

Time-Domain Speech Enhancement Using State-Space Models

Thesis FA/MA
Supervisor Tobias Raichle
Examiner Prof. Dr.-Ing. Bin Yang

Motivation

Speech enhancement studies improving the quality of spoken language and finds applications as a front-end in automatic speech recognition, telecommunication or hearing aids. Generally, speech enhancement covers multiple types of corruptions (noise, reverberations, echoes, compression artifacts etc.) but most previous works focus on denoising.

Selective state-space models (SSMs), like the recently proposed Mamba [1], are an efficient alternative to the ubiquitous Transformer. They have been shown to scale well and benefit from long sequence length, which makes them well suited to speech enhancement. In part due to their novelty, previous works have not fully explored the use of SSMs for speech enhancement [2, 3]. Particularly the use in time-domain speech enhancement has not been thoroughly covered.

Objectives

- Implement a framework for using SSMs for time-domain speech enhancement
- Use different datasets for training the model
- Evaluate the model on a range of benchmarks and compare with existing models
- Identify problems in the approach and find solutions

Prerequisites

- Took the Deep Learning exam with good results
- Good programming skills in Python
- Experience in ML-frameworks (Preferably PyTorch)
- *Optional*: Experience in sequence modelling
- *Optional*: Participated in the ISS Deep Learning Lab

If this topic has sparked your interest, write me an email and we can discuss the proposal in more detail. Please include your current transcript and CV.

References

- [1] Albert Gu and Tri Dao. “Mamba: Linear-time sequence modeling with selective state spaces”. In: *arXiv preprint arXiv:2312.00752* (2023).
- [2] Rong Chao et al. “An Investigation of Incorporating Mamba for Speech Enhancement”. In: *arXiv preprint arXiv:2405.06573* (2024).
- [3] Xiangyu Zhang et al. “Mamba in Speech: Towards an Alternative to Self-Attention”. In: *arXiv preprint arXiv:2405.12609* (2024).