

韩冰

手机：(+86) 13262615051 · 邮箱：hanbing97@sjtu.edu.cn
性别：男 · 主页：<https://hunterhuan.github.io>



教育背景

上海交通大学，计算机科学与技术，博士	2022.6 - 至今
<ul style="list-style-type: none">听觉认知与计算声学实验室 (AudioCC Lab)，导师：钱彦旻主要关注：音频表征、音频合成、说话人识别、富音频分析	
上海交通大学，计算机科学与技术，硕士	2020.9 - 2022.6
<ul style="list-style-type: none">推荐免试入学	
上海交通大学，计算机科学与技术，本科	2016.9 - 2020.6
<ul style="list-style-type: none">IEEE 荣誉试点班、校级优秀毕业生	

科研经历

听觉认知与计算声学实验室 AudioCC Lab，计算机学院，上海交通大学

- 交交-音色 & 情感可控语音对话系统 Demo 链接
 - 构建基于 SFT LLM 的情感语音对话系统，利用情感提示 token 实现语音情感引导
 - 解耦情感及音色，实现任意音色的情感控制
 - 通过文本续写任务实现音频编码器与 LLM 的高效模态对齐
 - 使用 LLM 的 CFG 策略，正在探索同时兼顾 prompt 说话人风格和情感控制的语音合成系统。
- 基于自监督预训练的音频表征
 - 基于多尺度建模和信号引导，提升音频自监督训练的泛化性和高效性
 - 工业领域信号（声音和振动）的多采样率统一表征建模，工业信号 Benchmark 达到 SOTA
- 利用自监督学习的鲁棒高效说话人验证
 - 基于 DINO 及伪标签策略不依赖说话人标签构建说话人验证系统
 - 在不使用任何标签的情况下基本达到有标签训练的性能
- 采用局部建模强化自注意力的说话人验证模型
 - 强化 Transformer 局部建模能力，说话人等错误率降低 50% 以上

实习经历

混元-腾讯-算法实习生	2025.5 - 至今
<ul style="list-style-type: none">• 研究端到端语音对话系统<ul style="list-style-type: none">- 探索语音对话系统中的理解、生成一体化- 探索音频理解中的推理 CoT	
微软亚洲研究院 MSRA-算法实习生	2023.8 - 2024.4
<ul style="list-style-type: none">• 基于语言模型架构的零样本语音合成 Demo 链接<ul style="list-style-type: none">- 利用音素单调对齐的训练解码策略来加强音素和声学序列之间的联系，让自回归语音合成更稳定- 采用编解码器合并的方法对浅的量化层中的离散编码进行降采样，推理速度提升 3 倍以上	
SAMI-字节跳动-算法实习生	2023.2 - 2023.7
<ul style="list-style-type: none">• 基于扩散模型的音乐编辑算法 Demo 链接<ul style="list-style-type: none">- 首次提出音乐编辑任务，填补了音乐编辑的空白- 基于扩散模型实现长时音乐乐器的增、删、提取、替换等操作	
AI Lab-字节跳动-算法实习生	2019.10 - 2020.1
<ul style="list-style-type: none">• 小说人物语言理解系统（语音合成本文前端）<ul style="list-style-type: none">- 基于半监督和合成数据，构建了命名实体识别-性别判定-指代消解的小说人物关系分析框架	

奖项 & 比赛 & 开源

奖项：

- 入选首届中国电子学会-腾讯博士生科研激励计划（混元大模型专项） 2024
 - 获科研经费 10 万元，全国共 17 位入选
- ISCSLP 最佳学生论文 2024
- 华为奖学金-博士 2024
- 华为奖学金-硕士 2021
- 上海交通大学优秀毕业生 2020
- 国家励志奖学金 2019

比赛：

- DCASE 国际声音场景和事件检测与分类挑战赛 2024
 - 任务 1、任务 3，冠军
- DCASE 国际声音场景和事件检测与分类挑战赛 2023
 - 任务 3，亚军
- 对话短语音说话人日志挑战赛 ISCSLP 2022
 - 任务 1，冠军
- CN-Celeb 说话人识别挑战赛 2022
 - 任务 1，冠军
- Vox-Celeb 说话人识别挑战赛 2022
 - 任务 1，亚军

