SpeechLab @ BC-CUNY

The SpeechLab at Brooklyn College builds computational systems for understanding and manipulating speech and audio signals using machine learning and signal processing. We are funded by the National Science Foundation, with both PI's receiving CAREER awards, along with the Alfred P Sloan Foundation and Google. Our graduates are now working at Google, IBM, and ETS and faculty at the University of San Francisco and Queensborough Community College.

Professors

Michael Mandel

Rivka Levitan

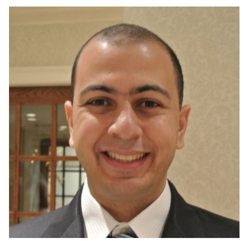
Get in touch: mim@sci.brooklyn.cuny.edu and levitan@sci.brooklyn.cuny.edu

Current PhD Students

Ali Syed

Rachel Rakov

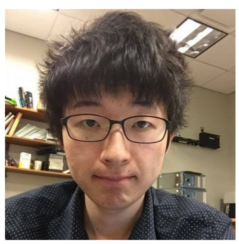
Felix Grezes



Hussein Ghaly

Andreas Weise

Soumi Maiti



Zhaoheng Ni

Trinh Viet Anh

Soheyl Salami

Why speech and audio?

Humans communicate effortlessly with one another through speech, even in the presence of adverse conditions like noise (imagine a cocktail party).

A smaller research community means it is easier to make an impact than in fields like machine vision and machine learning.

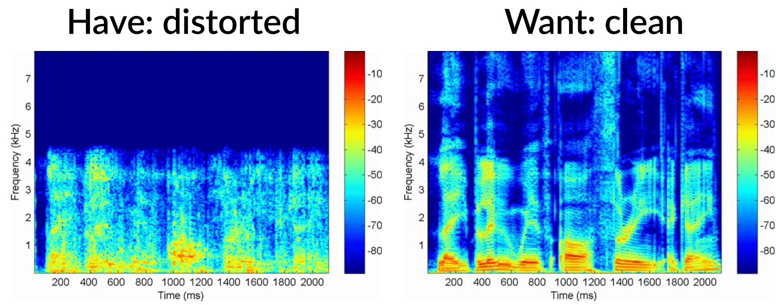
BCC Research estimates that the size of the global market for automatic speech recognition was $104 billion in 2017 and will increase to $185 billion in 2021.

The MarkeTrak survey estimates that 36 million Americans are hearing impaired, but only 10 million have a hearing aid and of those, many are unsatisfied with their ability to carry on a conversation in noisy environments.

Some current projects

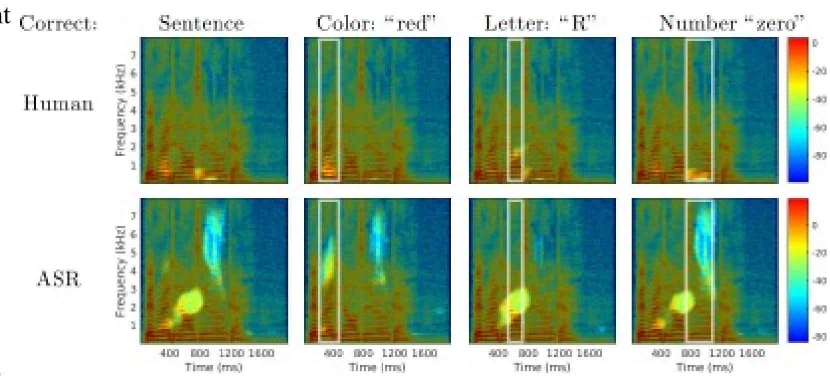
Speech enhancement by synthesis

Environmental noise is one of the largest problem for users of voice technologies, such as hearing aids, mobile phones, and automatic speech recognition. Current approaches to source separation and speech enhancement typically attempt to modify the noisy signal in order to make it more like the original, leading to distortions in target speech and residual noise. In contrast, this project uses the innovative approach of driving a speech synthesizer using information extracted from the noisy signal to create a brand new, high quality, noise-free version of the original sentence. This approach is able to improve the quality of speech in noisy conditions and in degraded conditions including reverberation, compression with mobile telelphony codecs, and simulated packet loss. When combined with neural vocoders like WaveNet and WaveGlow, it is able to generate predicted clean speech that is rated by listeners as having higher quality than enhancements produced by systems based on more traditional methods with oracle knowledge of the original clean speech.



Bubble noise to find important cues

Hearing is central to human interaction, but the hearing process is not easily observed. The objective of this project is to train models to identify portions of speech utterances that are important to their being correctly identified by human listeners, and to use predictions from these models to make automatic speech recognition (ASR) systems more noise robust by focusing on those regions. The ability to identify important regions of an utterance could significantly advance our understanding of healthy and impaired hearing. Our initial results have shown that that some regions of an utterance are more important or useful than others in identifying it by measuring the intelligibility of a given utterance in many different noisy mixtures. We are currently expanding this approach to utilize a deep learning framework we call the Bubble Cooperative Network, in which an agent augments the training data of an ASR system by adding noise where it is least disruptive, which requires understanding where important information is not.



Automating ecoacoustic analysis

Across North America, Arctic and boreal regions have been warming at a rate two to three times higher than the global average. At the same time, human development continues to encroach and intensify, primarily due to demand for natural resources, such as oil and gas. The vast and remote nature of Arctic-boreal regions typify their landscapes, environment, wildlife, and people, but their size and isolation also make it difficult to study how their ecosystems are changing. To overcome these challenges, autonomous recording networks can be used to characterize "soundscapes" - a collection of sounds that emanate from landscapes. Unlike traditional observing methods that are expensive, labor-intensive, and logistically challenging, sound-recording networks provide a cost-effective means to both monitor and understand the response of wildlife to environmental and anthropogenic changes across vast areas. One particular challenge with this sound-measurement approach is extracting useful ecological information from the large volumes of soundscape data that are collected. This project will develop the techniques necessary to overcome this challenge.