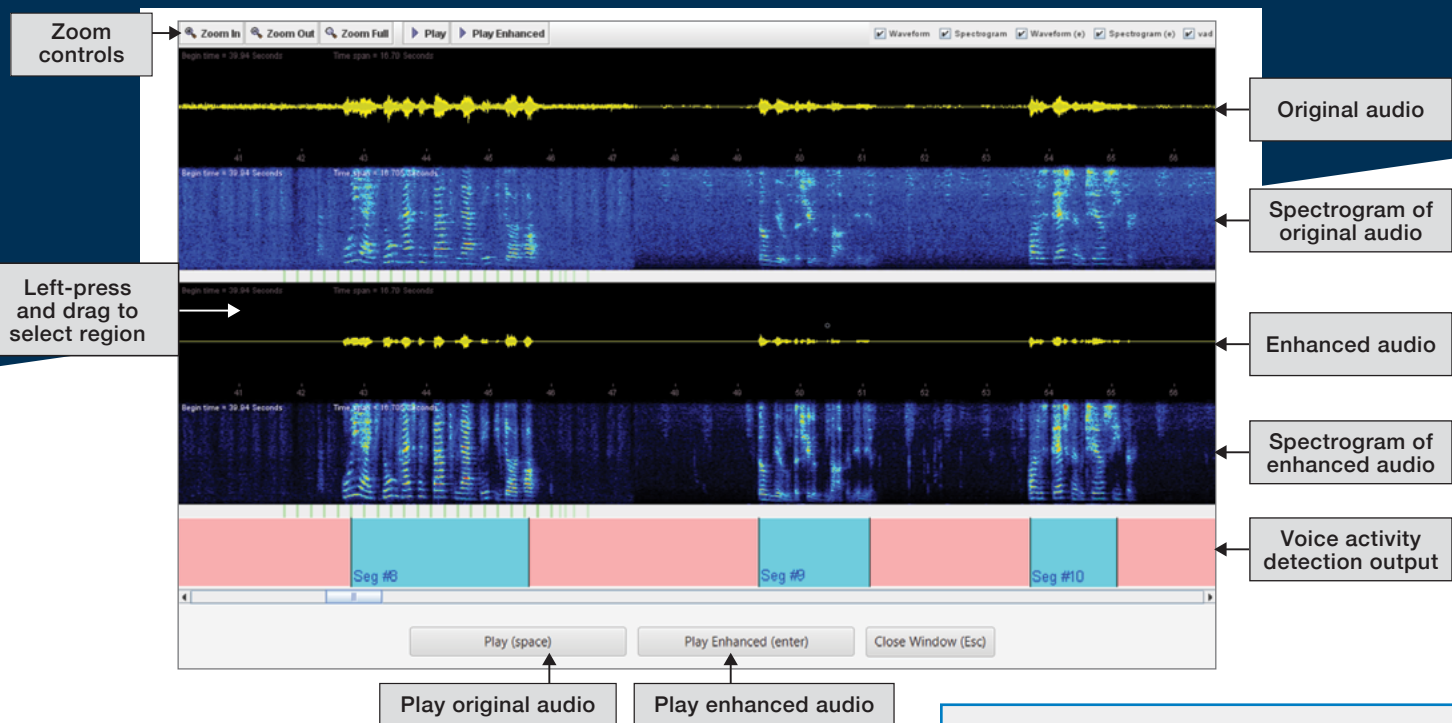


Speech Enhancement via Attention Masking Network (SEAMNET)

