TU Clausthal

SWZ Clausthal – Göttingen
Simulationswissenschaftliches Zentrum

INRECNET – INSECT RECOGNITION MODEL

INRECNET – INSEKTENERKENNUNGSMODELL

SWZ-Mini-Workshop “Simulation meets AI”

15.08.2024

Chingiz Seyidbayli, Prof. Dr. Andreas Reinhardt
Institute for Computer Science, TU Clausthal

Contents

- Introduction
- Speaker Embedding Models
- Models
- Dataset
- Evaluation & Results

INTRODUCTION

Introduction

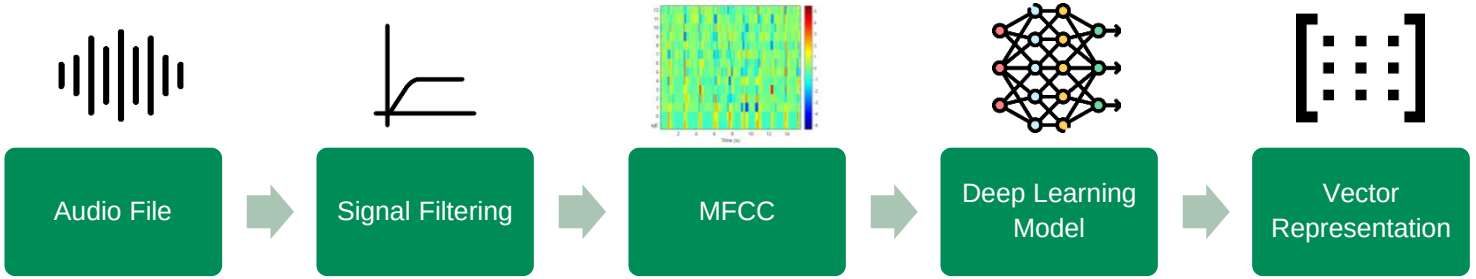
- The aim of the *BioIntAkt project* research is to develop an intelligent sensor system that uses artificial intelligence methods to autonomously and self-learningly classify and count insect species according to the sounds they make

- In the study, the creation of a *vector representation of the sound* used in systems such as human voice recognition, verification, etc. [1]-[3]

- The studies in the literature are generally based on the *classification of spectral features* obtained from the sounds of insect species [4]-[7]

SPEAKER EMBEDDING MODELS

Audio Vector Representations



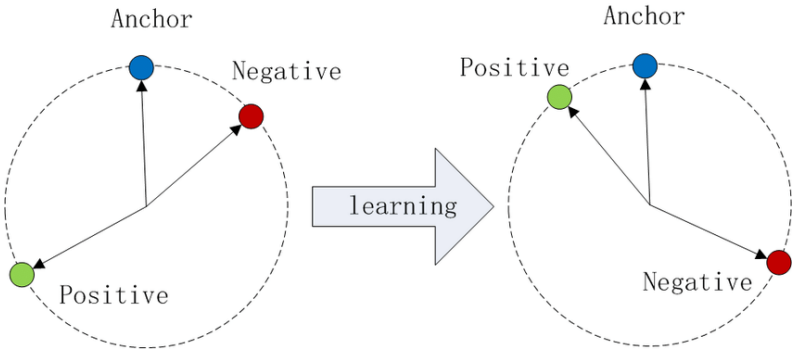
i-vector, x-vector, d-vector

Feature	x-vector	i-vector	d-vector
Purpose	Used in speaker recognition and verification	Used in speaker and language recognition	Used in speech and speaker recognition
Development	Based on deep learning (neural networks)	Based on traditional statistical models	Based on deep learning (neural networks)
Data Requirement	Requires large amounts of labeled data	Works with moderate amounts of labeled data	Requires large amounts of labeled data
Performance	High accuracy, especially in noisy environments	Good performance but less effective in noisy environments	High accuracy, adaptable to different tasks
Complexity	Complex due to deep learning techniques	Less complex, easier to implement	Complex due to neural networks
Usage Examples	Modern voice assistants, biometric systems	Early voice recognition systems, speaker ID	Advanced voice recognition, emotional analysis

Why d-vector representation?

