Stuttering Identification using Deep Learning

Shakeel A. Sheikh [1] Md Sahidullah [1] Fabrice Hirsch [2] Slim Ouni [1]

[1]*Université de Lorraine, CNRS, Inria, LORIA, F-54000, Nancy, France* [2]Université Paul-Valéry Montpellier, CNRS, Praxiling, Montpellier, France

Abstract

Stuttering identification (SI) is an interdisciplinary research problem in which a variety of research studies (in terms of auditory feature extraction and classification approaches) have been carried out with the goal of creating automated tools for its detection and identification. The speech domain has been drastically revolutionized thanks to advances in deep learning, but SI has received less attention so far. In this work for, we explore time delay neural networks (TDNN) which is suitable for capturing temporal information of disfluent utterances. We also investigate how multi-task (MTL) and adversarial (ADV) learning framework can help to learn robust stuttering features. In addition, we explore different pre-trained speech embeddings in SI and we achieved state-of-the-art performance on a large corpus. The pre-trained speech embeddings based SI methods has shown the best performance with an overall accuracy of 68.35% on SEP-28k dataset.

Introduction

- Stuttering is a neuro-developmental speech impairment defned by blocks, prolongation, repetitions and interjections.
- Approximately 70 million ≈ 1% World's population suffer from stuttering.

Motivation

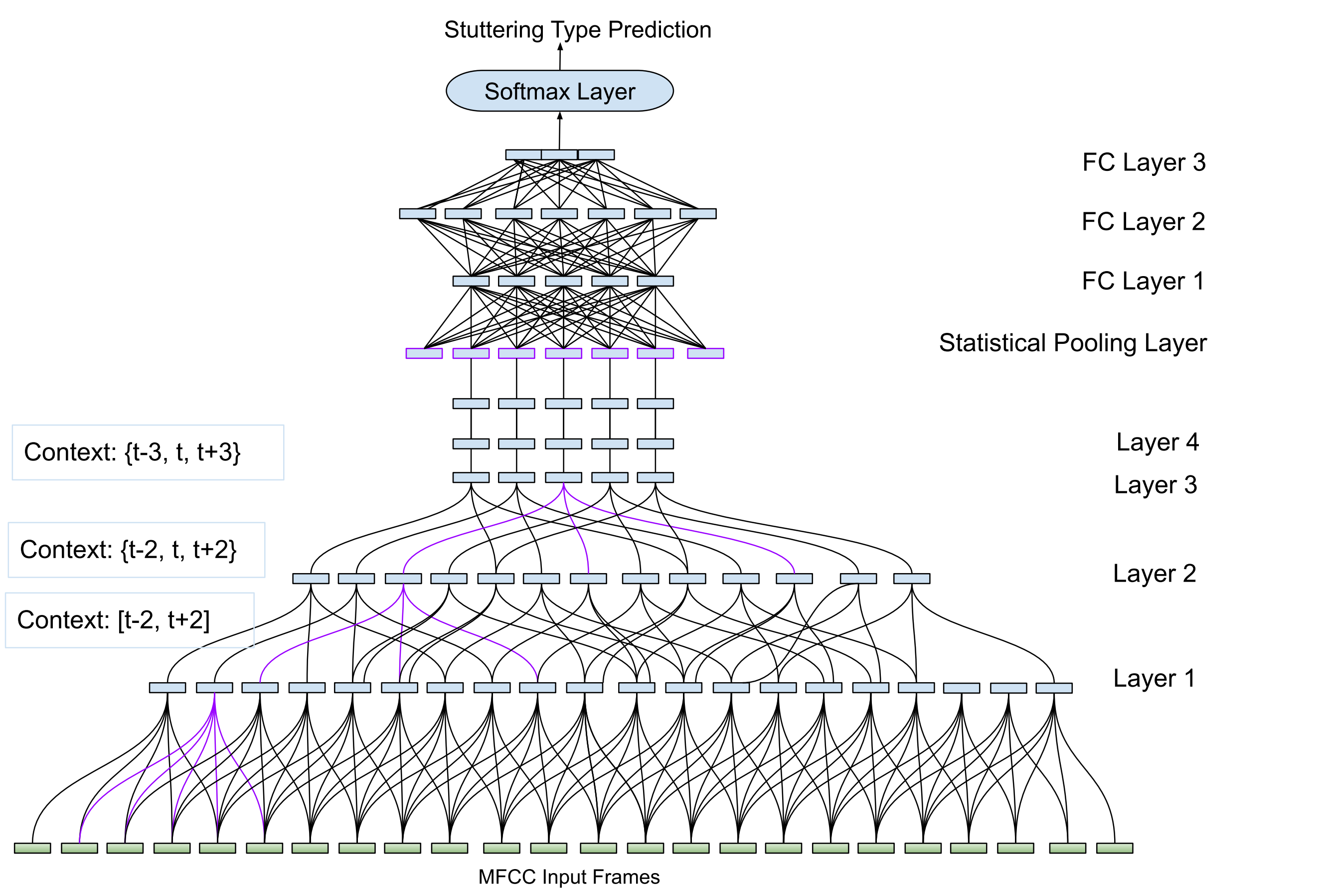
- Nearly impossible to access virtual assistants like Alexa, Apple Siri.
- Helpful for speech therapists.
- The success of statistical machine learning methods in limited in SI due to its complex nature among the different disfluent categories.
- Deep learning has been successfully applied in other speech domains like automatic speech recognition (ASR), emotion detection (ED), etc, but very less investiaged in SI.

