

## 教育背景

- 2013.09 **北京大学**, 智能科学与技术, 硕士学位
- 2016.06 研究方向: 语音识别与合成
- 2009.09 **北京邮电大学**, 智能科学与技术, 学士学位
- 2013.06 研究方向: 人工智能、自然语言处理

## 工作经历

- 2020.12 **贝壳**, 资深算法工程师, 负责语音识别、语音大模型
- 至今 负责智能家居交互、训练场、智能客服等项目; 语音大模型能力建设; 荣获年度蓝血贝壳。
- 2017.12 **滴滴**, 资深算法工程师, 负责语音识别
- 2020.12 负责行程中录音、智能客服、外呼等项目; 鲁棒语音识别、中英文混响语音识别能力建设。
- 2017.04 **百度**, 高级语音工程师, 负责语音合成
- 2017.12 研发百度地图明星音合成等项目; 语音合成声学建模。

## 专业技能

- 专业能力 语音识别与合成、多模态语音大模型
- 编程能力 Python、Git、Kaldi、Transformers

## 项目经历

### 语音大模型方向 (2023-2025)

- 2024.05 **语音大模型与对话数据**, 项目负责人, Project Space, 贝壳
- 2025.06
  - 搭建端到端语音大模型, 强大语音理解与生成能力、超低延迟, 带来交互体验跨越提升。
  - 解决对话数据不足问题, 构建语音对话生产与质检系统, 产出 10 万小时高质量对话数据。
  - 自研语音大模型 Ke-Omni, 在对话理解榜单VoiceBench上超越 Qwen2-Audio。
  - 语音-文本模态鸿沟研究, 粗粒度、细粒度定量分析、揭示模态鸿沟来源 (EMNLP 投稿)。
- 2025.03 **音频推理大模型**, 项目负责人, Github, 贝壳
- 2025.04
  - 构建音频推理模型, 通过强化学习引入深度思考过程, 提升复杂任务的理解和推理能力。
  - 基于 Qwen2.5-Omni, GRPO 强化学习训练, 在音频推理榜单 MMAU 上达到 SOTA。
- 2023 **新一代语音识别系统 Belle-Whisper**, 项目负责人, Github, 贝壳
- 2024
  - 打造领域泛化统一的语音识别, 取代深度定制, 更多模型参数、数据, 提升精度和鲁棒性。
  - 基于基座的低成本领域自适应, 在 VR 带看、客服等场景落地, 比深度定制绝对提升 2%。
- 2023 **面向大语言模型的数据与评测解决方案**, 项目负责人, 贝壳
- 2024
  - 搭建贝壳领域大语言模型的数据与评测平台, 支撑居住领域大模型的训练与评测。
  - 具备数据收集、去重、过滤、配比等功能及模型自挑选能力。天池大模型数据挑战季军。

### 语音识别方向 (2018-2023)

- 2022.05 **小贝训练场语音识别**, 语音负责人, 贝壳
- 2023.02
- 小贝训练场是提升经纪人作业能力的训练平台, 涵盖新房、二手、家装等业务范围。
  - 提升房产领域的语音识别精度, 搭建模型自适应训练平台, 以及热词、后处理干预模块。
  - 分钟级的新词识别与快速干预, 满足业务个性化需求, 整体识别精度由 92% 提升到 **97%**。
- 2020.12 **智能家居语音交互**, 语音负责人, 贝壳
- 2022.04
- 打造“小海智家”智能家居产品, 构建智能家居语音交互能力, 实现近场和远场语音识别。
  - 家居语音交互场景面临噪声、混响以及领域不匹配的问题, 严重影响识别性能。
  - 为解决噪声混响问题, 语音叠加噪声、混响模拟真实环境, 识别精度由 87% 提升到 **92%**。
  - 为解决领域不匹配问题, 通过语言模型自适应训练, 识别精度进一步提升至 94%。
- 2018.07 **行程中录音鲁棒语音识别**, 项目负责人, 滴滴
- 2020.04
- 行程中录音语音识别是保障出行安全的基础, 面临环境复杂、远场、无数据等问题。
  - 基于 LAS 框架构建鲁棒语音识别, 数据增强扩充 (加车噪、加混响), 提升模型鲁棒性。
  - 研发半监督训练框架, 筛选线上大量弱标签数据, 对模型快速优化, 识别精度提升 5%。
  - 行程中录音语音识别率从 50% 提升到 **84%**, 1 个月内落地可用, 并实现 **100%** 全覆盖。
- 2017.12 **基于端到端方法的中英文夹杂 (CODE-SWITCHING) 语音识别**, 核心研发, 滴滴
- 2018.10
- 普遍存在的中英文夹杂现象给语音识别带来很大挑战, 传统级联方案系统复杂、精度低。
  - 基于 Attention 的端到端语音识别框架, 端到端实现中英文夹杂的语音识别。
  - 探究了建模单元 (char、subword), 解码策略, 多任务学习 (联合 CTC、融合语种信息)。
  - 在 SEAME 数据上达到 SOTA, MER **34%**。应用到多方言混合语音识别, 整体提升 **10%**。

## 语音合成方向 (2013-2017)

- 2017.05 **基于神经网络的语音合成及在极小量数据下的明星音合成**, 核心研发, 百度
- 2017.12
- 利用大规模语音数据, 构建基于循环神经网络的声学模型和时长模型。
  - 通过模型自适应与数据平衡, 0.5 小时极小量明星音实现了高质量明星音合成 MOS **3.8**。
  - 明星音在百度地图导航上广泛应用, 数量超过竞品, 并得到地图部门的高度认可。
- 2013.10 **基于深层循环神经网络的语音合成声学建模研究**, 研究课题, 北大
- 2016.04
- 为解决 GMM 趋向均值导致的过平滑问题, 在语音合成声学建模中引入循环神经网络。
  - 探讨不同层级文本特征对声学建模的影响; 引入韵律信息, 进一步提高声学模型的性能。
  - 基于深层循环神经网络实现语音合成, 声学特征误差由基线系统的 5.8 dB 降低到 **5.0 dB**。

---

## Publications

- A comparable study of modeling units for end-to-end mandarin speech recognition. *ISCSLP*, 2018.
- Towards end-to-end code-switching speech recognition. *arXiv preprint arXiv:1810.13091*, 2018.
- Gigaspeech: An evolving, multi-domain asr corpus with 10,000 hours of transcribed audio. *arXiv preprint arXiv:2106.06909*, 2021.
- Kespeech: An open source speech dataset of mandarin and its eight subdialects. *NeurIPS*, 2021.
- Chathome: Development and evaluation of a domain-specific language model for home renovation. *arXiv preprint arXiv:2307.15290*, 2023.
- Advancing speech language models by scaling supervised fine-tuning with over 60,000 hours of synthetic speech dialogue data. *arXiv preprint arXiv:2412.01078*, 2024.
- Sari: Structured audio reasoning via curriculum-guided reinforcement learning. *arXiv preprint arXiv:2504.15900*, 2025.