- Here are some reasons why d-vector can outperform other methods, especially when used with modern deep learning approaches:
  - The Power of Deep Learning
  - Feature Learning
  - Overall Performance and Durability
  - Adaptation and Transfer Learning
  - Context Knowledge

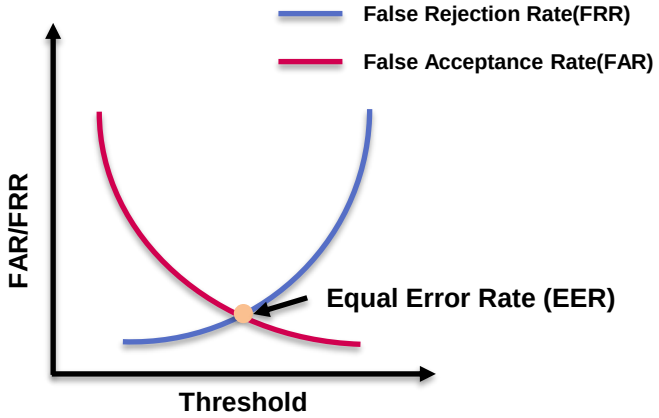
Triplet Loss and Equal Error Rate (EER)



$$L = \sum_{i=0}^{N} [s_i^{an} - s_i^{ap} + \alpha]_+$$

Triplet Loss Function

s_i^{ap} *is similarity score between anchor and positive sample*
s_i^{an} *is similarity score between anchor and positive sample*
α *is a margin parameter*



$$EER = min\{1 - TPR(\theta) - FPR(\theta) = 0 \mid \theta \in thresholds\}$$

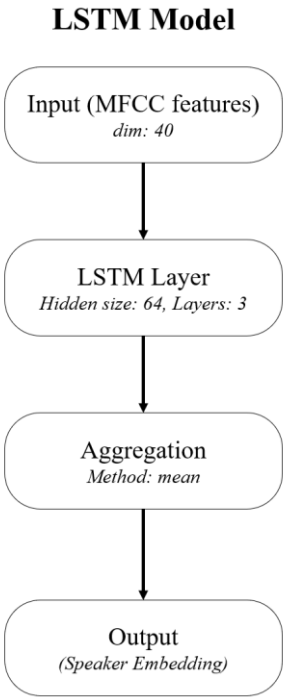
Equal Error Rate Function

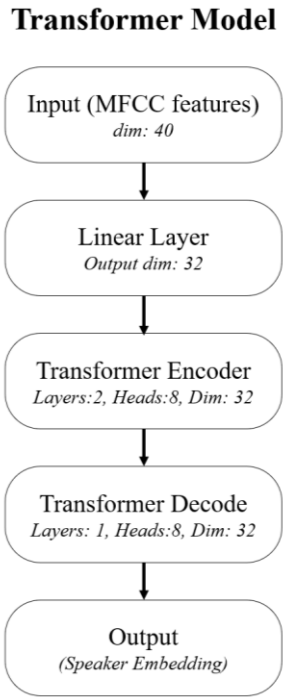
$TPR(\theta)$ *is the true positive rate at threshold* θ
$FPR(\theta)$ *is the false positive rate at threshold* θ

Image Source: https://en.wikipedia.org/wiki/Triplet_loss#/media/File:The-Triplet-loss-in-cosine-similarity.png

MODELS

Models


LSTM Model
Input (MFCC features)
dim: 40
LSTM Layer
Hidden size: 64, Layers: 3
Aggregation
Method: mean
Output
(Speaker Embedding)