Main Contributions

1. *StutterNet*: an end-to-end architecture for SI.
2. *Advancing StutterNet via multi-tasking (MTL) and adversarial (ADV) learning for robust SI*
3. *Stuttering identification with speech embeddings.*

Architecture for SI

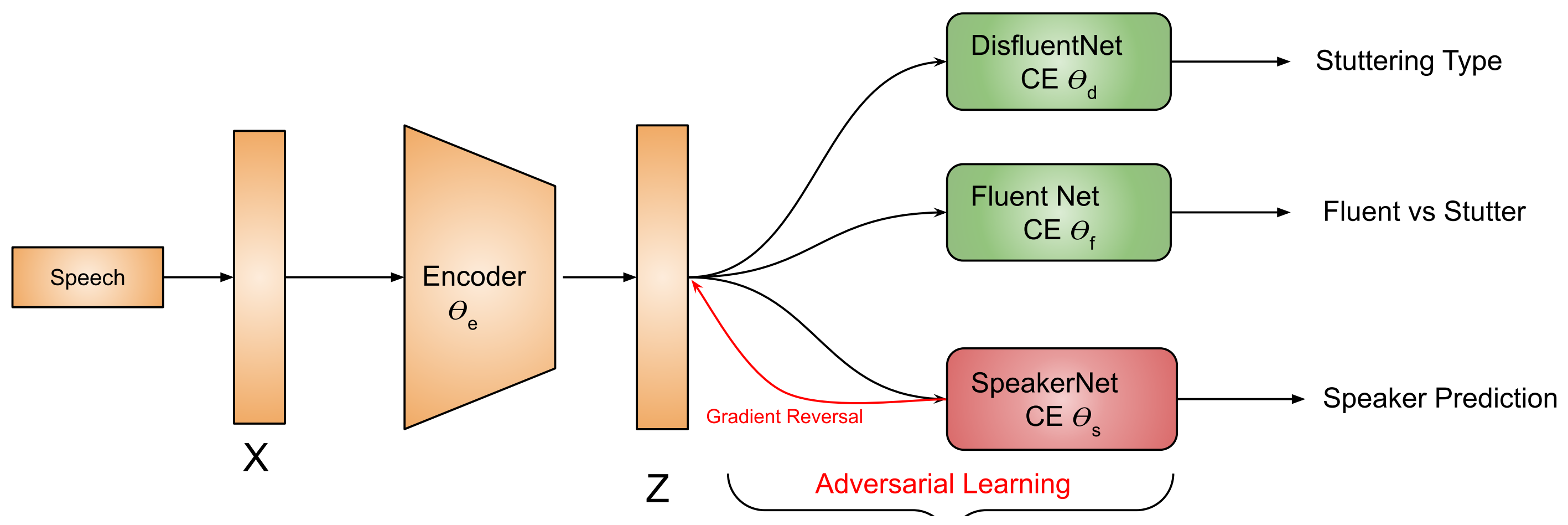
StutterNet:

