Haoqi Li

haoqi.github.io Google Scholar LinkedIn haoqilele [at] gmail [dot] com

PROFILE

- **Senior Research Scientist** with industry experience at Netflix, Hume AI, and AWS AI, backed by 10+ years of research in **speech/audio AI** and **affective computing**.
- Experience developing and deploying **speech systems**, with technical leadership in controllable speech generation and speech quality assessment, alongside work in multilingual ASR, speaker verification, and evaluation of content-understanding agents.
- Authored **20+ peer-reviewed publications** in top-tier venues. **Ph.D.** from USC's SAIL Lab focused on understanding human emotion and behavior through speech and language using deep learning.

WORKING EXPERIENCE

Netflix Feb. 2024 - Present

Senior Research Scientist, Globalization — Netflix Research & DSE *Los Gatos, CA*

Work on **ML for speech and language** in media localization, including data preparation, model development, LLM post-training, and evaluation. Bring an end-to-end understanding of workflows spanning AI dubbing, speech recognition, subtitle generation, translation, and quality assessment.

Synthetic Voice

- **Generative AI Architecture:** Technical lead for synthetic voice initiatives spanning LLM-based autoregressive (AR), non-autoregressive (NAR), and hybrid architectures, shaping the technical roadmap. Developed proprietary dubbing models that outperformed commercial systems in human mean opinion score (MOS) evaluations.
- **End-to-End Data Engineering:** Built and scaled data preparation pipelines to ingest, clean, and extract paired audio and text ($\sim 100k$ hrs) from heterogeneous movie assets.
- **Linguistic Modeling & Prosody:** Optimized Japanese TTS models to address context-dependent Kanji pronunciation and intonation errors with limited training data.
- **Speech & Multimodal Evaluation:** Combined objective metrics with human evaluations to assess speech quality, investigated acoustic biases in speech LLM judges, and evaluated multimodal lip-sync quality for dubbing. Developed **SpeechCritic**, a diagnostic speech judge using supervision calibrated with limited human preferences to produce multidimensional judgments and detailed, evidence-grounded rationales.
- **Cross-Lingual TTS & Accent Reduction:** Developed region-aware masking for flow-matching TTS, reducing perceived accent by **18–51%** on seen language pairs and **35–55%** on unseen source languages versus bilingual adaptation alone, with no measurable loss in intelligibility or naturalness.

LLM-based Multilingual Speech Recognition

- **Model Training & Evaluation:** Trained and evaluated LLM-based multilingual ASR models for subtitle and dubbing workflows and diagnosed model and training pipeline issues. The team's system outperformed the production baseline in evaluations spanning **30+ languages** and diverse content types.
- **ASR Robustness & Reliability:** Applied targeted token-level preference optimization (Anti-doom) in collaboration with colleagues to reduce repetition loops in generated transcripts and improve output reliability.

LLM-as-a-Judge & Content-Understanding Agent Evaluation

- **Judge Development:** Lead development of prototype judges for content-understanding agents, combining citation grounding, evidence faithfulness, and rubric-based tag judgments with supporting rationales.
- **Editorial Verification:** Combined retrieved evidence, taxonomy checks, and rule-based human-review routing to evaluate editorial tags.

Mentorship

- Mentored junior researchers and summer interns in speech and multimodal AI.

Hume AI Inc. Feb. 2023 - Feb. 2024
Senior Research Scientist *Sunnyvale, CA*

- **Empathic Voice Interface:** Technical lead for controllable speech generation and conversational systems for Hume AI's first empathic voice interface; developed and launched its speech generation engine.
- **Nuanced Emotion Recognition:** Developed speech emotion recognition systems to capture nuanced human expression and support emotionally responsive interactions.

Amazon AWS AI Jan. 2021 - Feb. 2023
Applied Scientist *Sunnyvale, CA*

Speaker Verification and Spoofing Detection

- Developed ECAPA-TDNN-based models for speaker recognition and speech spoofing detection and launched the production model for **Amazon Connect Voice ID**.
- Built parallel data-generation pipelines using open-source and commercial TTS systems to create large-scale spoofing-detection datasets.
- Developed self-supervised speaker verification methods using Siamese networks and self-supervised regularization.

Accented Speech Recognition

Improved ASR production models by >3% on African American accented speech while maintaining overall performance by implementing novel debiasing and fairness techniques.

- Architected domain adversarial training frameworks with accent classifiers integrated into existing end-to-end ASR systems within ESPnet.
- Developed TTS and data-splicing augmentation strategies, generating synthetic speech with customized phonetic rules and concatenating accented speech segments to expand training data.