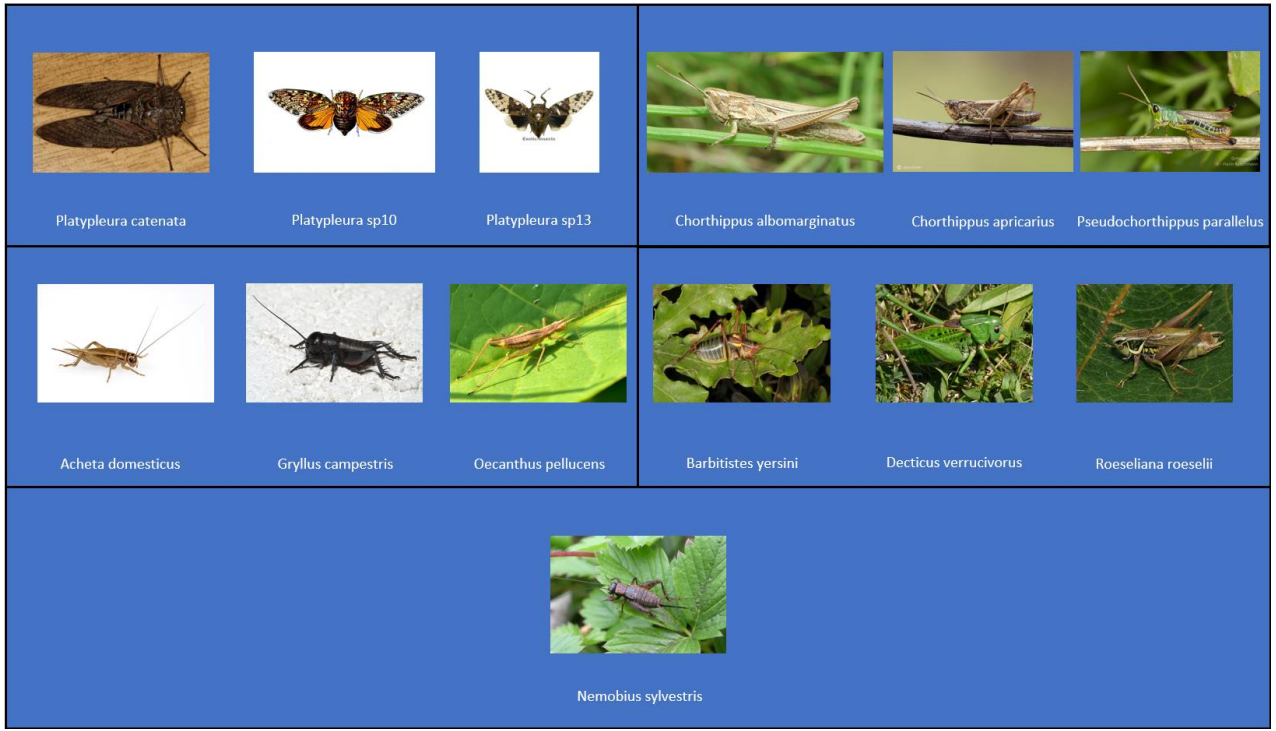

Transformer Model
Input (MFCC features)
dim: 40
Linear Layer
Output dim: 32
Transformer Encoder
Layers:2, Heads:8, Dim: 32
Transformer Decode
Layers: 1, Heads:8, Dim: 32
Output
(Speaker Embedding)

DATASET

Dataset

- InsectSet47 Dataset have been used for experiment. The dataset contains sound files of 5 different insect families;
  - Cicadidae,
  - Acrididae
  - Gryllidae,
  - Tettigoniidae,
  - Trigonidiidae

