


EDUCATION	<p>Columbia University New York, NY <i>M.S. Computer Science (Machine Learning Track)</i> Dec. 2022</p> <ul style="list-style-type: none"> • 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. <p><i>B.A. Drama and Theatre Arts</i> May 2018</p>
EXPERIENCE	<p>TikTok/ByteDance San Jose, CA <i>Research Scientist</i> Jul. 2024 – Present <i>Speech Synthesis Engineer</i> Dec. 2022 – Jun. 2024 <i>Software Engineer Intern</i> May 2022 – Nov. 2022</p> <ul style="list-style-type: none"> • Conduct research as part of Seed team on generative speech AI for TikTok and ByteDance products. <ul style="list-style-type: none"> ◦ Contribute 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. <ul style="list-style-type: none"> ◦ 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 for 100M daily active users (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. <ul style="list-style-type: none"> ◦ Expanded model capabilities including improved accent and prosodic controllability, instruction following, sarcasm and humor understanding, decoupled acoustic-text emotion reasoning, and singing. ◦ Deployed multi-timbre voice agents on global Dola AI platform (10M DAU), driving engagement gains in key English-speaking markets, including +21.49% average voice messages sent per user (Philippines), +11.44% (Malaysia), and up to +48.36% call duration (UK). • 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	<p>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 speech, vision, and text modalities.</p>
PUBLICATIONS	<p>Seed Team, ByteDance, "<i>Seed-TTS: A family of high-quality versatile speech generation models</i>," arXiv:2406.02430, Jun. 2024. [Paper, Demo, Code (1.5K+  on GitHub)]</p> <p>Philip Anastassiou*, Zhenyu Tang*, Kainan Peng, Dongya Jia, Jiaxin Li, Ming Tu, Yuping Wang, Yuxuan Wang, Mingbo Ma (*equal contribution), "<i>VoiceShop: A unified speech-to-speech framework for zero-shot voice editing</i>," arXiv:2404.06674, Apr. 2024. [Paper, Demo]</p>
PATENTS	<p>Philip Anastassiou, Zhenyu Tang, Jiaxin Li, Kainan Peng, Dongya Jia, Qiao Tian, Mingbo Ma, Yuping Wang, Yuxuan Wang, "<i>Identity-preserving zero-shot many-to-many accent and speech style conversion via bottleneck-to-bottleneck and diffusion modeling</i>," CN120108409A, ByteDance Ltd., 2025.</p>
SERVICES	<p>Program Committee/Paper Reviewer: 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).</p>
PROJECTS	<p>VAE-GAN for Speech-to-Speech Style Transfer Dec. 2021</p> <ul style="list-style-type: none"> • 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	<p>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.</p>