

# Speech Representation Learning Through The Lens of TTA for SE

08. February 2026

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

## Motivation

LaDen [1] computes pseudo-labels in the latent space of WavLM [2], a large self-supervised (SSL) encoder, to perform test-time adaptation (TTA) for speech enhancement (SE). As WavLM's pre-training incorporates a denoising task, its embeddings are very robust in the presence of noise. The goal of this thesis is to explore how other SSL paradigms and models can be trained or fine-tuned to produce more reliable embeddings in the presence of diverse corruptions. As training these models is very computationally expensive, one part of this thesis is an analysis of how they scale with size and data.

## Objectives

- Conduct a literature search for candidate SSL and fine-tuning paradigms
- Evaluate existing approaches
- Fine-tune published models
- *Optional*: Train models from scratch
- Conduct a scaling analysis

## Prerequisites

- Took the Deep Learning exam with very 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] Tobias Raichle, Niels Edinger, and Bin Yang. "Test-Time Adaptation for Speech Enhancement Via Domain Invariant Embedding Transformation". In: *IEEE Open Journal of Signal Processing* (2026), pp. 1–10. DOI: 10.1109/OJSP.2026.3656059.
- [2] Sanyuan Chen et al. "Wavlm: Large-scale self-supervised pre-training for full stack speech processing". In: *IEEE Journal of Selected Topics in Signal Processing* 16.6 (2022), pp. 1505–1518.