Dataset

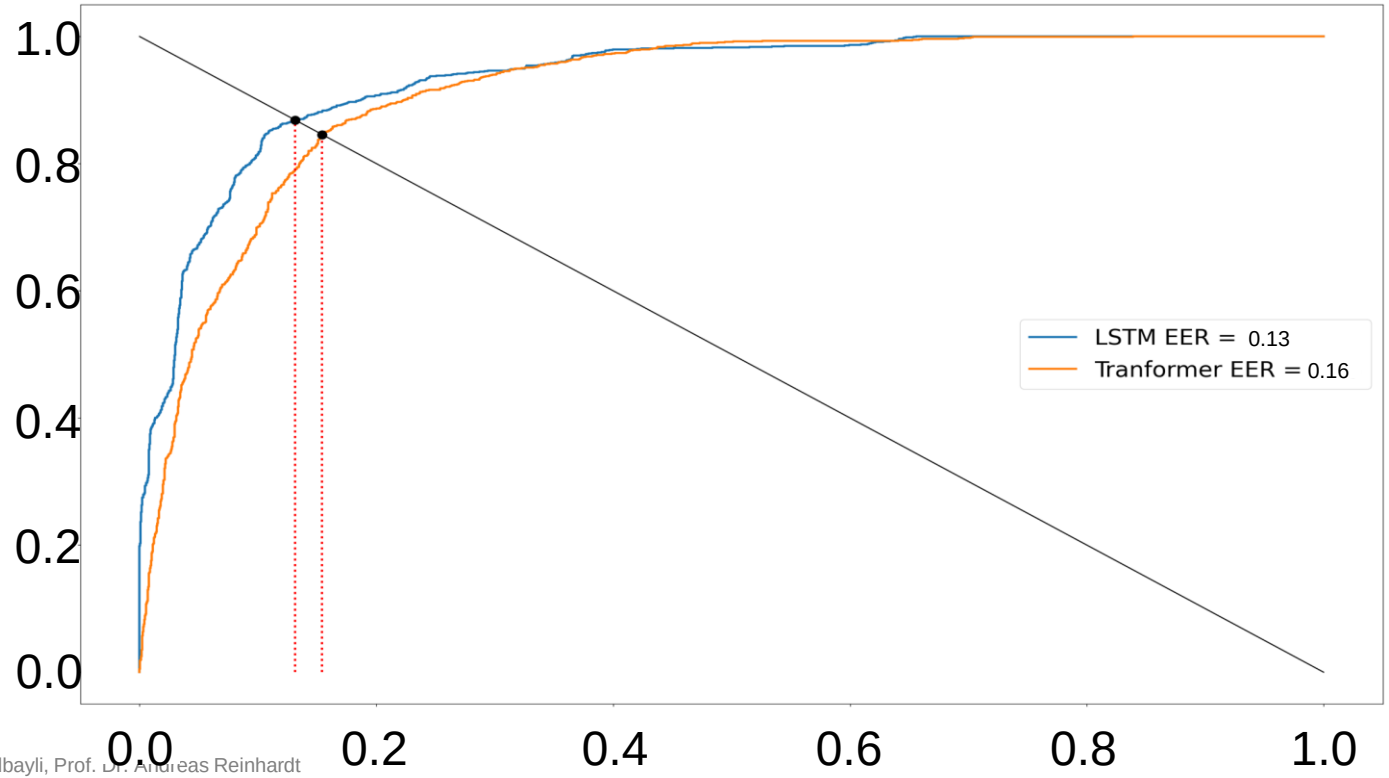

Platypleura catenata
Platypleura sp10
Platypleura sp13
Chorthippus albomarginatus
Chorthippus apricarius
Pseudochorthippus parallelus
Acheta domesticus
Gryllus campestris
Oecanthus pellucens
Barbitistes yersini
Decticus verrucivorus
Roeseliana roeselii
Nemobius sylvestris

