants like Alexa, Siri, all these systems in the sense that they are pre-programmed with a long list of words, a complex set of grammars. And they can talk about things without really referring to the meaning.

ROBOT: I play an instrument called Music Boxes. They are--

ANGELO CANGELOSI: This is a prerecorded, pre-programmed set of words and grammatical rules. And the Pepper, in some way, has no understanding of the meaning of words. It just can babble.

ROBOT: Don't be shy. Put your hands on top of mine.

PAUL PIWEK: In more recent years, in the last decade or so, in the wider area of artificial intelligence, machine learning has become very popular. In particular, the use of what are known as neural nets. What the machine learning algorithm or system does is it represents the English sentence as a number or a mathematical structure. Basically, these neural nets at some point became very successful for recognizing images. So if you give it an image, it will tell you, oh, this is an image with a house in it or something like that.

BARIS SERHAN: Where is the red hammer? Pick up the blue cup.

ANGELO CANGELOSI: The iCub Project is now a 10 to 12 years old project. The iCub that you see here will be 9 years old on the 31st of March '18. So it's a 9 years old boy. But I think it speaks still probably like a 2 years old child.

BARIS SERHAN: Learn cup.

ROBOT: Cup.

BARIS SERHAN: Learn ball.

ROBOT: I like to learn.

ANGELO CANGELOSI: Our approach is called developmental robotics. This copies the ideas of developmental psychology. So we understand and we study how children learn looking at psychology experiments.

We implemented the same protocols and the same strategies into our robots. And up to here, we are achieving progress in technology.

MOTHER: Where the cooker?

BABY: Cooker.

MOTHER: Cooker. Very clever. Where's the kettle?

BABY: Kettle.

MOTHER: Kettle.

ANGELO CANGELOSI: One of the challenges is to understand speech. So to decode a wave sound into actual words. But I think the biggest challenge is to attach a meaning to these sounds. So as a baby, I can hear words like dog and red. But the real challenge is, how do I connect the word dog to the actual perception of interaction with an animal? And red to the actual perception of the color for the specific object?

MOTHER: Where is the yellow one? Can you see yellow? OK. Let me see it.

BABY: [BABY TALK]

GRANDMOTHER: That's right.

BARIS SERHAN: Learn ball.

ROBOT: I like to learn. This is a ball.

ANGELO CANGELOSI: In this demonstration, we are replicating the behavior of 2 years old children between 18 and 24 months of age. This is when children are starting to name objects, individual words or one or two words in maximum.

FATHER: What is it? What's that?

BABY: [BABY TALK]

MOTHER: Good girl.

FATHER: This one. You know what it's called?

BABY: Cucumbah.

FATHER: Cucumber.

BARIS SERHAN: Find ball.

ROBOT: OK. Now I'm looking for a ball.

ANGELO CANGELOSI: So you see in the demonstration that the robot is paying attention to the object. The tutor is moving the object around so that the robot can keep its attention there. And then via joint attention with the human teacher, the robot pays attention to the same direction of vision, which is the object focus here. You can now say the name of the object.

BARIS SERHAN: What is this?

ROBOT: I think this is a ball.

ANGELO CANGELOSI: The simultaneous experience of seeing the object and hearing the label, the word for the object, allows the artificial intelligence algorithm behind the robot architecture called artificial neural networks to learn the association.

So it's a pure speech, sound, and visual architecture. Associative learning, like in behaviour learning the human brain. And this shows that the robot is the basis for further later acquisition of complex skills like grammar combining two or three words together.

BARIS SERHAN: Stop detection.

ROBOT: OK, I'll stop it.

ANGELO CANGELOSI: Our approach is really highly interdisciplinary, means it requires the collaboration of experts from different fields. In particular, we'll require expertise from roboticists who build our machines. From programmers and AI people, artificial intelligence, who can program the AI techniques. But of course, psychologists and linguists because they are those that tell us what's our current understanding of the way language and meaning and sound works. And of course, how this applies, for example, in child development for language acquisition.

MOTHER: Where's the toaster?

BABY: [BABY TALK].

MOTHER: Toaster, you clever girl.

PAUL PIWEK: What I've really learned from doing research on human-computer interaction is and especially computer interaction using language is that you very quickly realize that things are a lot more complicated than you initially might have expected.

MOTHER: But there's no water coming out of that tap, is there?

PAUL PIWEK: So usually when I say hello to somebody in the street, they might say, hi or hello or how are you back. So it is a fairly limited number of things. But you can see where it gets tricky.

So when I say, OK, what have you been up to today? Then it's unlikely that you're going to find the answer to that question in a huge collection of things that have been said previously. Because it's really what my plans and intentions and beliefs are right now that matter here. And those are not necessarily explicitly represented in any text or even in any prior dialogue that exists.

MOTHER: I know you'd like it to have water. But we can just pretend.

ANGELO CANGELOSI: I don't believe that it's possible to achieve human level capabilities. At least in the medium term or maybe during my lifetime, just because of the complexity of the task. Language is what makes us special amongst other skills, let's say, as a human species. And I think we are very far from achieving this in my lifetime at least.

BARIS SERHAN: What is this?

ROBOT: Sorry, I do not know this object.