

AI can use human perception to help tune out noisy audio

February 7 2024, by Tatyana Woodall



Credit: Pixabay/CC0 Public Domain

Researchers have developed a new deep learning model that promises to significantly improve audio quality in real-world scenarios by taking advantage of a previously underutilized tool: Human perception.

Researchers found that they could use the subjective ratings of sound quality made by people and combine that with a [speech](#) enhancement model to lead to better speech quality as measured by objective metrics.

The new model outperformed other standard approaches at minimizing the presence of noisy audio—unwanted sounds that may disrupt what the listener actually wants to hear. Most importantly, the predicted quality scores the model generates were found to be strongly correlated to the judgments humans would make.

Conventional measures to limit [background noise](#) have used AI algorithms to extract noise from the desired signal. But these objective methods don't always coincide with listeners' assessment of what makes speech easy to understand, said Donald Williamson, co-author of the study and an associate professor in computer science and engineering at The Ohio State University.

"What distinguishes this study from others is that we're trying to use perception to train the model to remove unwanted sounds," said Williamson. "If something about the signal in terms of its quality can be perceived by people, then our model can use that as additional information to learn and better remove noise."

The study, [published](#) in the journal *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, focused on improving monaural speech enhancement, or speech that comes from a single audio channel, such as one microphone.

This study trained the new model on two datasets from previous research that involved recordings of people talking. In some cases, there were background noises like TV or music that could obscure the conversations. Listeners rated the speech quality of each recording on a scale of 1 to 100.

This team's model derives its impressive performance from a joint-learning method that incorporates a specialized speech enhancement language module with a prediction model that can anticipate the mean opinion score that human listeners might give a noisy signal.

Results showed that their new approach outperformed other models in leading to better speech quality as measured by objective metrics such as perceptual quality, intelligibility and human ratings.

But using [human perception](#) of sound quality has its own issues, Williamson said.

"What makes noisy audio so difficult to evaluate is that it's very subjective. It depends on your hearing capabilities and on your hearing experiences," he said. Factors like having a hearing aid or a cochlear implant also impact how much the average person perceives from their sound environment, he said.

Since enhancing the quality of noisy speech is crucial for improving [hearing aids](#), speech recognition programs, speaker verification applications and hands-free communication systems, it's important that these differences in perception be small enough to prevent noisy audio from being less than user-friendly.

As the complex relationship between [artificial intelligence](#) and the real world continues to evolve, Williamson imagines that, similar to augmented reality devices for images, future technologies may augment audio in [real-time](#), adding or removing certain parts of the sound environment to improve a consumer's overall listening experience.

To help get to that point, the researchers plan to keep using human subjective evaluations to bolster their model to handle even more complex audio systems and ensure it keeps up with the ever-fluctuating

expectations of human users.

"In general, the entire machine learning AI process needs more human involvement," he said. "I'm hoping the field will recognize that importance and continue to support going down that path."

More information: Khandokar Md. Nayem et al, Attention-Based Speech Enhancement Using Human Quality Perception Modeling, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2023). [DOI: 10.1109/TASLP.2023.3328282](https://doi.org/10.1109/TASLP.2023.3328282)

Provided by The Ohio State University

Citation: AI can use human perception to help tune out noisy audio (2024, February 7) retrieved 9 May 2024 from <https://techxplore.com/news/2024-02-ai-human-perception-tune-noisy.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.