EVALUATION & RESULTS

Evaluation

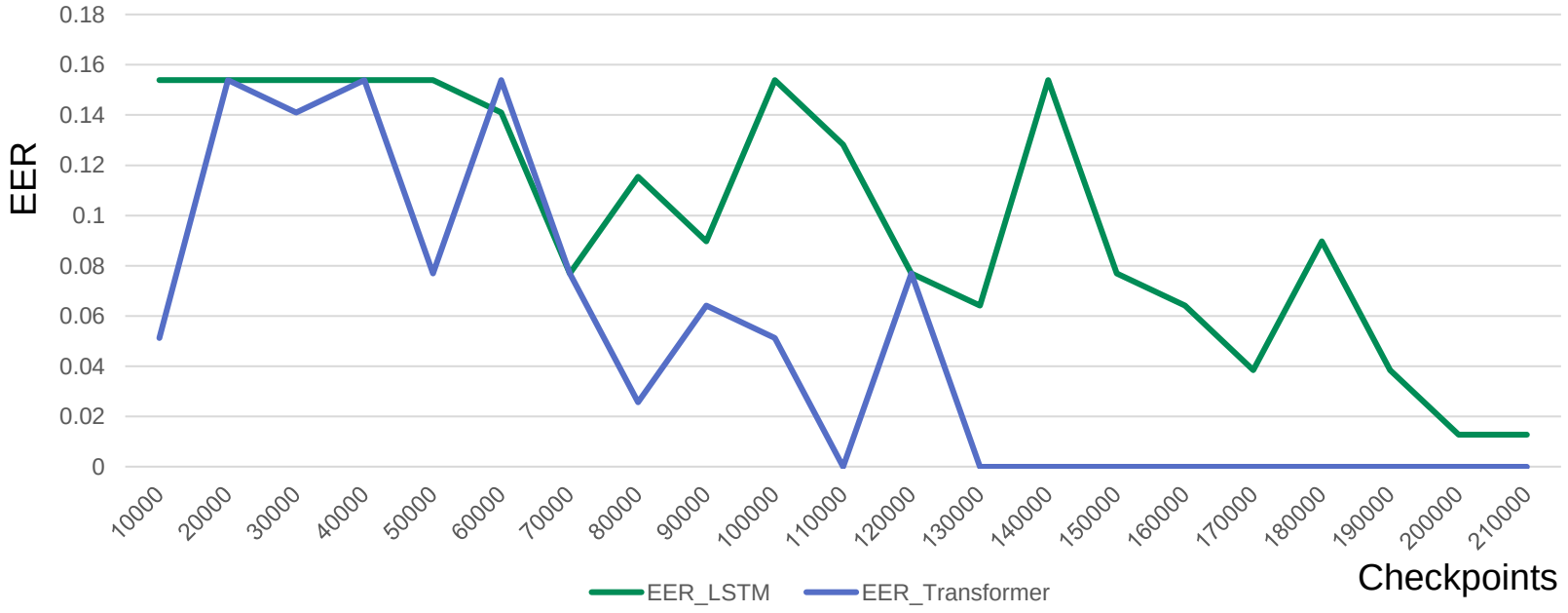
- Triplet loss was used in the study. During training, three files were randomly selected and the training process was carried out according to whether they were in the same class or not. Checkpoints obtained every 10000 triplets were recorded. With the checkpoints obtained, EER was calculated on 42 randomly selected files. In addition, the model obtained at the 210000th checkpoint was tested on 4371 audio pairs and an EER of 15% was obtained with LSTM and 16% with Transformer.

Result


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
LSTM EER = 0.13
Tranformer EER = 0.16

Result

- Models performance on the 42 random selected pairs



Conclusion and future works

- InsectSet47 dataset was used as train and test dataset
- LSTM model achieve to 15%, Transformer model achieve to 16% EER on 4371 audio pairs at the 210000th triplet
- LSTM model achieve to 13% EER at the 300000th triplet
- Speaker recognition models work for nature sound

- Re-train and test models on silence removed audio files
- Using real life noise on the background while training and testing
- Application of embedding creation model to the diarization models

References

[1] – Z. Bai, X. Zhang. "Speaker recognition based on deep learning: An overview," in Neural Networks, vol. 140, pp. 65–99, 2021.

[2] – N. Vaessen, D. Van Leeuwen, "Fine-tuning wav2vec2 for speaker recognition," in ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 7967–7971.

[3] – Novoselov, S., et al. "Robust speaker recognition with transformers using wav2vec 2.0," in arXiv preprint arXiv:2203.15095, 2022.

[4] – Truong, T., et al. "A deep learning-based approach for bee sound identification," in Ecological Informatics, vol. 78, pp. 102274, 2023.

[5] – Varma, A., et al, "Acoustic Classification of Insects using Signal Processing and Deep Learning Approaches," in 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN), 2021, pp. 1048–1052.

[6] – Zhang, M., et al, "A novel insect sound recognition algorithm based on MFCC and CNN," in 2021 6th International Conference on Communication, Image and Signal Processing (CCISP), 2021, pp. 289–294.

[7] – M. Faiß, D. Stowell. "Adaptive representations of sound for automatic insect recognition," in PLOS Computational Biology, vol. 19, no. 10, pp. e1011541, 2023.

Thank you for your attention!

Chingiz Seyidbayli
chingiz.seyidbayli@tu-clausthal.de

Image source: djd / Tourismusbetrieb der Stadt Oberharz am Brocken