

# Gene-Ping Yang

✉ geneping2009@gmail.com | 🏠 homepages.inf.ed.ac.uk/s2064029/ | in geneping-yang | 📧 CZyca70AAAAJ

## Education

### University of Edinburgh

Ph.D. in CSTR, School of Informatics, Supervisor: Prof. Hao Tang  
Thesis: Geometric Structures of Phonetic Representation in Self-Supervised Models

Edinburgh, Scotland

Oct. 2020 - July 2025

### National Taiwan University

M.S. in Computer Science, Supervisor: Prof. Lin-shan Lee and Prof. Hung-yi Lee  
Thesis: Speech Separation with Time-and-Frequency Cross-Domain Joint Embedding and Clustering  
B.S. in Electrical Engineering

Taipei, Taiwan

Sep. 2017 - June 2019

Sep. 2013 - June 2017

## Research Interests

**Self-Supervised Pre-Training:** Cross-modal alignment of self-supervised speech and text representations.

**Speech Tokenization:** Joint segmentation and discretization of continuous speech into discrete tokens.

**Automatic Speech Recognition:** Optimizing implicit speech-text alignment for robust recognition systems.

**Speech Enhancement and Separation:** Speech extraction and separation in noisy and reverberant environments.

## Work Experience

### Meta

Research Scientist

Menlo Park, CA

August 2025 - Present

- **Enable Rich Human Speech Patterns for Conversational TTS:** drive research into expressive ASR and TTS by incorporating non-verbal speech events, significantly enhancing the conversational flow and emotional resonance of voice interactions.
- **TTS i18n:** lead the development and research on TTS for wearable devices, enhancing inclusivity by expanding language support to reach a broader global audience.
- **TTS Post-Training:** develop a post-training pipeline using SFT and distributed GRPO to optimize speech prosody, naturalness and expressiveness, significantly closing the gap between synthetic and human speech.

### Microsoft Research

Research Intern, Mentor: Sebastian Braun

Redmond, WA

July 2024 - Sep. 2024

- **Multi-microphone speech enhancement and separation:** designed for distributed asynchronous devices, utilizing novel cross-attention methods that outperform previous microphone aggregation and neural beamforming techniques. (Publication 1)

### Apple

Research Scientist Intern, Mentor: Zhen Huang, Stefan Braun, Loren Lugosch

Cambridge, MA

July 2023 - Sep. 2023

- **Speech foundation model:** Developed an encoder-decoder pre-training framework using multilingual speech pseudo labels, demonstrating significant improvement in downstream ASR, particularly with large-scale unlabeled data. (Paper under preparation)

### Amazon

Applied Scientist Intern, Mentor: Yuzong Liu, Yue Gu, Qingming Tang

Sunnyvale, CA

Sep. 2022 - Dec. 2022

- **Self-supervised model distillation:** Developed a task-agnostic objective for distilling self-supervised models, incorporating redundancy reduction and contrastive learning techniques to minimize data bias. (Publication 5)
- **Quantization-aware training for self-supervised models:** Applied a novel quantization-aware training method to quantize both model weights and activations, preserving optimal model performance. (Publication 3)

### Ministry of Science and Technology

Head Teaching Assistant for Formosa Grand Challenge, Organizer: Prof. Hung-yi Lee

Taipei, Taiwan

July 2017 - Oct. 2017

- **Chinese Question Answering Challenge:** Organized a Chinese question-answering challenge, which included collecting data from TV shows, preparing models with PyTorch, TensorFlow, and Keras, and creating documentation to facilitate understanding and implementation of state-of-the-art question-answering models.

## Microsoft

Research and Design Intern

Taipei, Taiwan

July 2016 - Oct. 2016

- **Real-time face tracking and emotion recognition:** Developed a real-time face tracking system, integrating a machine learning model for accurate emotion recognition.
- Robot development with embedded systems: Built robots using MediaTek LinkIt Smart 7688 chips, implementing automated instructions through Azure for enhanced functionality.

## Academic Experience

---

### The Centre for Speech Technology Research (CSTR)

PhD, Supervisor: Prof. Hao Tang

University of Edinburgh

Oct. 2020 - July 2025

- **Self-supervised learning:** Autoregressive predictive coding (Publication 6)
- **Unsupervised speech segmentation / phone-like unit discovery:** Leveraged self-supervised speech features (HuBERT and Wav2vec 2.0) with constrained unsupervised HMMs. (Publication 2)
- **Unsupervised ASR:** Applied optimal transport to align speech embedding with phonetic embedding. (Publication 4)
- **Speech-Text alignment for ASR:** Advanced encoder-decoder framework with supervised attention. (Publication 7)

### Speech Processing and Machine Learning Lab

Master & undergrad research, Supervisor: Prof. Lin-shan Lee and Prof. Hung-yi Lee

