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, and analysis of algorithms.
- Conducted research advised by Professor Peter Belhumeur on developing self-supervised optical character recognition model whose training procedure is guided by separate language model via student-teacher knowledge distillation paradigm without use of ground-truth labels.

B.A. Drama and Theatre Arts — May 2018

Experience

TikTok/ByteDance — San Jose, CA
Research Scientist — Jul. 2024 – Present
Speech Synthesis Engineer — Dec. 2022 – Jun. 2024
Software Engineer Intern — May 2022 – Nov. 2022

- Conduct research on Seed/Doubao team (previously Speech, Audio, and Music Intelligence) on state-of-the-art neural speech synthesis and audio processing algorithms for various TikTok and ByteDance products to empower content creation and consumption.

Research

My research interests include deep generative modeling, self-supervised representation and transfer learning, zero-shot learning, and knowledge distillation. I'm especially interested in neural end-to-end learning for audio and natural language processing tasks.

Publications

Seed Team, ByteDance, *"Seed-TTS: A family of high-quality vesatile speech generation models,"* arXiv:2406.02430, Jun. 2024. [Paper, Demo, Code (1K+ stars)]

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-based modeling,"* CN202311659609.5, pending CN patent application filed by ByteDance Ltd., 2024.

Prof. Services

Program Committee: AAAI Conference on Artificial Intelligence (2025); **Paper Reviewer:** IEEE Signal Processing Letters (2024).

Presentations

Nokia Bell Labs — Murray Hill, NJ
Experiments in Art and Technology Program ("Squashed" Software) — Feb. 2019

- Developed software for Nokia Bell Labs E.A.T. program at Columbia Computer Music Center to produce musically desirable digital artifacts in audio signals with lossy LAME MP3 encoder.

Lamont-Doherty Earth Observatory — Palisades, NY
Research as Art Exhibition ("Novel Song" Software) — May 2018

- Developed software to convert text of fictional novels into music with VADER Python package for sentiment analysis to interpolate emotional valence scores into harmonic relationships in RTcmix.

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]

Deep Convolutional Spectral Autoencoder for Neural Audio Synthesis — May 2021

- Implemented deep convolutional autoencoder in Python using TensorFlow for audio synthesis of musical notes based on Engel, et al. (2017), trained on subset of Google Magenta's *NSynth* corpus.

Skills

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