开源工具：

- WeSpeaker-说话人表征建模工具 **核心贡献者**
 - 自监督说话人训练
 - 后端打分优化
 - 新模型 & 优化函数

个人优势

- 尝试过的任务稍广，对音频的理解、合成等任务均有一定了解，熟悉主流自监督训练（Self-supervised Learning）、Speech VAE、Codec、自回归及非自回归语音合成等基线模型
- 除语音外，对非语音的音频建模也有一定了解

主要文章列表

谷歌学术主页（引用量: 544 ; H 指数: 16）

论文在投：

- **B. Han**, A. Jiang, Y. Qian, “Exploring Self-Supervised Audio Models for Generalized Anomalous Sound Detection,” submitted to TASLP
- H. Sun, **B. Han**, Y. Qian, “ContextSpeech: A Large-Scale Real-Human Speech Corpus with Context-Aware Descriptions”, submitted to NeurIPS.
- C. Le, **B. Han**, Y. Qian, “SimulMEGA: MoE Routers are Advanced Policy Makers for Simultaneous Speech Translation”, submitted to NeurIPS.
- P. Fan, A. Jiang, S. Zhang, Z. Lv, **B. Han**, etc. “FISHER: A Foundation Model for Multi-Modal Industrial Signal Comprehensive Representation,” in the review.

已发表论文：

- **B. Han**, J. Dai, X. Song, Y. Qian, “InstructME: An instruction guided music edit framework with latent diffusion models,” in IJCAI 2024
- L. Meng, L. Zhou, S. Liu, S. Chen, **B. Han**, etc. “Autoregressive speech synthesis without vector quantization” in ACL 2025
- **B. Han**, L. Zhou, S. Liu, Y. Qian, “VALL-E R: Robust and Efficient Zero-Shot Text-to-Speech Synthesis via Monotonic Alignment,” in NeurIPS 2024 Workshop

- **B. Han**, Z. Chen, Y. Qian, “Self-Supervised Learning with Cluster-Aware-DINO for High-Performance Robust Speaker Verification,” in Trans.ASLP 32, 529-541
- Z. Chen, **B. Han**, Y. Qian “Attention-based encoder-decoder end-to-end neural diarization with embedding enhancer” in Trans.ASLP 32, 1636-1649
- **B. Han**, W. Huang, Y. Qian, “Data-Efficient Low-Complexity Acoustic Scene Classification via Distilling and Progressive Pruning,” in ICASSP 2025
- **B. Han***, Z. Lv*, Y. Qian, “Exploring Large Scale Pre-Trained Models for Robust Machine Anomalous Sound Detection,” in ICASSP 2024
- **B. Han**, Z. Chen, Y. Qian, “Local Information Modeling with Self-Attention for Speaker Verification,” in ICASSP 2022
- **B. Han**, Z. Chen, Y. Qian, “Exploring Binary Classification Loss for Speaker Verification,” in ICASSP 2023
- Z. Chen*, **B. Han***, X. Xiang, H. Huang, B. Liu, Y. Qian, “Build a SRE Challenge System: Lessons from VoxSRC 2022 and CNSRC 2022,” in Interspeech 2023
- **B. Han**, Z. Chen, Y. Qian, “Self-Supervised Speaker Verification Using Dynamic Loss-Gate and Label Correction,” in Interspeech 2022
- S. Wang, Z. Chen, **B. Han**, H. Wang, etc., “Advancing speaker embedding learning: Wespeaker toolkit for research and production” in Speech Communication 162, 103104
- X. Zheng, A. Jiang, **B. Han**, Y. Qian, etc., “Improving anomalous sound detection via low-rank adaptation fine-tuning of pre-trained audio models” in SLT 2024
- W. Huang, **B. Han**, Z. Chen, S. Wang, Y. Qian, “Prototype and Instance Contrastive Learning for Unsupervised Domain Adaptation in Speaker Verification” in ISCSLP 2024 (Best Student Paper)
- **B. Han**, Z. Chen, Y. Qian, “Improving DINO-based Self-Supervised Speaker Verification by Progressive Cluster-aware Training,” in ICASSP SASB 2023
- **B. Han**, Z. Chen, Z. Zhou, Y. Qian, “The SJTU System for Short-duration Speaker Verification Challenge 2021,” in Interspeech 2021