National Taiwan University

Feb. 2017 - Aug. 2020

- **Modeling for speech separation:** Integrated time-domain and frequency-domain features to design a feature space that facilitates joint feature clustering. (Publication 10)
- **Improved permutation invariant training:** Developed a novel algorithm to improve speaker permutation optimization for speaker-invariant speech separation. (Publication 9)

## Publications

---

### 1. Distributed Asynchronous Device Speech Enhancement via Windowed Cross-Attention

WASPAA 2025

Gene-Ping Yang, Sebastian Braun

### 2. A Simple HMM with Self-Supervised Representations for Phone Segmentation

SLT 2024

Gene-Ping Yang, Hao Tang

### 3. On-Device Constrained Self-Supervised Learning for Keyword Spotting via Quantization Aware Pre-Training and Fine-tuning

ICASSP 2024 Lecture

Gene-Ping Yang, Yue Gu, Sashank Macha, Qingming Tang, Yuzong Liu

### 4. Towards Matching Phones and Speech Representations

ASRU 2023

Gene-Ping Yang, Hao Tang

### 5. On-device Constrained Self-Supervised Speech Representation Learning for Keyword Spotting via Knowledge Distillation

Interspeech 2023 Oral

Gene-Ping Yang, Yue Gu, Qingming Tang, Dongsu Du, Yuzong Liu

### 6. Autoregressive Predictive Coding: A Comprehensive Study

JSTSP 2022

Gene-Ping Yang, Sung-Lin Yeh, Yu-An Chung, James Glass, Hao Tang

### 7. Supervised Attention In Sequence-to-Sequence Models for Speech Recognition

ICASSP 2022 Lecture

Gene-Ping Yang, Hao Tang

### 8. Stabilizing Label Assignment for Speech Separation by Self-Supervised Pre-Training

Interspeech 2021

Sung-Feng Huang, Shun-Po Chuang, Da-Rong Liu, Yi-Chen Chen, Gene-Ping Yang, Hung-yi Lee

### 9. Interrupted and Cascaded Permutation Invariant Training for Speech Separation

ICASSP 2020 Lecture

Gene-Ping Yang, Szu-Lin Wu, Yao-Wen Mao, Hung-yi Lee, Lin-shan Lee

### 10. Improved Speech Separation with Time-and-Frequency Cross-domain Joint Embedding and Clustering

Interspeech 2019 Oral

Gene-Ping Yang, Chao-I Tuan, Hung-Yi Lee, Lin-shan Lee

## 11. T-Mimi: A Transformer-based Mimi Decoder for Real-Time On-Phone TTS

ICASSP 2026

Haibin Wu, Bach Viet Do, Naveen Suda, Julian Chan, Madhavan C R, Gene-Ping Yang, Yi-Chiao Wu, Naoyuki Kanda, Yossef Adi, Xin Lei, Yue Liu, Florian Metze, Yuzong Liu

## 12. Beyond Words: Towards Effective Modeling of Non-Verbal Vocalizations in Automatic Speech Recognition

Under Review

Gene-Ping Yang, Haibin Wu, Peng Su, Ruizhe Huang, Suwon Shon, ..., Yuzong Liu

## Teaching

---

### Maching Learning , Lecturer: Hao Tang

University of Edinburgh

Lead discussions in five tutorial sessions, covering optimizations, learning, and hands-on implementation.

### Applied Deep Learning , Lecturer: Yun-Nung Chen

National Taiwan University

Lead coursework on anime face generation based on text descriptions, implementing conditional GANs and testing various GAN objectives such as WGAN, improved WGAN, and ACGAN.

### Machine Learning and Having it Deep and Structured, Lecturer: Hung-yi Lee

National Taiwan University

Involve in the design of coursework on sequence labeling (phone prediction) using joint training of CNN and RNN, video caption generation with a seq2seq-based model, and automatic game playing through deep reinforcement learning.

### Machine Learning , Lecturer: Hung-yi Lee

National Taiwan University

Lead coursework on predicting PM 2.5 levels in the air using a hand-crafted linear regression model with gradient descent, based on historical air quality indicators.

## Honors

---

2025 IEEE Travel Grant for WASPAA

2022 IEEE Signal Processing Society Travel Grant for ICASSP

2020–2025 Full funding award for doctoral research

2019–2020 Advanced Speech Technologies Scholarship

2019 Foundation For The Advancement of Outstanding Scholarship

2019 International Speech Communication Association (ISCA) Travel Grant

2019 Appier scholarship for Artificial Intelligence and Information Technology

2019 Higher Education Sprout Project

## Services

---

Reviewer ICLR, WACV, AACL, IJCAI, IJCNN, Computer Speech & Language, IEEE Signal Processing Letters, ICASSP, Interspeech, ASRU