Philip Anastassiou

www.philipanastassiou.com | philip.anastassiou@pm.me |

Education

Columbia University — New York, NY

M.S. Computer Science (Machine Learning Track) — Dec. 2022

- Coursework included: artificial intelligence, machine learning, advanced topics in neural networks and deep learning, advanced topics in spoken language processing and speech recognition, natural language processing, computer vision, databases, analysis of algorithms, and independent research.

B.A. Drama and Theatre Arts — May 2018

Experience

Adobe — San Jose, CA

Senior Applied Scientist, Generative AI/ML — May 2026 – Present

- Conduct research as part of **Firefly Foundry** team on **generative audio and multimodal AI**.

TikTok/ByteDance — San Jose, CA

Research Scientist — Jul. 2024 – May 2026

Speech Synthesis Engineer — Dec. 2022 – Jun. 2024

Software Engineer Intern — May 2022 – Nov. 2022

- Conducted research as part of **Seed** team on **generative speech AI** for TikTok and ByteDance products.
  - Contributed to development of voice agents within **Speech Interaction and Learning** group, leveraging large language models, diffusion and flow models, and reinforcement learning.
- Co-developed **Doubao Real-Time Voice Model**, an **end-to-end joint speech-text model for real-time conversations** with human-like naturalness and enhanced emotional intelligence.
  - **Surpassed OpenAI's GPT-4o Advanced Voice Mode** (4.36/5 > 3.18/5) in MOS evaluations for Mandarin speakers and deployed on **Doubao AI platform (100M DAU)**.
  - **Led R&D of English adaptation of real-time speech-to-speech model**, spanning data curation, pre-training, supervised fine-tuning, reinforcement post-training, evaluation, and productization.
  - **Deployed multi-timbre voice agents on global Dola AI platform (10M DAU)**, driving engagement gains in key English-speaking markets.
- Co-developed **Seed-TTS**, an audio foundation model for **human-like speech generation** with state-of-the-art **zero-shot in-context learning capabilities**.
- Led **VoiceShop** project for simultaneous, disentangled multi-attribute voice editing, **deploying two style conversion voice filters on TikTok (1B DAU)**.

Research

My research interests include deep generative modeling, self-supervised representation and transfer learning, zero-shot learning, and knowledge distillation. I'm broadly interested in unifying neural audio generation and understanding to develop general auditory intelligence across modalities.

Publications

Seed Team, ByteDance, *"Seed-TTS: A family of high-quality versatile speech generation models,"* arXiv:2406.02430, Jun. 2024. [Paper, Demo, Code (1.5K+ ⭐ on GitHub)]

Philip Anastassiou*, Zhenyu Tang*, Kainan Peng, Dongya Jia, Jiaxin Li, Ming Tu, Yuping Wang, Yuxuan Wang, Mingbo Ma (*equal contribution), *"VoiceShop: A unified speech-to-speech framework for zero-shot voice editing,"* arXiv:2404.06674, Apr. 2024. [Paper, Demo]

Patents

Philip Anastassiou, Zhenyu Tang, Jiaxin Li, Kainan Peng, Dongya Jia, Qiao Tian, Mingbo Ma, Yuping Wang, Yuxuan Wang, *"Identity-preserving zero-shot many-to-many accent and speech style conversion via bottleneck-to-bottleneck and diffusion modeling,"* CN120108409A, ByteDance Ltd., 2025.

Services

Program Committee/Paper Reviewer: ICML Workshop on Machine Learning for Audio (2026), Association for Computational Linguistics (2026), IEEE International Conference on Acoustics, Speech, and Signal Processing (2026), AAAI Conference on Artificial Intelligence (2026, 2025), IEEE Transactions on Audio, Speech, and Language Processing (2026, 2025), ACM International Conference on Multimedia (2025), IEEE Signal Processing Letters (2025, 2024).

Projects

VAE-GAN for Speech-to-Speech Style Transfer — Dec. 2021

- Implemented proposed variational autoencoder-generative adversarial network (VAE-GAN) architecture with domain-specific decoders for non-autoregressive speech-to-speech style transfer based on AlBadawy et al. (2020) and Bonnici et al. (2021). [Code]

Skills

Languages: Python, Java, C, MATLAB, SQL, Unix shell scripting.
Software: Git, PyTorch, PyTorch Lightning, TensorFlow, Keras, Pandas, PySpark, SciPy, NumPy, Matplotlib, Scikit-Learn, FFmpeg, Librosa, Kaldi, ESPNet, Praat, NLTK, Conda, LaTeX.