Signal Analysis and Interpretation Laboratory (SAIL)
University of Southern California Aug. 2014 - Nov. 2020
Research Assistant *Los Angeles, CA*

Worked on deep learning techniques for modeling human behaviors from speech and language to improve well-being and identify psychological disorders. Investigated sparse networks, transfer learning and unsupervised learning under constrained low-resource data conditions. Representative works are:

- Linking emotions to behaviors: A novel framework that investigates the inferential capacity and contextual importance between emotions and behaviors via deep transfer learning.

- Beh2Vec: Developed contextual speech representations for behavior modeling using triplet-enhanced networks.
- SD-DNN: Sparsely connected and disjointly trained deep neural networks for low-resource behavioral annotation.

Amazon Lab126 May 2020 - Aug. 2020
Applied Scientist Intern *Sunnyvale, CA*

Designed an audio-visual emotion recognition model with domain adaptation for human–robot interaction.

- Explored domain transfer learning techniques on emotional utterances collected under different emotion elicitation approaches, particularly with limited labeled target samples.
- First-authored a paper published at Interspeech 2021.

JD AI Research - JD.com May 2018 - Aug. 2018
Research Intern *Mountain View, CA*

Worked on voice conversion, end-to-end Text-to-Speech (TTS) and speaker-invariant emotion representation learning.

- Implemented and optimized Chinese Text-to-Speech systems based on VoiceLoop.
- Designed the neural network for extracting speaker-invariant emotion representations from speech via adversarial training.
- First-authored a paper published at ICASSP 2020.

Sony PlayStation Acoustic Analysis Group May 2017 - Aug. 2017
Research Intern *San Mateo, CA*

Worked on the prediction of funny movie scenes via reinforcement learning.

- Designed a reinforcement learning framework for affective detection from movie footage.
- First-authored a paper published at ICASSP 2018.

SELECTED PUBLICATIONS

The **full list** of publications is available on my **Google Scholar profile**.

[1] **Haoqi Li**, Shivam Mehta, Ravi Teja Gadde, Yinghong Lan. "Region-Aware Masking for Accent-Robust Cross-Lingual Text-to-Speech." arXiv preprint, 2026; submitted to ICASSP 2027.

[2] Mingyue Huo, Shivam Mehta, Bhavin Jawade, Yinghong Lan, **Haoqi Li**. "SpeechCritic: Learning a Diagnostic Speech Judge from Limited Human Preferences." arXiv preprint, 2026; submitted to ICLR 2027.

[3] Mingyue Huo, Shivam Mehta, Bhavin Jawade, Yinghong Lan, **Haoqi Li**. "Louder, Longer, Livelier: Acoustic Shortcuts and Underspecified Rationales in Speech LLM Judges." arXiv preprint, 2026; submitted to ACL ARR 2026.

[4] Mufan Sang, **Haoqi Li**, Fang Liu, Andrew O. Arnold, Li Wan. "Self-Supervised Speaker Verification with Simple Siamese Network and Self-Supervised Regularization." ICASSP, 2022.

[5] **Haoqi Li**, Yelin Kim, Cheng-Hao Kuo, Shrikanth Narayanan. "Acted vs. Improvised: Domain Adaptation for Elicitation Approaches in Audio-Visual Emotion Recognition." Interspeech, 2021.

[6] **Haoqi Li**, Brian Baucom, Shrikanth Narayanan, Panayiotis Georgiou. "Unsupervised Speech Representation Learning for Behavior Modeling using Triplet Enhanced Contextualized Networks." Computer Speech and Language, vol. 70, article 101226, 2021.

[7] **Haoqi Li**, Ming Tu, Jing Huang, Shrikanth Narayanan, Panayiotis Georgiou. "Speaker-invariant Affective Representation Learning via Adversarial Training." ICASSP, 2020.

[8] **Haoqi Li**, Brian Baucom, Panayiotis Georgiou. "Linking emotions to behaviors through deep transfer learning." PeerJ Computer Science, 6:e246, 2020.

[9] Raghuveer Peri, **Haoqi Li**, Krishna Somandepalli, Arindam Jati, Shrikanth Narayanan. "An Empirical Analysis of Information Encoded in Disentangled Neural Speaker Representations." Odyssey: The Speaker and Language Recognition Workshop, 2020.

[10] Prashanth Gurunath Shivakumar, **Haoqi Li**, Kevin Knight, and Panayiotis Georgiou. "Learning from Past Mistakes: Improving Automatic Speech Recognition Output via Noisy-Clean Phrase Context Modeling." APSIPA Transactions on Signal and Information Processing, vol. 8, 2019.

