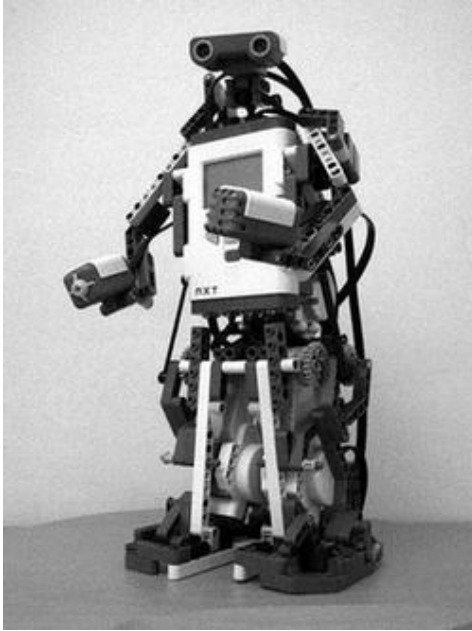


# Machines might talk with humans by putting themselves in our shoes



The humanoid robot ALPHA REX demonstrated the core principles of PRESENCE, as it synchronized its own behavior with that of a human. Image credit: Roger K. Moore.

**While robots can do some remarkable things, they don't yet possess the gift of gab. Since the 1970s, researchers have been trying to develop a speech-based human-machine interface, but improvements are gradual, and some fear that the performance of current systems may not reach an adequate level for real-world applications.**

Roger Moore, a computer scientist at the University of Sheffield in the UK, thinks that the current bottom-up architecture of speech-based human-machine interactions may be flawed. He is concerned because, although the quantity of training data for machines has increased exponentially, machines are still poor at understanding accented or conversational speech, and lack individuality and expression when speaking.

Moore has recently suggested an alternative model for speech-based human-machine interaction called PRESENCE (PREdictive SENsorimotor Control and Emulation). While the conventional reductionist architecture views spoken language as a chain of transformations from the mind of the speaker to the mind of the listener, PRESENCE takes a more integrative approach. As Moore explains, PRESENCE focuses on a recursive feedback control structure, where the machine empathizes with the human by imagining itself in the human's position, and then changes its speech patterns accordingly.

“The main difference between PRESENCE and current approaches to spoken language technology is that it offers the possibility of, one, unifying the processes of speech recognition and generation (thereby reducing the number of parameters that have to be estimated in setting up a system) and, two, linking low-level speech processing behaviors to high-level cognitive behaviors,” Moore told *PhysOrg.com*. “This should give a PRESENCE-based system a considerable advantage over more conventional systems that treat such processes as independent components, and then struggle to integrate them into a coherent overall system.”

Moore's model is inspired by recent results in neurobiology—such as the communicative behavior of all living systems, and the special cognitive abilities of humans—that aren't directly related to speech. Nevertheless, the results have provided a number of implications for human-machine speech, such as the

strong relationship between sensor and motor activity, and the power of negative feedback control and memory to predict and anticipate future events.

“A key idea behind the PRESENCE architecture is that behavior is driven by underlying beliefs, desires and intentions,” Moore explained. “As a consequence, behavior is interpreted with respect to one organism’s understanding of another organism’s beliefs, desires and intentions. That is, the ‘meaning’ of an observed action is derived from the estimated beliefs, desires and intentions that lie behind it—an individual is only able to make sense of another’s actions because they themselves can perform those actions. This is precisely a manifestation of the empathetic or mirror relationships that can exist between conspecifics (members of the same species).”

In a preliminary investigation, Moore constructed a humanoid robot called “ALPHA REX” that uses the PRESENCE hierarchical structure to demonstrate the relatively simple task of human-machine synchronization. As a human uttered the words “one, two” spoken at regular intervals, the robot generated taps. An overall control loop generated an error signal, which in turn modified the robot’s tapping rhythm until it matched the human’s words. Synchronization occurred by the eighth count, whereas a conventional model would require the robot to compute complex analytical solutions and suffer system delays. Further, because ALPHA REX could anticipate the human’s behavior, it tapped one extra time after the human ceased counting.

While it sounds simple, these kinds of coordination, reaction, and prediction abilities are necessary for the PRESENCE model, where behavior is quickly altered in response to the environment in order to achieve a desired state. As Moore explains, PRESENCE is less about speaking or listening, but about the human and machine interacting to meet each other’s needs. Again, this is in sharp contrast to conventional models that rely on the breakdown of components such as speech recognition, generation and dialogue.

Future machines that use PRESENCE could provide a variety of applications, such as robot companions or hands-free, eyes-free information retrieval. Moore predicts that PRESENCE machines could produce appropriate vocal intonations, volume levels, and a degree of emotion that is absent in current systems. He even suggests that the new machines could help unify currently divergent fields, such as speech science and technology; natural, life and computer sciences; and provide insight into fields in neurobiology that inspired PRESENCE itself.

Finally, Moore explains that it is very difficult to predict the speed and degree of progress in the future of human-machine speech.

“If we simply continue with the current research paradigm (which is mainly training on more data),” Moore said, “then for automatic speech recognition to compete with alternative technologies (e.g. keyboards etc.), it would need to be half as good as human speech recognition (i.e. it doesn’t need to be ‘super-human’)—and that is five times better than it is today. And the time until this would happen? In about 20 years if progress of the past 10 years can be sustained, or, if it can’t (which is most likely), then [possibly] never!”

Citation: Moore, Roger K. “PRESENCE: A Human-Inspired Architecture for Speech-Based Human-Machine Interaction.” *IEEE Transactions on Computers*, Vol. 56, No. 9, September 2007.

*Copyright 2007 PhysOrg.com.*

*All rights reserved. This material may not be published, broadcast, rewritten or redistributed in whole or part without the express written permission of PhysOrg.com.*

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study, research, no part*

*may be reproduced without the written permission. The content is provided for information purposes only.*