Adversarial Continual Learning to Transfer Self-Supervised Speech Representations for Voice Pathology Detection

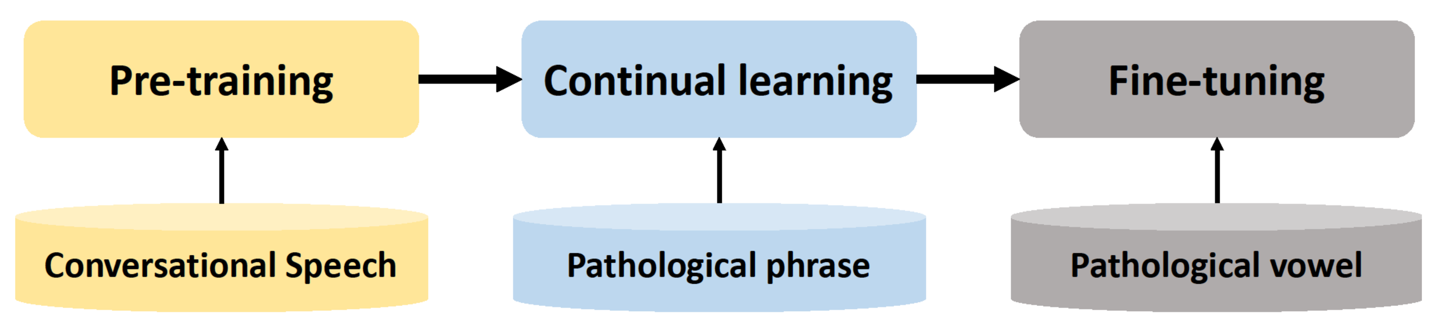
Dongkeon Park[1], Yechan Yu[2], Dina Katabi[3], Hong Kook Kim[1]

[1]GIST, AI Graduate School, [2]Superton Inc., [3]MIT, Department of Electrical Engineering and Computer Science

dongkeon@gm.gist.ac.kr, ato@supertone.ai, dina@csail.mit.edu, hongkook@gist.ac.kr

Introduction

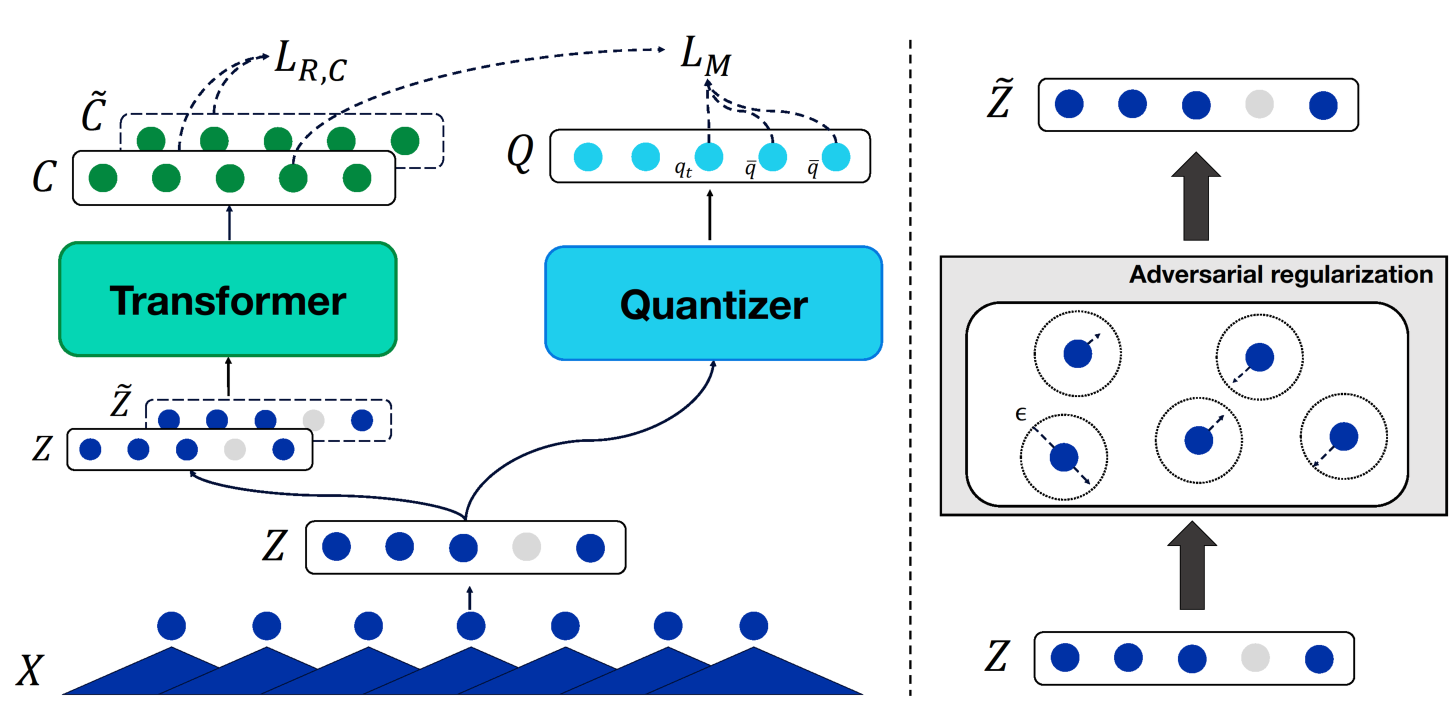
- **Voice Pathology Detection (VPD)**
  - Voice pathology is a common and significant problem
    - With voice quality, pitch, and loudness
    - Due to abnormally vibrate in vocal folds
- **Problem on fine-tuning the SSL model: Domain shift**
  - When fine-tuning pre-trained models to specialized or narrow target domains, the available datasets are significantly smaller
  - The distribution of the downstream task is different from that of the pre-training task
  - Leads to poor performance on the downstream task



