Automatic Longitudinal Investigation of Multiple Sclerosis Subjects

GÁBOR GOSZTOLYA[1,2], LÁSZLÓ TÓTH[2], VERONIKA SVINDT[3], JUDIT BÓNA[4], ILDIKÓ HOFFMANN[3,5]

[1] HUN-REN–SZTE Research Group on Artificial Intelligence, Szeged, Hungary
[2] University of Szeged, Institute of Informatics, Szeged, Hungary
[3] HUN-REN Research Center for Linguistics, Budapest, Hungary
[4] ELTE Eötvös Loránd University, Dept. of Applied Linguistics and Phonetics, Budapest, Hungary
[5] University of Szeged, Department of Linguistics, Szeged, Hungary

1. THE FOCUS OF THIS STUDY

- Multiple sclerosis, among other symptoms, might deteriorate the speech of the patient
- Due to this, automatic speech analysis can serve as a tool to detect the disease, or monitor its progression
- We employed a standard pathological speech processing workflow (wav2vec 2.0 embeddings as features, SVM as classifier in a nested cross-validation setup)
- We tested the embeddings taken from all layers of the fine-tuned block
- We found that statistically significant improvements could be achieved with the lower one-third of layers (8 layers out of 24)

2. MULTIPLE SCLEROSIS

MULTIPLE SCLEROSIS AND LANGUAGE

- A chronic inflammatory disease of the central nervous system
- Impairments in the patient's gross and fine motor skills
- cca. 60% of MS subjects have some cognitive impairments (cognitive flexibility, disorders of orientation, working-memory limitation, information processing speed) [1]
- cca. one-third of MS patients report (temporary or persistent) speech disorders
- Motor speech disorders (dysarthria, dysphonia); word finding difficulties; limitations of the higher level cognitive processes

THE MULTIPLE SCLEROSIS RECORDINGS USED

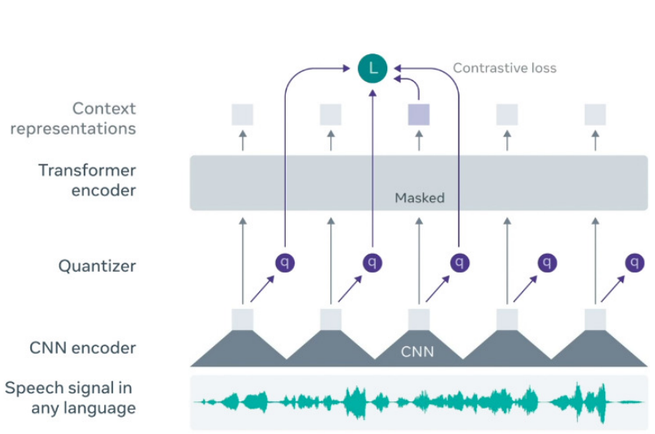
- 23 MS subjects (5/18 m/fm), 22 Healthy Controls (HC) (6/16 m/fm)
- No statistically significant differences between the two groups in demographic attributes (age / gender / years of education)

Two different speech tasks:

(1) Share their opinions about vegetarianism (**Opinion**)
(2) Retell a short historical anecdote just heard (**Narrative Recall**)

3. CLASSIFICATION AND EVALUATION

- We used a wav2vec 2.0 model (XLS-53) fine-tuned on 17 hours of the target language (Hungarian)
- Embeddings from the last layers of the **convolutional** and contextualized (**fine-tuned**) blocks as baseline, aggregated by mean and standard deviation (1024 and 2048 features)



- Embeddings from all 24 hidden layers from the fine-tuned block (mean and standard deviation, 2048 features)
- Support Vector Machines + linear kernel
- **Nested** cross-validation, repeated 5 times with different folds
- Area Under the ROC Curve (AUC) for evaluation
- Mann-Whitney U-test for significance testing

4. RESULTS WITH THE EMBEDDINGS OF THE LAST LAYERS

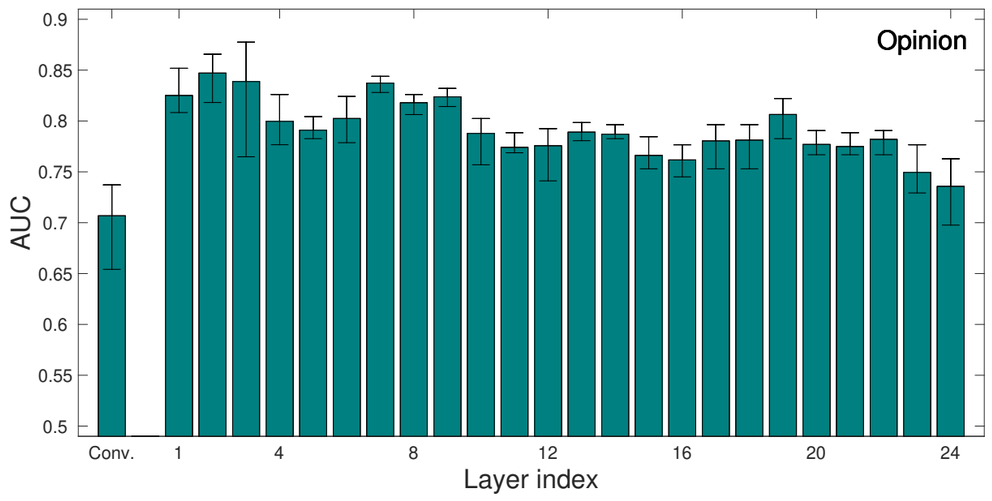
Speech task	Embedding type	AUC Mean	AUC Std.	AUC Range
Opinion	Convolutional	0.707	0.032	[0.654, 0.737]
	Fine-tuned	0.736	0.025	[0.698, 0.763]
Narrative Recall	Convolutional	0.724	0.008	[0.712, 0.733]
	Fine-tuned	0.806	0.014	[0.787, 0.824]

