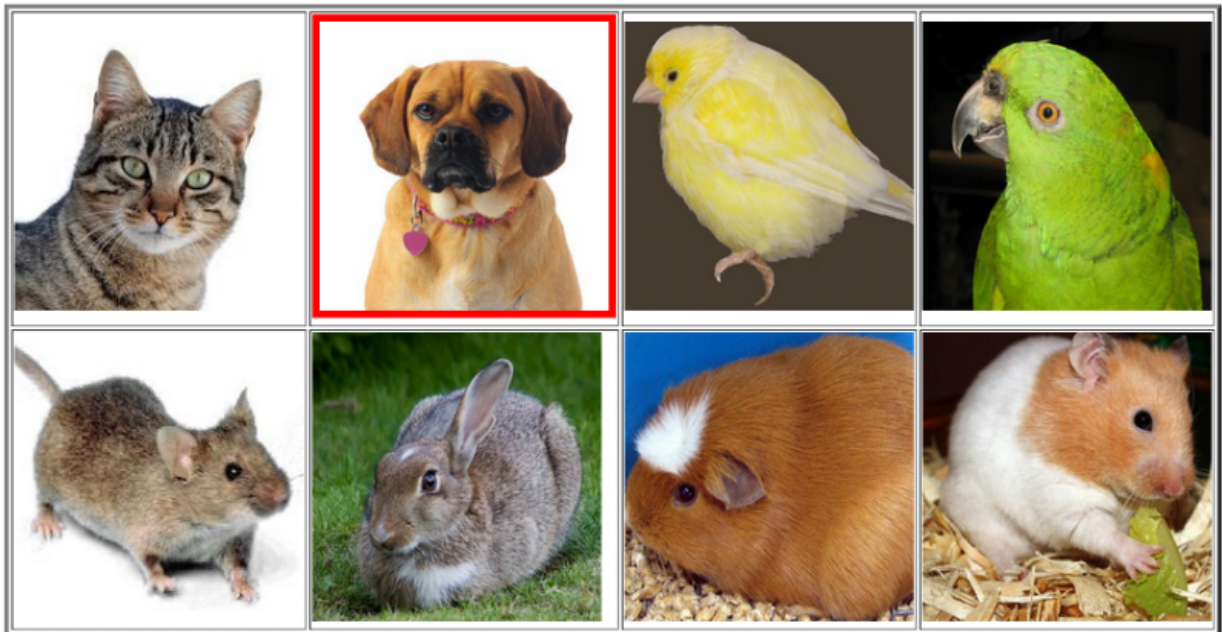


Researchers develop algorithms that let machines understand speech with human-like speed, accuracy

March 3 2016, by Aaron Dubrow



In the game, users describe pictures they see on their computer screen, and the agent tries to guess which picture they are talking about as fast and accurately as it can. Using its real-time, "incremental" (word-by-word) speech processing algorithms, the agent's speed of understanding and response is so fast that its game performance rivals that of human-human teams playing the same game. Credit: David DeVault, USC Institute for Creative Technologies

Today, talking to a computer or an app can be an infuriating experience.

But is it possible that one day we will have conversations with computers that feel as fluid and natural as talking to another person?

David DeVault, a research assistant professor at the University of Southern California's Institute for Creative Technologies, believes so—and his research is developing high-speed language processing systems whose speed and efficiency can rival that of human speakers in specific settings.

It turns out that one of the key challenges in creating more human-like voice interfaces is the rapid speed with which human speakers understand and respond to each other in a live conversation.

"While we human speakers can often understand and respond to what someone is saying to us in a fraction of a second, a typical voice interface will require much longer—often a second or two—to try to understand what you have said and respond in an appropriate way," DeVault says.

Interacting with the relatively slow pace of current voice interfaces is one reason many people continue to find them inefficient and frustrating to use.

With support from the National Science Foundation, DeVault and his students are studying new techniques that can streamline human-machine conversations by enabling the system to perform all of the necessary computer processing steps in real-time while the user is talking. The resulting systems are often able to figure out what the speaker means and how it should respond well before that person finishes speaking.

In a recent research paper, DeVault and his students, Ramesh Manuvinakurike and Maiké Paetzel, described the creation and evaluation of a high-performance game-playing agent called Eve.

In the game, users describe the pictures they see on their computer screen and the agent tries to guess which picture they are talking about as fast and accurately as it can. By using "incremental" (word-by-word) speech processing algorithms, the agent's speed of understanding and response is so fast that its game performance rivals that of human-human teams playing the same game.

When compared to alternative versions of the agent that wait until a user's speech is finished to try to understand and respond, users rate their interactions with the more incremental version of Eve as more efficient, more natural, and having better shared understanding.

"These findings underscore the importance of enabling systems to not only understand what users are saying, but to do so as quickly as a human would," DeVault said.

The research received a Best Paper Award at the 16th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL 2015). It points toward the creation of voice interfaces for other applications that users may find more natural and efficient to use.

As voice interfaces continue to become faster, better at understanding what we mean and less frustrating to talk to, they will be increasingly adopted in a wide range of important applications including information access, education, healthcare, entertainment and training.

"We are at the beginning of a sea change in what we can achieve through conversation with computers," DeVault says. "At this point we are seeing just a glimmer of what is about to become possible."

More information: "So, which one is it?" The effect of alternative incremental architectures in a high-performance game-playing agent

(Maike Paetzel, Ramesh Manuvinakurike, David DeVault), In Proceedings of SIGDIAL 2015, 2015. [ict.usc.edu/pubs/So ,%20which%20one%20is%20it%20-%20The%20effect%20of%20alter native%20incremental%20architectures%20in%20a%20high-performance%20game-playing%20agent.pdf](http://ict.usc.edu/pubs/So,%20which%20one%20is%20it%20-%20The%20effect%20of%20alter native%20incremental%20architectures%20in%20a%20high-performance%20game-playing%20agent.pdf)

Provided by National Science Foundation

Citation: Researchers develop algorithms that let machines understand speech with human-like speed, accuracy (2016, March 3) retrieved 20 July 2024 from <https://phys.org/news/2016-03-algorithms-machines-speech-human-like-accuracy.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.