Yuan Gao

PhD candidate in Speech and Audio Processing Lab. (SAP)
Supervisor: Tatsuya Kawahara
gao.yuan.75x@st.kyoto-u.ac.jp

Research Interest

★ Speech emotion recognition

- Interplay with personality.
- Incorporating semantic information.
- Multistage finetuning.
- Domain adversarial learning.

★ Speech LLM

- Speaking style captioning.

★ Automatic speech recognition

- Emotion specific adapter.
- Error correction.

Education

2022.10 – now **Ph.D in Department of Intelligence Science and Technology**
Supervisor: *Prof. Tatsuya Kawahara*.

2020.10 – 2022.09 **Master in Human Information Science Area**
Supervisor: *Prof. Okada Shogo*.

2019.10 – 2021.06 **Master in College of Intelligence and Computing**
Supervisor: *Prof. Longbiao Wang*.

Working Experience

2024.08 – 2024.09 **Research intern** in NTT
Research: Japanese speech understanding using LLM

Honors

2022.10 – now **Spring fellowship**,
awarded by Japan Science and Technology Agency (JST).

2020.10 – 2022.09 **Tianjin University-JAIST Collaborative Educational Program Scholarship**,
awarded by Japan Advanced Institute of Science and Technology (JAIST).

Skills

Languages	Chinese Native English Bilingual Japanese Elementary
Coding	Python, Matlab, LaTeX, …

Research Publications

Journal Articles

1. Y. Gao, H. Shi, C. Chu, and T. Kawahara, "Multi-attribute learning for multi-level emotion recognition from speech," *IEEE Transactions on Affective Computing*, (under review).
2. Y. Gao, L. Wang, J. Liu, J. Dang, and S. Okada, "Adversarial domain generalized transformer for cross-corpus speech emotion recognition," *IEEE Transactions on Affective Computing*, 2023.

Conference Proceedings

1. Y. Gao, H. Shi, C. Chu, and T. Kawahara, "Enhancing two-stage finetuning for speech emotion recognition using adapters," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2024.
2. Y. Gao, H. Shi, C. Chu, and T. Kawahara, "Speech emotion recognition with multi-level acoustic and semantic information extraction and interaction," in *Proc. Interspeech 2024*, 2024, pp. 1060–1064.
3. H. Shi, Y. Gao, Z. Ni, and T. Kawahara, "Serialized speech information guidance with overlapped encoding separation for multi-speaker automatic speech recognition," in *2024 IEEE Spoken Language Technology Workshop (SLT)*, IEEE, 2024, pp. 193–199.
4. Y. Gao, C. Chu, and T. Kawahara, "Two-stage Finetuning of Wav2vec 2.0 for Speech Emotion Recognition with ASR and Gender Pretraining," in *Proc. INTERSPEECH 2023*, 2023, pp. 3637–3641.
5. J. Tian, D. Hu, X. Shi, *et al.*, "Semi-supervised multimodal emotion recognition with consensus decision-making and label correction," in *Proceedings of the 1st International Workshop on Multimodal and Responsible Affective Computing*, 2023, pp. 67–73.
6. Y. Gao, S. Okada, L. Wang, J. Liu, and J. Dang, "Domain-invariant feature learning for cross corpus speech emotion recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2022, pp. 6427–6431.
7. Y. Gao, J. Liu, L. Wang, and J. Dang, "Domain-adversarial autoencoder with attention based feature level fusion for speech emotion recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2021, pp. 6314–6318.
8. Y. Gao, J. Liu, L. Wang, and J. Dang, "Metric learning based feature representation with gated fusion model for speech emotion recognition.," in *Proc. INTERSPEECH*, 2021, pp. 4503–4507.
9. J. Liu, Z. Liu, L. Wang, Y. Gao, L. Guo, and J. Dang, "Temporal attention convolutional network for speech emotion recognition with latent representation.," in *Proc. INTERSPEECH*, 2020, pp. 2337–2341.