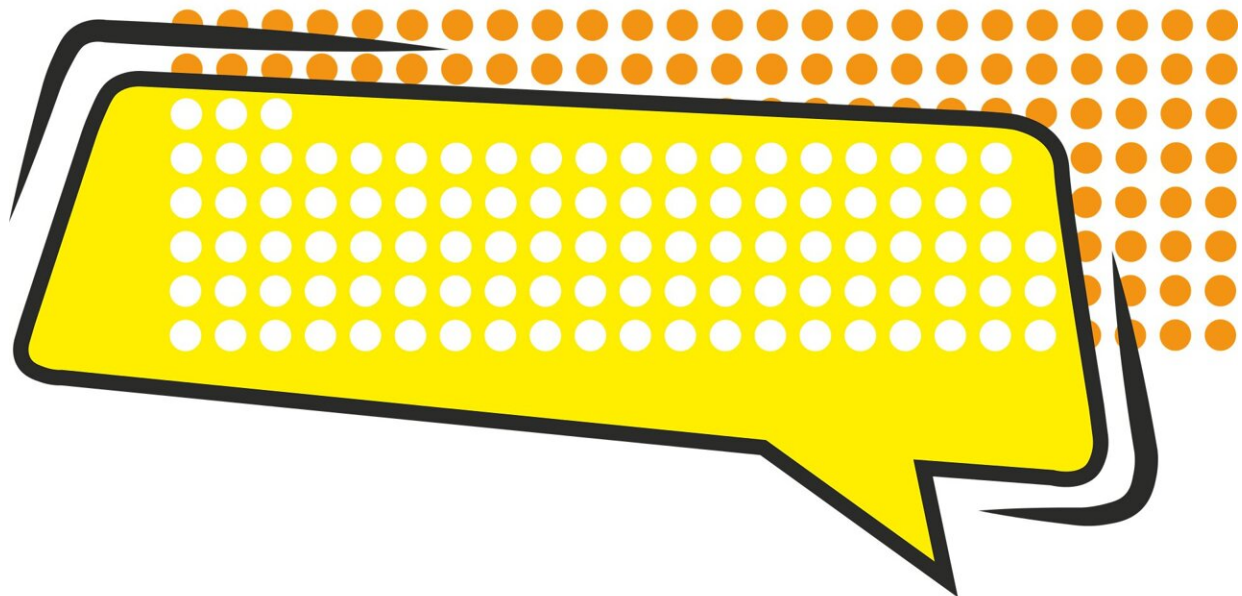


Research could bring automatic speech recognition to 2,000 languages

January 10 2023, by Aaron Aupperlee



Credit: Pixabay/CC0 Public Domain

Only a fraction of the 7,000 to 8,000 languages spoken around the world benefit from modern language technologies like voice-to-text transcription, automatic captioning, instantaneous translation and voice recognition. Carnegie Mellon University researchers want to expand the number of languages with automatic speech recognition tools available to them from around 200 to potentially 2,000.

"A lot of people in this world speak diverse languages, but language technology tools aren't being developed for all of them," said Xinjian Li, a Ph.D. student in the School of Computer Science's Language Technologies Institute (LTI). "Developing technology and a good language model for all people is one of the goals of this research."

Li is part of a research team aiming to simplify the data requirements languages need to create a [speech](#) recognition model. The team—which also includes LTI faculty members Shinji Watanabe, Florian Metze, David Mortensen and Alan Black—presented their most recent work, "ASR2K: Speech Recognition for Around 2,000 Languages Without Audio," at Interspeech 2022 in South Korea.

Most speech recognition models require two [data sets](#): text and audio. Text data exists for thousands of languages. Audio data does not. The team hopes to eliminate the need for audio data by focusing on linguistic elements common across many languages.

Historically, speech recognition technologies focus on a language's phoneme. These distinct sounds that distinguish one word from another—like the "d" that differentiates "dog" from "log" and "cog"—are unique to each language. But languages also have phones, which describe how a word sounds physically. Multiple phones might correspond to a single phoneme. So even though separate languages may have different phonemes, their underlying phones could be the same.

The LTI team is developing a speech recognition model that moves away from phonemes and instead relies on information about how phones are shared between languages, thereby reducing the effort to build separate models for each language. Specifically, it pairs the model with a [phylogenetic tree](#)—a diagram that maps the relationships between languages—to help with pronunciation rules. Through their model and the tree structure, the team can approximate the speech model for

thousands of languages without audio data.

"We are trying to remove this [audio data](#) requirement, which helps us move from 100 or 200 languages to 2,000," Li said. "This is the first research to target such a large number of languages, and we're the first team aiming to expand language tools to this scope."

Still in an early stage, the research has improved existing language approximation tools by a modest 5%, but the team hopes it will serve as inspiration not only for their future work but also for that of other researchers.

For Li, the work means more than making language technologies available to all. It's about cultural preservation.

"Each language is a very important factor in its culture. Each [language](#) has its own story, and if you don't try to preserve languages, those stories might be lost," Li said. "Developing this kind of [speech recognition](#) system and this tool is a step to try to preserve those languages."

More information: Xinjian Li et al, ASR2K: Speech Recognition for Around 2000 Languages without Audio, *Interspeech 2022* (2022). [DOI: 10.21437/Interspeech.2022-10712](https://doi.org/10.21437/Interspeech.2022-10712)

Provided by Carnegie Mellon University

Citation: Research could bring automatic speech recognition to 2,000 languages (2023, January 10) retrieved 21 April 2024 from <https://techxplore.com/news/2023-01-automatic-speech-recognition-languages.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.