**Abstract:**

Deep learning is used in multi-channel (MC) and single microphone acoustic echo cancellation (AEC) without altering the correlation to the loudspeaker signals and microphone signals while delivering a good performance.

In this article, we use different models to train the MC AEC Setups. Since acoustic echo is a common issue in communication systems, where sound reflects back into microphones, disrupting clarity. The proposed method utilizes a Single-Input Multi-Output (SIMO) framework, which enhances the system's capability in handling acoustic signals. To construct a high-quality and realistic dataset, we employ Pyroomacoustics, a powerful room acoustics simulation tool that accurately models reverberation and sound propagation in various environments. Furthermore, the model is training using the Noise-to-Signal Deviation (NSD) technique, which plays a crucial role in refining the system’s efficiency. This approach ensures an improved output by effectively mitigating distortions and enhancing the overall performance of the echo cancellation system.

\*\* This part should be from base model ig but i dont think we should put since our values arent better that thiers…..

Experiments are conducted on various unmatched scenarios and results show that the proposed method significantly outperforms previous methods. Moreover, a lightweight version of the proposed model with 0.29 million train able parameters also shows good performance, which is essential for resource-limited and real-time applications.

**Introduction**

* Multi channel Acoustic Echo Cancellation using deep learning is a fairly new area where very little research has been done.
* We have trained and evaluated several deep learning models and intend to show our findings in this paper.
* Convolutions are able to reduce noise but fail to take the temporal aspect.
* Recurrent NN can be used to consider the temporal parts of the wave file.
* Hence an combination of CNN and RNN gives a better output.
* Further it is noticed an encoder-decoder structure gives better output, than simply stacking multiple convolutions.
* The skip connections and the decoding from bottleneck is concatenated in every layer to give output, which learns to abstract and keep the fine details in a wave file.
* We have a synthetic dataset which is created by combining multiple speech from far end sources and near sources in pyroomacoustics.