

Speech technology enables kids to control video game

September 6 2016

Kids needed to say just two words - "jump" and "go" - to control a video game called Mole Madness, but Disney researchers had to design a speech technology system capable of sorting through the overlapping speech, social side talk and creative pronunciations of young children to make it work.

The keyword-spotting system developed by Disney Research works better for this game application than commercial [speech recognition](#) systems, which are derived largely from adult speech.

"The system's responsiveness and accuracy helped [children](#) enjoy the rapid-paced, multi-player game," said Jill Fain Lehman, senior research scientist at Disney Research.

"Speech recognition applications have become increasingly commonplace as the technology has matured, but understanding what kids say when they play remains difficult," said Jessica Hodgins, vice president at Disney Research. "This latest work by our researchers could make it possible to design any number of speech-based game or entertainment applications for children, including interactions with robots."

Lehman and her colleagues, Nikolas Wolfe and Andre Pereira, will present their findings at the Workshop on Child Computer Interaction Sept. 6-7 in San Francisco and at the International Conference on Intelligent Virtual Agents Sept. 20-23 in Los Angeles.

"Kids don't necessarily pronounce words quite like adults and when they are playing together, as they like to do, they often engage in side banter, or exclamations of excitement, or simply talk over each other," Lehman said. "That makes it tough for a speech-based system, even one that just has to detect the words 'go' and 'jump' as in Mole Madness."

In the cooperative two-player game, the players have to move an animated mole through its environment, gathering rewards as they avoid obstacles. To move the mole horizontally, one player says "go," while the other player moves the mole vertically by saying "jump."

During game play, the players often say their commands simultaneously. In other cases, they make statements to each other, such as "Don't say 'go' yet," that can be misinterpreted by the system. Sometimes, they're just making observations, such as "He's funny." They also sometimes speak very quickly, or speak slowly, or change pronunciations - "g-g-g-g-go" - in an effort to exert greater control over the game.

To train their keyword-spotting system, the researchers had 62 children ages 5-10 play the game, both in pairs and paired with a robot called Sammy, while a human "wizard" listened in another room and tried to map each "go" and "jump" into a button press on a game controller. The system uses separate models of go, jump, mixed, social speech and background noise, built from 150-millisecond segments of the training data. Though "go" and "jump" normally take about 300 milliseconds to say, the system used the 150-millisecond window to increase responsiveness and thus make the game more compelling.

Overall, the system was 85 percent accurate in recognizing the keywords. Almost 40 percent of the words overlapped in the child-child pairings and 26 percent overlapped in the child-robot pairings. Children spoke the keywords faster than normal 32 percent of the time when playing with each other and with the robot, but slower-than-normal 27 percent of

the time with the robot and 17 percent with another child.

A commercial continuous speech recognition system was about 35 percent less accurate overall than the keyword spotter, having particular trouble recognizing "go," recognizing overlapping speech and recognizing fast speech.

When 24 additional children ages 4-9 subsequently played the automated game, the system had some trouble understanding "jump," perhaps because the group included four-year-olds, who weren't represented in the original training data. Most of the very young children adapted to the system's difficulties by simply pronouncing the word more carefully or repeating it more often so that the game could proceed.

Though the keyword spotter wasn't perfect, it could respond faster than the human wizard, which made the game more compelling. When three mothers of [young children](#) reviewed the video and rated the engagement in the game of each player, they judged that the children were less than halfway between "could take it or leave it" and "very much into the game" when the human wizard was relaying their commands, but judged to be solidly enjoying the [game](#) when the automated system was in control.

"This technology can be reproduced with other vocabulary, allowing designers and developers to build novel children's applications that use limited speech as an input method," Lehman said.

Provided by Disney Research

Citation: Speech technology enables kids to control video game (2016, September 6) retrieved 24 April 2024 from <https://phys.org/news/2016-09-speech-technology-enables-kids-video.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.