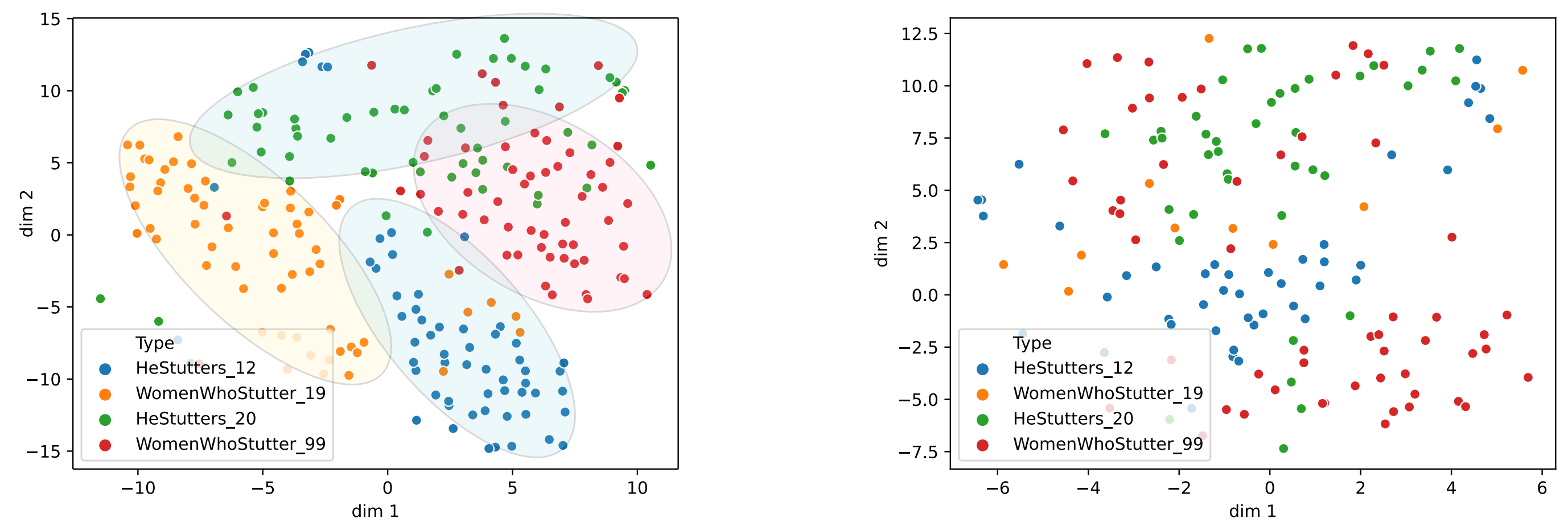


Multi-task and Adversarial Learning:

