

AI outperforms humans in speech recognition

October 20 2020, by Monika Landgraf



Thanks to its superior speech recognition system, KIT's Lecture Translator will provide better results with minimum latency in future. Credit: KIT

Following a conversation and transcribing it precisely is one of the biggest challenges in artificial intelligence (AI) research. For the first

time now, researchers of Karlsruhe Institute of Technology (KIT) have succeeded in developing a computer system that outperforms humans in recognizing such spontaneously spoken language with minimum latency. This is reported on arXiv.org.

"When people talk to each other, there are stops, stutterings, hesitations, such as 'er' or 'hmmm,' laughs and coughs," says Alex Waibel, Professor for Informatics at KIT. "Often, words are pronounced unclearly." This makes it difficult even for people to make accurate notes of a conversation. "And so far, this has been even more difficult for AI." KIT scientists and staff of KITES, a start-up company from KIT, have now programmed a [computer system](#) that executes this task better than humans and quicker than other systems.

Waibel already developed an automatic live translator that directly translates university lectures from German or English into the languages spoken by foreign students. This "Lecture Translator" has been used in the lecture halls of KIT since 2012. "Recognition of spontaneous [speech](#) is the most important component of this system," Waibel explains, "as errors and delays in recognition make the translation incomprehensible. On conversational speech, the human [error rate](#) amounts to about 5.5%. Our system now reaches 5.0%." Apart from precision, however, the speed of the system to produce output is just as important so students can follow the lecture live. The researchers have now succeeded in reducing this [latency](#) to one second. This is the smallest reported latency reached by a speech recognition system of this quality to date, says Waibel.

Error rate and latency are measured using the standardized and internationally recognized, scientific "switchboard-benchmark" test. This benchmark (defined by US NIST) is widely used by international AI researchers in their competition to build a machine that comes close to humans in recognizing spontaneous speech under comparable conditions,

or even outperforming them.

According to Waibel, fast, high accuracy speech [recognition](#) is an essential step for further downstream processing. It enables dialog, translation, and other AI modules to provide better voice based interaction with machines.

More information: Nguyen et al., Super-Human Performance in Online Low-latency Recognition of Conversational Speech. arXiv:2010.03449 [cs.CV]. arxiv.org/abs/2010.03449

Provided by Karlsruhe Institute of Technology

Citation: AI outperforms humans in speech recognition (2020, October 20) retrieved 30 January 2023 from <https://techxplore.com/news/2020-10-ai-outperforms-humans-speech-recognition.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.