

SPEECH SEPARATION

Wang, researchers train computers to hear through background noise

Machine-based speech separation, often referred to as “the cocktail party problem,” refers to the problem of using computers and other devices to separate target speech from interference caused by background noise. Monaural speech separation, accomplished from input made with a single microphone or other source, is central to many real-world applications, such as robust speech and speaker recognition, audio information retrieval and hearing aid design. However, despite decades of effort, monaural speech separation remains one of the most significant challenges in signal and speech processing.

Traditional speech separation algorithms have fallen into two categories: speech enhancement and beamforming. Speech enhancement is primarily a signal-processing based approach that estimates the target speech based upon broad statistics of speech and noise, while beamforming utilizes a sensor or microphone array.

More recently, however, researchers have begun to use high performance computing to formulate speech separation as a binary classification problem – dividing the elements of a given set into two groups. With such a formulation, considerable advances have been made in computational auditory scene analysis on monaural speech separation.

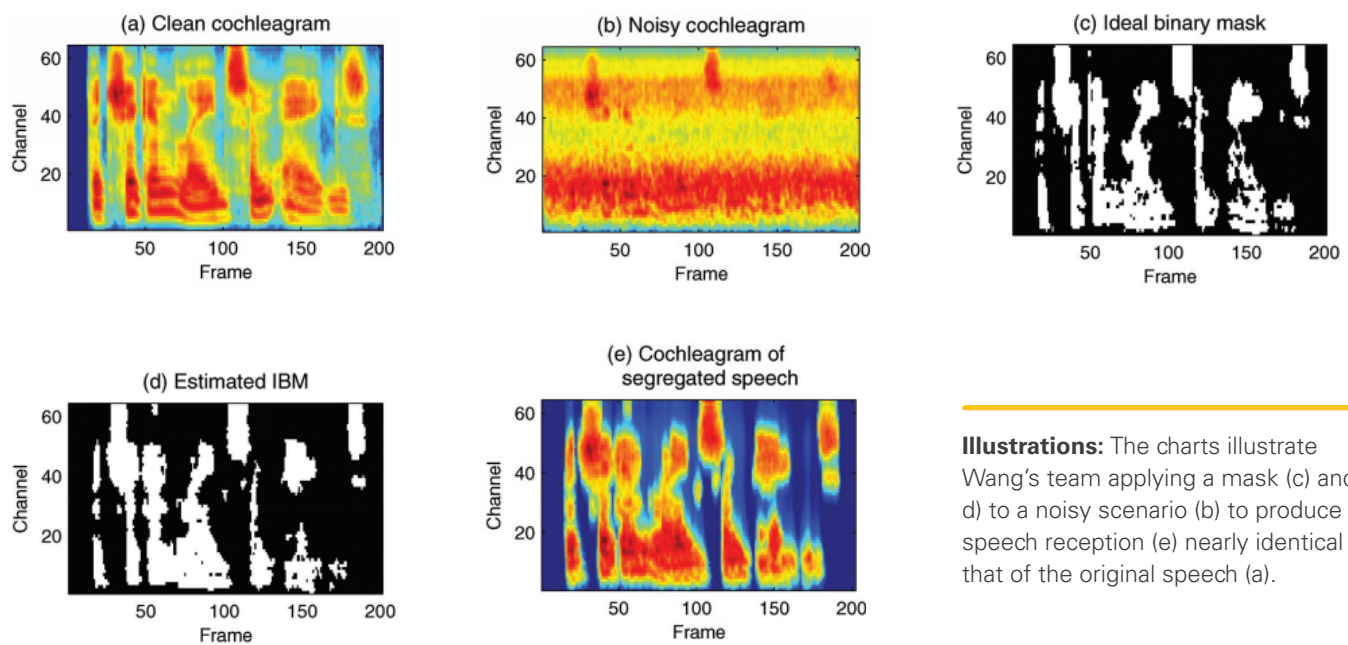
By utilizing resources at the Ohio Supercomputer Center, a research team headed by DeLiang Wang, Ph.D., professor of Computer Science and Engineering at The Ohio State University, already has made significant progress in identifying two important components of the supervised learning task for speech separation. First, the researchers identified a set

of complementary discriminative features that works well for speech separation, and, second, they identified deep neural networks as better choices than many other alternatives.

“We are now systematically studying the use of deep neural networks for feature learning and classification for the purpose of speech separation,” said Wang. “Specifically, we are training the classifiers on a large number of acoustic conditions in order to separate speech from a variety of noises.”

To improve generalization of the system to unseen acoustic environments, the researchers employ a large number of speakers and noises.

“We’ve managed to obtain significantly improved generalization performance through training on hundreds of hours of audio data,” explained Yuxuan Wang, an Ohio State doctoral student who is leading the research. “Solving the resulting large-scale machine learning problem is greatly facilitated by the Ohio Supercomputer Center.”



Illustrations: The charts illustrate Wang’s team applying a mask (c) and (d) to a noisy scenario (b) to produce speech reception (e) nearly identical to that of the original speech (a).

Project Lead: DeLiang Wang, Ph.D., The Ohio State University

Research Title: Large-scale computing
for classification-based speech separation

Funding Source: Air Force Office of Scientific Research, Kuzer

Website: <https://cse.osu.edu/people/wang.77>



Ohio Supercomputer Center

An OH·TECH Consortium Member