New AI noise-canceling headphone technology lets wearers pick which sounds they hear

November 9 2023, by Stefan Milne

A team led by researchers at the University of Washington has developed deep-learning algorithms that let users pick which sounds filter through their headphones in real time. Pictured is co-author Malek Itani demonstrating the system. Credit: University of Washington

Most anyone who's used noise-canceling headphones knows that hearing the right noise at the right time can be vital. Someone might want to erase car horns when working indoors, but not when walking along busy streets. Yet people can't choose what sounds their headphones cancel.

Now, a team led by researchers at the University of Washington has developed deep-learning algorithms that let users pick which sounds filter through their headphones in real time. The team is calling the system "semantic hearing." Headphones stream captured audio to a connected smartphone, which cancels all environmental sounds.

Either through voice commands or a smartphone app, headphone wearers can select which sounds they want to include from 20 classes, such as sirens, baby cries, speech, vacuum cleaners and bird chirps. Only the selected sounds will be played through the headphones.

The team presented its findings Nov. 1 at UIST '23 in San Francisco. In the future, the researchers plan to release a commercial version of the system.

"Understanding what a bird sounds like and extracting it from all other sounds in an environment requires real-time intelligence that today's noise canceling headphones haven't achieved," said senior author Shyam Gollakota, a UW professor in the Paul G. Allen School of Computer Science & Engineering.

"The challenge is that the sounds headphone wearers hear need to sync with their visual senses. You can't be hearing someone's voice two seconds after they talk to you. This means the neural algorithms must process sounds in under a hundredth of a second."

Because of this time crunch, the semantic hearing system must process sounds on a device such as a connected smartphone, instead of on more robust cloud servers. Additionally, because sounds from different directions arrive in people's ears at different times, the system must preserve these delays and other spatial cues so people can still meaningfully perceive sounds in their environment.

Tested in environments such as offices, streets and parks, the system was able to extract sirens, bird chirps, alarms and other target sounds, while removing all other real-world noise. When 22 participants rated the system's audio output for the target sound, they said that on average, the quality improved compared to the original recording.

In some cases, the system struggled to distinguish between sounds that share many properties, such as vocal music and human speech. The researchers note that training the models on more real-world data might improve these outcomes.

Additional co-authors on the paper were Bandhav Veluri and Malek Itani, both UW doctoral students in the Allen School; Justin Chan, who completed this research as a doctoral student in the Allen School and is now at Carnegie Mellon University; and Takuya Yoshioka, director of research at AssemblyAI.

More information: Bandhav Veluri et al, Semantic Hearing: Programming Acoustic Scenes with Binaural Hearables, *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (2023). DOI: 10.1145/3586183.3606779

Provided by University of Washington