- **Motivation: task adaptive pre-training (TAPT)**
  - Inspired by based on TAPT [1], which aims to bridge the gap between pre-training and the target domain by continually learning the pre-trained model on the target dataset
  - TAPT does not necessarily improve performance and degrade the performance of the downstream task as the number of continual learning iterations increases

Proposed Method

- **Continual Learning Based on adversarial regularization on TAPT (A-TAPT)**



- Perturbations to the transformer

$$L_{R,C} = \frac{1}{n}\sum_{t=1}^{T} \max_{\|\tilde{z}_t - z_t\|_p \leq \epsilon} l_s(sim(C(\tilde{z}_t), q_t), sim(C(z_t), q_t))$$

- Perturbations to the quantizer

$$L_{R,Q} = \frac{1}{n}\sum_{t=1}^{T} \max_{\|\tilde{z}_t - z_t\|_p \leq \epsilon} l_s(Gumbel(Q(\tilde{z}_t)), Gubmel(Q(z_t)))$$

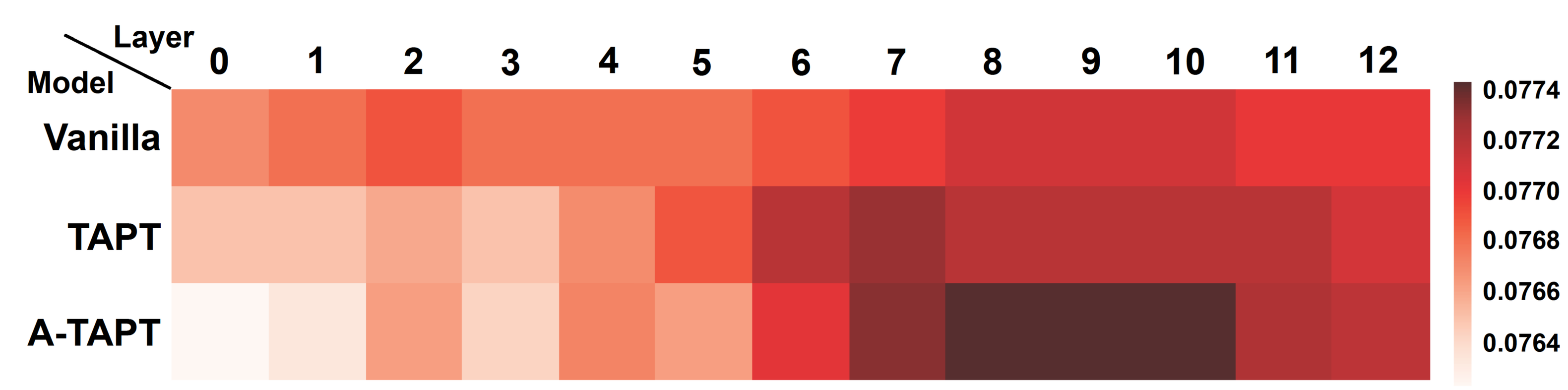
- **A-TAPT method incorporates adversarial regularization into the process of continual learning**
  - Enabling the model to adapt to domain shift through input perturbations
  - Generalize better and overfitting less than the model trained by TAPT