[11] **Haoqi Li***, Sandeep Nallan Chakravarthula*, Shao-yen Tseng, Maija Reblin, Panayiotis Georgiou (*Equal Contribution). "Predicting Behavior in Cancer-Afflicted Patient and Spouse Interactions using Speech and Language." Interspeech, 2019.

[12] **Haoqi Li**, Naveen Kumar, Ruxin Chen, Panayiotis Georgiou. "A Deep Reinforcement Learning Framework for Identifying Funny Scenes in Movies." ICASSP, 2018.

[13] **Haoqi Li***, Shao-yen Tseng*, Brian Baucom, and Panayiotis Georgiou (*Equal Contribution). "Honey, I Learned to Talk": Multimodal Fusion for Behavior Analysis. ACM International Conference on Multimodal Interaction (ICMI), 2018.

[14] **Haoqi Li**, Brian Baucom, Panayiotis Georgiou. "Unsupervised Latent Behavior Manifold Learning from Acoustic Features: audio2behavior." ICASSP, 2017.

[15] **Haoqi Li**, Brian Baucom, Panayiotis Georgiou. "Sparsely Connected and Disjointly Trained Deep Neural Networks for Low Resource Behavioral Annotation: Acoustic Classification in Couples' Therapy." Interspeech, 2016.

PATENTS

[1] Ruxin Chen, Naveen Kumar, **Haoqi Li**. "Deep reinforcement learning framework for characterizing video content." U.S. Patent No. 11,386,657. 12 Jul. 2022.

[2] Ruxin Chen, Naveen Kumar, **Haoqi Li**. "Deep reinforcement learning framework for sequence level prediction of high dimensional data." U.S. Patent Application No. 17/852,602.

EDUCATION

University of Southern California, Los Angeles, USA
Ph.D. in Electrical Engineering — Aug. 2014 - Nov. 2020
Dissertation: Behavior understanding from speech under constrained conditions: exploring sparse networks, transfer and unsupervised learning
M.S. in Computer Science — Aug. 2017 - May 2018

Chinese Academy of Sciences, Beijing, China
M.S. in Signal and Information Processing — Aug. 2011 - Jun. 2014

Xidian University, Xi'an, China
B.S. in Electronic Information Engineering, GPA: 91.3/100 (Rank 1/220) — Aug. 2007 - Jun. 2011

TEACHING EXPERIENCE

Teaching Assistant, University of Southern California Fall 2019

EE 660 Machine Learning from Signals: Foundations and Methods
Instructed by Prof. Keith Jenkins

Teaching Assistant, University of Southern California Spring 2018

EE 503 Probability for Electrical and Computer Engineers
Instructed by Prof. Vijay Kumar

ACADEMIC SERVICE

Conference Reviewer

International Conference on Acoustics, Speech, & Signal Processing (ICASSP)
Conference of the International Speech Communication Association (INTERSPEECH)
IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)
IEEE Spoken Language Technology Workshop (SLT)
International Conference on Affective Computing & Intelligent Interaction (ACII)
ACM International Conference on Multimodal Interaction (ICMI)
Annual Meeting of the Association for Computational Linguistics (ACL)
Empirical Methods in Natural Language Processing (EMNLP)
International Workshop on Spoken Dialogue Systems Technology (IWSDS)

Journal Reviewer

IEEE Transactions on Affective Computing
IEEE/ACM Transactions on Audio, Speech, and Language Processing
IEEE Transactions on Multimedia
IEEE Signal Processing Letters
IEEE Access
Computer Speech and Language
ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)
APSIPA Transactions on Signal and Information Processing
EURASIP Journal on Audio, Speech, and Music Processing

HONORS AND AWARDS

ICASSP 2020 Student Travel Grant — IEEE SPS, 2020
National Scholarship for Graduate Students — Ministry of Education of China, 2013
Merit Student of University of Chinese Academy of Sciences — 2012
Outstanding Graduate in Shaanxi Province — 2011
Outstanding Graduate of Xidian University — 2011
National Scholarship of China (Top 1%) — Ministry of Education of China, 2009
Undergraduate Mathematical Contest in Modeling (CUMCM), 1st Prize in Shaanxi Province — 2009
Outstanding Students Scholarship of Xidian University — 2009-2011
Excellent Freshman Scholarship of Xidian University — 2008

TECHNICAL SKILLS

Programming Languages	Python, Bash scripting
Tools	PyTorch, ESPnet, Kaldi, Linux, AWS, Git
Modeling	Speech AI, LLMs, Diffusion & Flow Matching Models
Evaluation	LLM-as-a-Judge, Human Evaluation