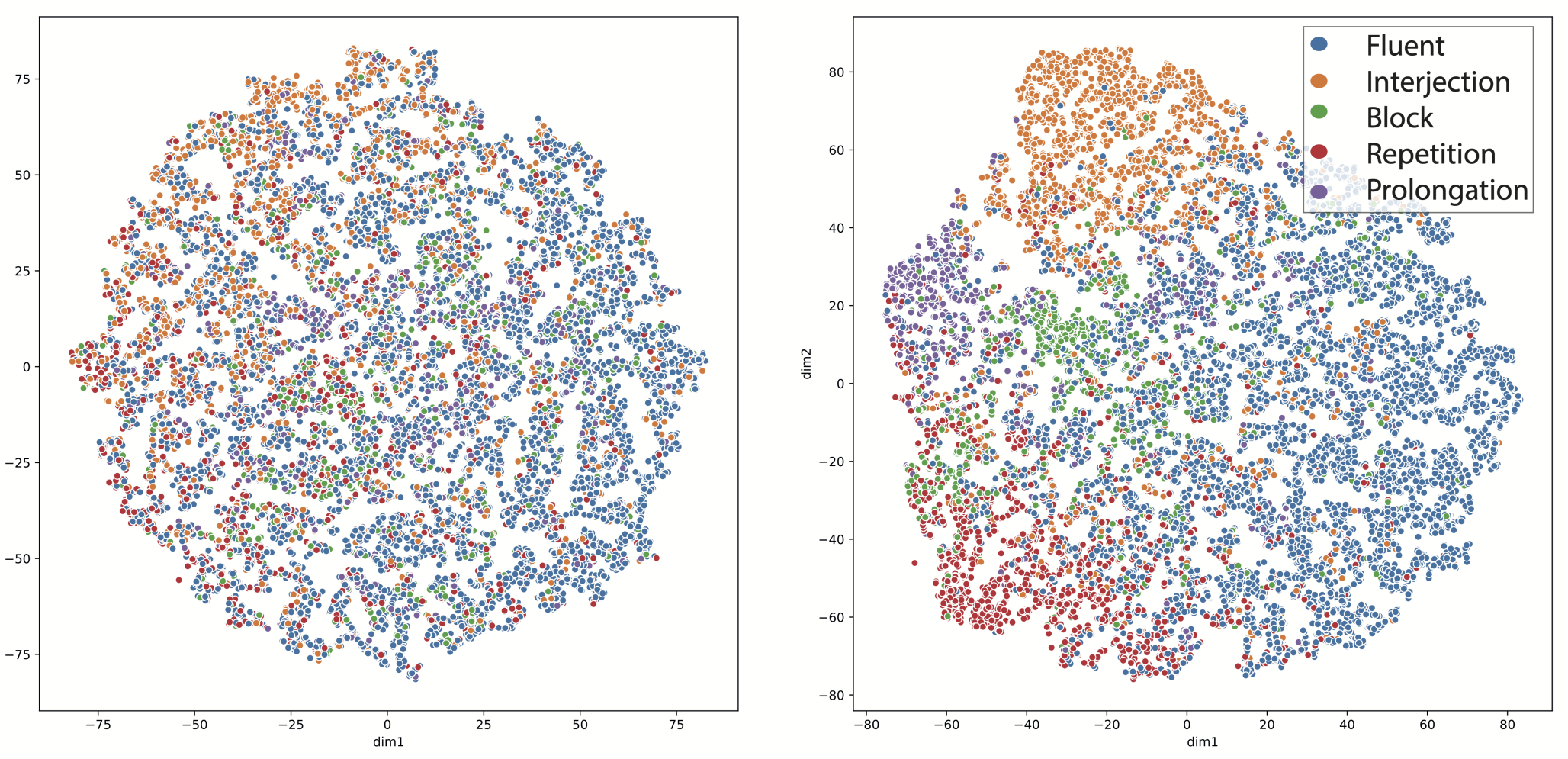
MTL objective function

$$\begin{aligned}\mathcal{L}(\theta_\mathrm{e},\theta_\mathrm{f},\theta_\mathrm{d},\theta_\mathrm{s}) &=(1-\lambda)*\mathcal{L}_\mathrm{stutter}(\theta_\mathrm{e},\theta_\mathrm{f},\theta_\mathrm{d})+\lambda*\mathcal{L}_\mathrm{speaker}(\theta_\mathrm{e},\theta_\mathrm{s})\\ \mathcal{L}_\mathrm{stutter}(\theta_\mathrm{e},\theta_\mathrm{f},\theta_\mathrm{d}) &=\mathcal{L}_\mathrm{fluent}(\theta_\mathrm{e},\theta_\mathrm{f})+\mathcal{L}_\mathrm{disfluent}(\theta_\mathrm{e},\theta_\mathrm{d})\end{aligned} \tag{1}$$