Experiments

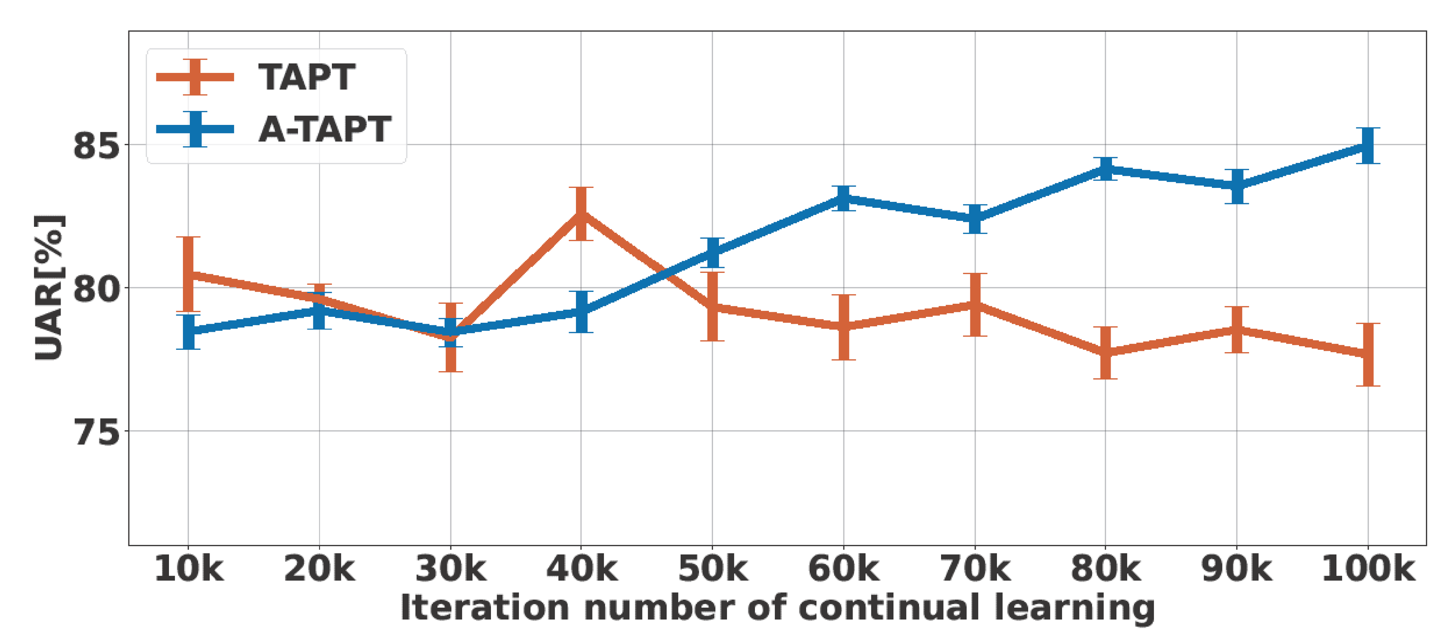
- **Experiment on Saarbrucken Voice Database (SVD)**
  - This improvement was because the continual learning reduced the domain gap between the conversational speech for the pre-training and the pathological speech for the downstream task

Model	ALL	ORGANIC	Avg.
SVM [2]	69.74	76.44	73.09
ResNet50 [2]	69.27	70.87	70.07
Vanilla	81.12±1.12	81.77±0.97	81.45
TAPT	82.21±0.94	83.14±0.74	82.68
A-TAPT ($L_{R,C}$)	**84.97±0.62**	**85.92±0.55**	**85.45**

- **Contribution of each transformer layer**
  - The top layers in a transformer
    - Contribute the most for content and semantic tasks
  - The lower layers
    - Have great impact on speaker characteristics



- **Continual learning iterations**
  - TAPT
    - It degraded with a large standard deviation at iteration range of 30K~50K
  - A-TAPT
    - It resulted in consistent improvement in the UAR as the number of continual learning iterations increased



- **Comparison the performance of with $L_{R,C}$ and $L_{R,Q}$**
  - Applying $L_{R,C}$, resulting in performance degradation
  - z_t and $\tilde{z}_t$ may belong to different codevectors

Model	ϵ	ALL	ORGANIC
A-TAPT ($L_{R,Q}$)	10^{-4}	80.47±1.38	81.41±1.24
A-TAPT ($L_{R,C}$)	10^{-4}	**84.97±0.62**	**85.92±0.55**
A-TAPT ($L_{R,Q} + L_{R,C}$)	10^{-4}	81.49±1.31	82.24±1.13

ACKNOWLEDGEMENT

- This work was supported in part by Institute of Information & communications Technology Planning & evaluation (IITP) grant funded by the Korea government (MSIT) (No.2022-0-00963, Localization Technology Development on Spoken Language Synthesis and Translation of OTT Media Contents) and by the GIST-MIT Research Collaboration grant funded by the GIST in 2023.

REFERENCES

[1] S. Gururangan et al., "Don't stop pretraining: Adapt language models to domains and tasks," in *Proc. ACL, Virtual*, 2020, pp. 8342–8360.
[2] M. Huckvale and C. Buciuleac, "Automated detection of voice disorder in the Saarbrucken voice database: Effects of pathology subset and audio materials," in *Proc. Interspeech*, Brno, Czechia, 2021, pp. 4850–4854.

Gwangju Institute of Science and Technology

ICASSP 2024 KOREA