

<b>EDUCATION</b>	<p><b>Columbia University</b>  <i>M.S. Computer Science (Machine Learning Track)</i></p> <ul style="list-style-type: none"> <li>• 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.</li> </ul> <p><i>B.A. Drama and Theatre Arts</i></p>	<p>New York, NY          Dec. 2022            May 2018</p>
<b>EXPERIENCE</b>	<p><b>TikTok/ByteDance</b>  <i>Research Scientist</i>  <i>Speech Synthesis Engineer</i>  <i>Software Engineer Intern</i></p> <ul style="list-style-type: none"> <li>• Conduct research as part of <b>Seed</b> team on <b>generative speech AI</b> for TikTok and ByteDance products.             <ul style="list-style-type: none"> <li>◦ Contribute to development of voice agents within <b>Speech Interaction and Learning</b> group, leveraging large language models, diffusion and flow models, and reinforcement learning.</li> </ul> </li> <li>• Co-developed <b>Doubao Real-Time Voice Model</b>, an <b>end-to-end joint speech-text model for real-time conversations</b> with human-like naturalness and enhanced emotional intelligence.             <ul style="list-style-type: none"> <li>◦ <b>Surpassed OpenAI's GPT-4o Advanced Voice Mode</b> (4.36/5 &gt; 3.18/5) in MOS evaluations for Mandarin speakers and deployed on <b>Doubao AI platform</b> for <b>100M daily active users</b> (DAU).</li> </ul> </li> <li>• <b>Led R&amp;D of English adaptation of real-time speech-to-speech model</b>, spanning data curation, pre-training, supervised fine-tuning, reinforcement post-training, evaluation, and productization.             <ul style="list-style-type: none"> <li>◦ <b>Expanded model capabilities</b> including improved accent and prosodic controllability, instruction following, sarcasm and humor understanding, decoupled acoustic-text emotion reasoning, and singing.</li> <li>◦ <b>Deployed multi-timbre voice agents on global Dola AI platform (10M DAU)</b>, driving engagement gains in key English-speaking markets, including <b>+21.49%</b> average voice messages sent per user (Philippines), <b>+11.44%</b> (Malaysia), and up to <b>+48.36%</b> call duration (UK).</li> </ul> </li> <li>• Co-developed <b>Seed-TTS</b>, an audio foundation model for <b>human-like speech generation</b> with state-of-the-art <b>zero-shot in-context learning capabilities</b>.</li> <li>• Supported audio data curation pipeline for <b>Seedance 2.0</b>, ByteDance's flagship multi-modal audio-video generation model, outperforming Sora 2 Pro and Veo 3.1 in internal T2V and I2V benchmarking.</li> <li>• Led <b>VoiceShop</b> project for simultaneous, disentangled multi-attribute voice editing, <b>deploying two style conversion voice filters on TikTok (1B DAU)</b>.</li> </ul>	<p>San Jose, CA          Jul. 2024 – Present          Dec. 2022 – Jun. 2024          May 2022 – Nov. 2022</p>
<b>RESEARCH</b>	<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>	
<b>PUBLICATIONS</b>	<p><b>Seed Team</b>, ByteDance, "<i>Seed-TTS: A family of high-quality versatile speech generation models</i>," arXiv:2406.02430, Jun. 2024. [<a href="#">Paper</a>, <a href="#">Demo</a>, <a href="#">Code</a> (1.5K+ ★ on GitHub)]</p> <p><b>Philip Anastassiou*</b>, 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. [<a href="#">Paper</a>, <a href="#">Demo</a>]</p>	
<b>PATENTS</b>	<p><b>Philip Anastassiou</b>, 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>	
<b>SERVICES</b>	<p><b>Program Committee/Paper Reviewer:</b> <a href="#">Association for Computational Linguistics</a> (2026), <a href="#">IEEE International Conference on Acoustics, Speech, and Signal Processing</a> (2026), <a href="#">AAAI Conference on Artificial Intelligence</a> (2026, 2025), <a href="#">IEEE Transactions on Audio, Speech, and Language Processing</a> (2026, 2025), <a href="#">ACM International Conference on Multimedia</a> (2025), <a href="#">IEEE Signal Processing Letters</a> (2025, 2024).</p>	
<b>PROJECTS</b>	<p><b>VAE-GAN for Speech-to-Speech Style Transfer</b></p> <ul style="list-style-type: none"> <li>• Implemented proposed variational autoencoder-generative adversarial network (VAE-GAN) architecture with domain-specific decoders for non-autoregressive speech-to-speech style transfer based on <a href="#">AlBadawy et al. (2020)</a> and <a href="#">Bonnici et al. (2021)</a>. [<a href="#">Code</a>]</li> </ul>	<p>Dec. 2021</p>
<b>SKILLS</b>	<p><b>Languages:</b> Python, Java, C, MATLAB, SQL, Unix shell scripting.  <b>Software:</b> Git, PyTorch, PyTorch Lightning, TensorFlow, Keras, Pandas, PySpark, SciPy, NumPy, Matplotlib, Scikit-Learn, FFmpeg, Librosa, Kaldi, ESPNet, Praat, NLTK, Conda, <math>\LaTeX</math>.</p>	