

## **Improved Decoding of Attentional Selection in Multi-Talker Environments with Self-Supervised Learned Speech Representation**

Cong Han<sup>1</sup>

Vishal Choudhari, Yinghao Li and Nima Mesgarani

<sup>1</sup> Columbia University

Auditory attention decoding (AAD) is a technique used to identify and amplify the talker that a listener is focused on in a noisy environment. This is done by comparing the listener's brainwaves to a representation of all the sound sources to find the closest match. The representation is typically the waveform or spectrogram of the sounds. The effectiveness of these representations for AAD is uncertain. In this study, we examined the use of self-supervised learned (SSL) speech representation in improving the accuracy and speed of AAD. We recorded the brain activity of three subjects using invasive electrocorticography (ECoG) as they listened to two conversations and focused on one. We used WavLM to extract a latent representation of each talker and trained a spatiotemporal filter to map brain activity to intermediate representations of speech. During the evaluation, the reconstructed representation is compared to each speaker's representation to determine the target speaker. Our results indicate that speech representation from WavLM provides better decoding accuracy and speed than the speech envelope and spectrogram. Our findings demonstrate the advantages of self-supervised learned speech representation for auditory attention decoding and pave the way for developing brain-controlled hearable technologies.