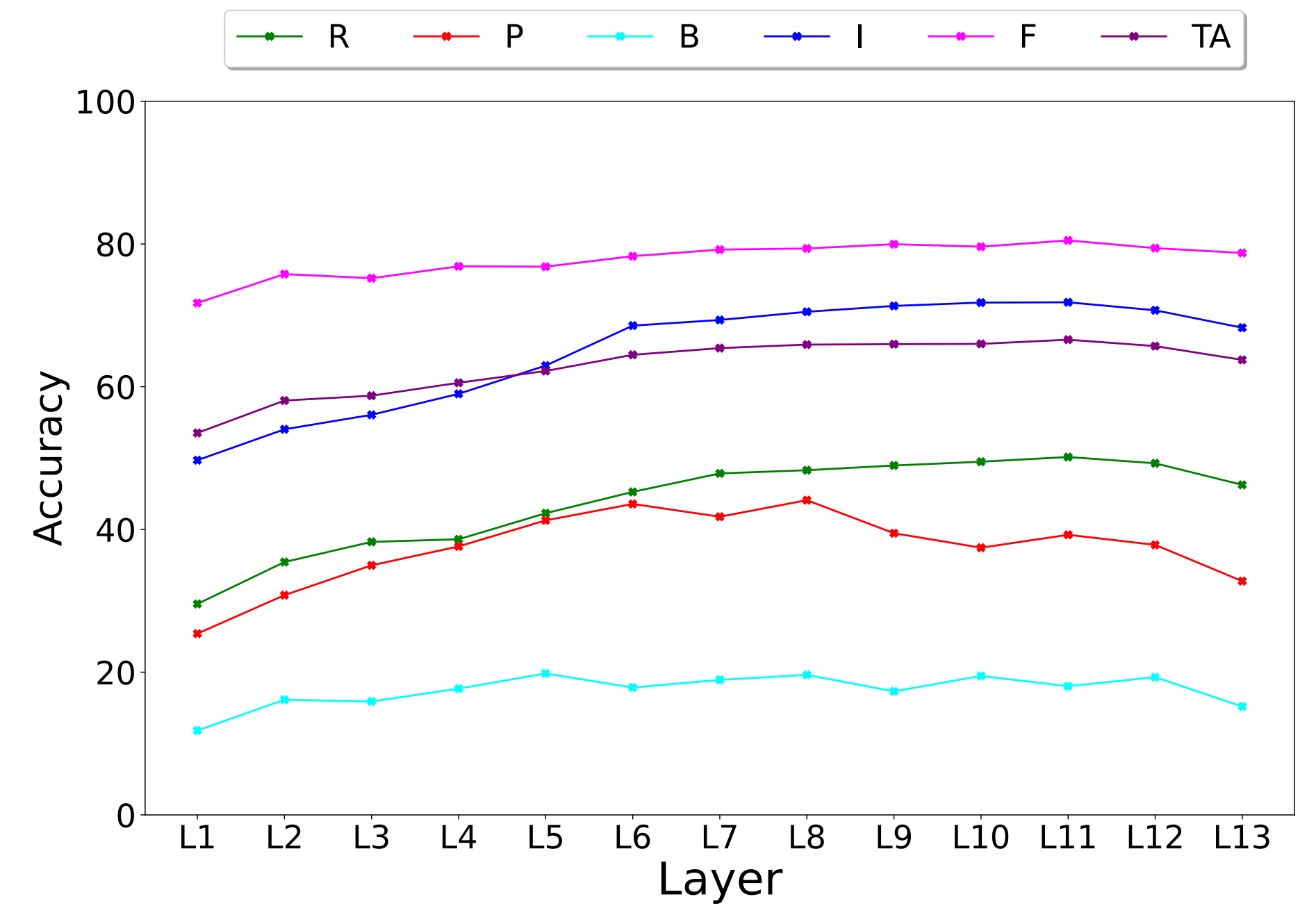




SI using Speech Embeddings



- SOTA speech embeddings like ECAPA-TDNN and Wav2Vec2.0 are widely used in ASR, ED, etc.
- Analyzed Wav2Vec2.0 speech embeddings are more suitable for SI.



Results and Discussion

Table 1. SD results on SEP-28k dataset (TA: Total accuracy, B: Block , F: Fluent , R: Repetition , P: Prolongation , I: Interjection, BL: Baseline, NN: 3 layered fully connected neural network).

Model	R	P	B	I	F	TA
StutterNet [3]	21.99	27.78	1.98	49.99	88.18	60.33
BL (Multi Branch)	28.70	37.89	9.58	57.65	74.43	57.04
MB StutterNet + MTL	31.59	31.62	10.23	58.92	72.14	56.09
MB StutterNet + ADV	27.24	32.89	8.33	56.36	77.10	57.51
NN + Wav2Vec2.0	46.79	40.79	**23.86**	69.54	**84.32**	**68.35**

- BL shows a relative improvement of 26% in disfluent classes.
- λ acts a control parameter for the podcast information to flow through the network.
- The well-formed podcast clusters in the MTL indicate that the model is attempting to learn podcast dependent stuttering information. The clusters, on the other hand, are not observable in the adversarial scenario, and the model is attempting to learn these meta-data invariant robust stutter representations.
- Wav2Vec2.0 captures rich stutter discriminative features.
- Overall improvement with Wav2Vec2.0 is 19.83% in SI.

References

[1] Liam Barrett et al. Systematic review of machine learning approaches for detecting developmental stuttering. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:1160–1172, 2022.

[2] S. A. Sheikh et al. Introducing ecapa-tdnn and wav2vec2. 0 embeddings to stuttering detection. In *Proc. Interspeech 2022 (Under Review)*.

[3] S A. Sheikh et al. Stutternet: Stuttering detection using tdnn. In *Proc. 29th EUSIPCO*, 2021.

[4] S. A. Sheikh et al. Robust stuttering detection via multi-task and adversarial learning. In *Proc. 30th EUSIPCO (Under Review)*, 2022.

[5] Shakeel Sheikh, Md Sahidullah, Fabrice Hirsch, and Slim Ouni. Machine learning for stuttering identification: Review, challenges & future directions. *arXiv preprint arXiv:2107.04057*, 2021.