

# Yicheng Gu

yichenggu@link.cuhk.edu.cn | yicheng.gu@aalto.fi | github.com/VocodexElysium

## KEY SKILLS

---

**Solid Domain Knowledge:** Extensive knowledge of Deep Learning and Digital Signal Processing; specifically applied to Neural Vocoder, Neural Audio Codec, and Digital Audio Effects.

**Multi-disciplinary:** Experience encompassing Research, Open Source Systems, Dataset Development, Commercial Sound Design, and Music Production.

## EDUCATION

---

### The Chinese University of Hong Kong, Shenzhen

*Bachelor of Engineering in Computer Science and Engineering*

Shenzhen, China

*Sep. 2022 – Present*

- GPA: 3.965 / 4.0, Ranking: 1 / 298

### Aalto University

*Exchange Student in Computer Science*

Espoo, Finland

*Sep. 2024 – Jun. 2025*

- GPA: 4.75 / 5.0

## EXPERIENCE

---


### Human Language Technology Lab

*Research Assistant, School of Data Science, CUHK-Shenzhen*

Shenzhen, China

*Oct. 2022 – Present*

Supervisor: Prof. Zhizheng Wu

- Singing Voice Conversion
  - \* Investigated characteristics and the complementary role of different Content-Based Features for the Singing Voice Conversion system.
- Speech Synthesis
  - \* Proposed a large-scale and diversified dataset, Emilia, to facilitate speech research, which has been utilized by various international institutions.
- Neural Vocoder
  - \* Built a Discriminator based on the Constant-Q Transform (CQT) and Continuous Wavelet Transform (CWT) to improve the Vocoder's synthesis quality. The methods have been implemented and supported by NVIDIA BigVGAN .

### Shanghai AI Laboratory

*Research Assistant*

Shanghai, China

*Dec. 2023 – Mar. 2025*

Supervisor: Dr. Yanhong Zeng

- Video-to-Audio Generation
  - \* Integrated IP-Adapter and Sound Event Detection model to existing Audio Generation pipeline, obtaining both Audio-Visual Synchronization and Text-Controllability.

### Acoustic Lab

*Visiting Scholar, DICE, Aalto University*

Espoo, Finland

*Sep. 2024 – Jun. 2025*

Supervisor: Prof. Lauri Juvela

- Digital Audio Effects
  - \* Proposed the SOTA Neural Autotune benchmarking against Melodyne.
  - \* Proposed the first large-scale and diverse dataset for Virtual Analog Modeling.

## PROJECTS

---

### Amphion

*An Open-Source Audio, Music and Speech Generation Toolkit*

- Migrate and adapt various well-known, widely used, or SOTA vocoders into our system.
- Integrate comprehensive Objective Evaluation Metrics to the framework.

### FoleyCrafter

*Bring Silent Videos to Life with Lifelike and Synchronized Sounds*

- Integrate baseline model and develop the temporal adapter for controlling timestamp details.

## PUBLICATIONS

---

**Yicheng Gu**, Xueyao Zhang, Liumeng Xue, Haizhou Li, Zhizheng Wu, "An Investigation of Time-Frequency Representation Discriminators for High-Fidelity Vocoder," TASLP.

**Yicheng Gu**, Xueyao Zhang, Liumeng Xue, Zhizheng Wu, "Multi-Scale Sub-Band Constant-Q Transform Discriminator for High-Fidelity Vocoder," ICASSP 2024.

**Yicheng Gu**, Chaoren Wang, Zhizheng Wu, Lauri Juvela, "Neurodyne: Neural Pitch Manipulation with Representation Learning and Cycle-Consistency GAN," Interspeech 2025.

**Yicheng Gu\***, Runsong Zhang\*, Lauri Juvela, Zhizheng Wu, "Diff-SSL-G-Comp: Towards a Large-Scale and Diverse Dataset for Virtual Analog Modeling," DAFx 2025.

Xueyao Zhang\*, Liumeng Xue\*, **Yicheng Gu\***, Yuancheng Wang\*, et al., "Amphion: An Open-Source Audio, Music and Speech Generation Toolkit," SLT 2024.

Haorui He\*, Zengqiang Shang\*, Chaoren Wang\*, Xuyuan Li\*, **Yicheng Gu**, et al., "Emilia: An Extensive Multilingual and Diverse Speech Dataset for Large-Scale Speech Generation," SLT 2024.

Xueyao Zhang, **Yicheng Gu**, et al., "Leveraging Content-based Features from Multiple Acoustic Models for Singing Voice Conversion," Machine Learning for Audio Workshop at NeurIPS 2023.

Xueyao Zhang, Zihao Fang, **Yicheng Gu**, et al., "Leveraging Diverse Semantic-based Audio Pretrained Models for Singing Voice Conversion," SLT 2024.

Yiming Zhang, **Yicheng Gu**, Yanhong Zeng, et al., "FoleyCrafter: Bring Silent Videos to Life with Lifelike and Synchronized Sounds," Submitted to IJCV.

Haorui He\*, Zengqiang Shang\*, Chaoren Wang\*, Xuyuan Li\*, **Yicheng Gu**, et al., "Emilia: A Large-Scale, Extensive, Multilingual, and Diverse Dataset for Speech Generation," Submitted to TASLP.

## HONORS AND AWARDS

---

- The Nobel Class (Top 1, 2024)
- The Academic Performance Scholarship, Class A (Top 1%, 2024)
- The Academic Performance Scholarship, Class B (Top 3%, 2023)
- "LanHuaYing" Scholarship (Top 10 admitted students in Zhejiang Province, 2022)
- University Entrance Scholarship (Top 0.7% in Zhejiang Province, 2022)

## CROSS-DISCIPLINARY ABILITIES

---

**Music Production:** experienced in Composing, Mixing, and Mastering for different genres including Pop, Electronic, etc.

**Sound Design:** experienced in Sound Designing for movie, game, etc.