- The results are competitive, and similar to those of previous studies
- Fine-tuned embeddings are somewhat more suitable for MS detection than convolutional ones
- The AUC scores for the Narrative Recall task were higher for both embedding types than for the Opinion one
- Standard deviation values were also smaller ($\Rightarrow$ more robust classification performance)

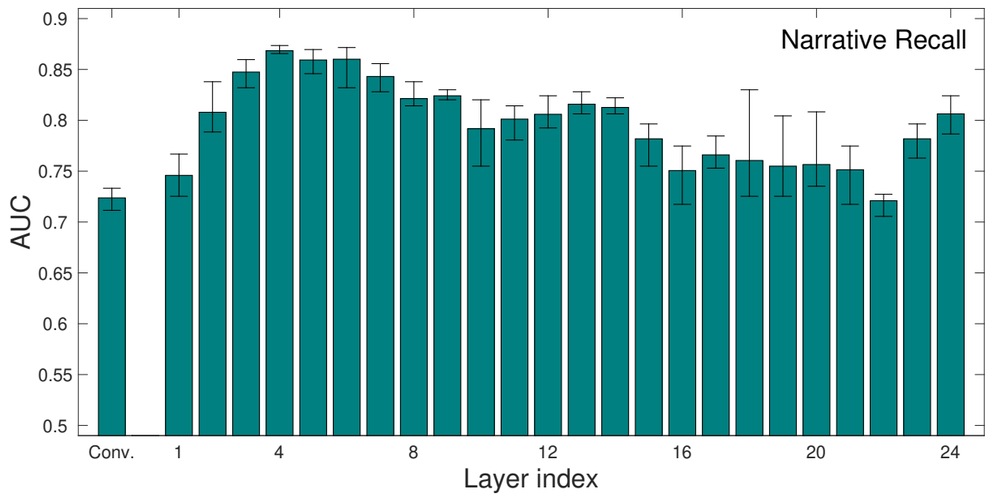
Main references

[1] Dobson et al., "Multiple sclerosis – a review", European Journal of Neurology 2019.
[2] Baevski et al., "wav2vec 2.0: A framework for self-supervised learning of speech representations", Advances in Neural Information Processing Systems, 2020.

5. RESULTS WITH THE EMBEDDINGS OF THE INNER LAYERS



- All inner layers outperformed both baseline approaches
- The improvement is significant in all cases vs. the convolutional layer, and in 20 cases out of 23 vs. the last fine-tuned one



- A significant improvement was achieved vs. the last layer of the convolutional block in 20 cases (out of 23)
- ...but only in 6 cases (out of 23) was the improvement statistically significant vs. the last fine-tuned layer (all in the 1...9 region)

RESULTS FOR SPECIFIC LAYERS

Speech task	Embedding type	AUC Mean	AUC Std.	AUC Range
Opinion	Fine-tuned (#2) $^{**/**}$	0.847	0.018	[0.818, 0.866]
	Fine-tuned (#4) $^{**/**}$	0.800	0.023	[0.777, 0.826]
	Fine-tuned (#6) $^{**/**}$	0.802	0.019	[0.779, 0.824]
	Fine-tuned (#8) $^{**/**}$	0.818	0.008	[0.806, 0.826]
Narrative Recall	Fine-tuned (#2) $^{**/—}$	0.808	0.022	[0.789, 0.838]
	Fine-tuned (#4) $^{**/**}$	0.868	0.004	[0.866, 0.874]
	Fine-tuned (#6) $^{**/**}$	0.860	0.016	[0.832, 0.872]
	Fine-tuned (#8) $^{**/—}$	0.821	0.010	[0.814, 0.838]

- Notation: convolutional / fine-tuned (*: $p < 0.05$, **: $p < 0.01$)
- In most cases a statistically significant improvement was achieved
- Classification performance is also more robust (smaller std.), especially for the 4th layer for the Narrative Recall speech task

IMPROVEMENTS FOR LAYER REGIONS

We also investigated the scores for layer **regions** (40 AUC scores each)

Speech task	Embedding type	AUC Mean	AUC Std.	AUC Range
Opinion	Fine-tuned (#1...#8) $^{**/**}$	0.820	0.028	[0.778, 0.867]
	Fine-tuned (#9...#16) $^{**/**}$	0.783	0.022	[0.749, 0.827]
	Fine-tuned (#17...#24) $^{**/**}$	0.773	0.025	[0.729, 0.810]
Narrative Recall	Fine-tuned (#1...#8) $^{**/*}$	0.832	0.040	[0.740, 0.872]
	Fine-tuned (#9...#16) $^{**/—}$	0.798	0.026	[0.741, 0.828]
	Fine-tuned (#17...#24) $^{**/—}$	0.762	0.033	[0.719, 0.817]

- We outperformed the convolutional embeddings in all cases
- The lowest region is robust for both speech tasks ($p < 0.01$)
- The upper layers (9...23) did not bring a significant improvement for Narrative Recall (or led to a significant drop in the AUC scores)

6. CONCLUSIONS

- We used embeddings taken from the inner hidden layers of the fine-tuned block of a wav2vec 2.0 model
- We obtained statistically significant improvements in most cases, the most effective region being the lowest one
- Combining the embeddings could be a possible extension of this work

This study was supported by the NRDI Office of the Hungarian Ministry of Innovation and Technology (grants K-132460 and TKP2021-NVA-09), and within the framework of the Artificial Intelligence National Laboratory Program (RRF-2.3.1-21-2022-00004).