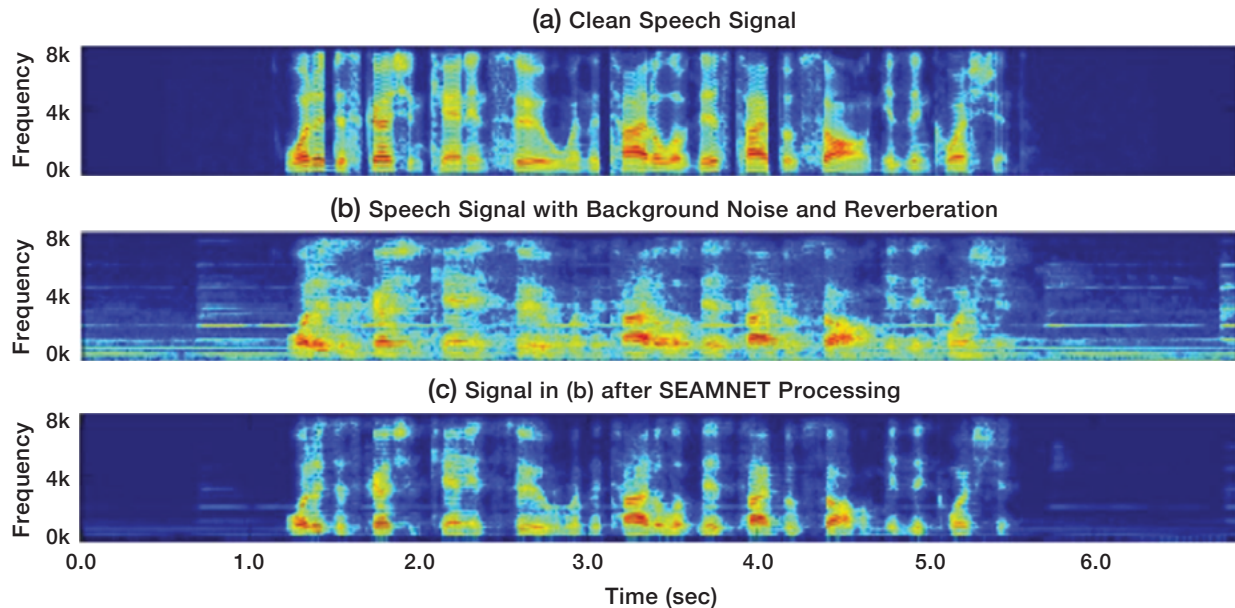


A graphical user interface of the SEAMNET enhancement system has been implemented on a laptop. The software allows a user to listen to and visualize specific portions of the noisy and enhanced audio signals.

SEAMNET is a neural network-based, end-to-end, single-channel speech enhancement system that could help improve the intelligibility of mobile communication devices, “smart” virtual assistants, and assistive speech technologies. It is designed to suppress both the noise and reverberation that can significantly degrade the performance of typical automated speech and speaker recognition systems.

KEY FEATURES

- Allows users to control the level of suppression applied during speech enhancement
- Provides joint noise and reverberation suppression; currently used speech/speaker recognition systems dampen background noise but not reverberation that is inherent in all speech signals
- Computational efficiency allows SEAMNET to operate even with systems limited in speed



An illustrative example of the SEAMNET system: Panel (a) shows the spectrogram of a clean speech signal. Panel (b) illustrates the spectrogram of the same speech in the presence of background noise and reverberation. Panel (c) shows the output of SEAMNET when applied to the noisy signal in (b).

Problem

Speech signals are typically captured in the presence of distortions such as reverberation and/or additive noise. For human listeners, these distortions result in increased listening effort and reduced intelligibility. For automated applications, such as speech and speaker recognition, distortions decrease system performance.

Solution

SEAMNET software reduces the effects of reverberation and noise on speech signals. It takes as input a noisy and/or reverberant speech waveform and outputs an enhanced version of the signal, processed to improve perceptual quality for human listening. SEAMNET leverages a deep

neural network to predict the underlying clean speech signal from a noisy observed version. It suppresses interfering sources in the input signal via a multiplicative mask in a learned embedding space. The enhancement and autoencoder capabilities within the overall network allow the user to dynamically control the trade-off between noise suppression and speech quality.

System testing has shown SEAMNET to outperform state-of-the-art competitors, both in terms of objective speech quality metrics and subjective listening tests. During the listening tests, human subjects found the dynamic control particularly beneficial in reducing the amount of perceived disturbing artifacts in the output signal.

U.S. PATENT
#11,227,586

More Information

B.J. Borgström and M.S. Brandstein, "Speech Enhancement via Attention Masking Network (SEAMNET): An End-to-End System for Joint Suppression of Noise and Reverberation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 515–526, 2021.

**INTERESTED IN WORKING WITH
MIT LINCOLN LABORATORY?**



Scan the QR code to learn more
www.ll.mit.edu/partner-us

Contact the Technology Transfer Office
tto@ll.mit.edu