郑瑞晨 | Rui-Chen Zheng

联系方式: 181-8926-6050 | zhengruichen@mail.ustc.edu.cn
个人主页: https://zhengrachel.github.io/

教育经历

➢ **中国科学技术大学 语音及语言信息处理国家工程研究中心** *2021.09 – 2026.06(预计)*

硕博连读

- 导师：凌震华 教授
- 毕业论文题目: 结合发音特征的语音合成和增强方法
- GPA: 3.9/4.3 (前 3%)

➢ **中国科学技术大学 电子工程与信息科学系(6系) 电子信息工程专业** *2017.09 – 2021.06*

工学学士

- 毕业论文: 无文本语音合成方法实践
- GPA: 3.89/4.3 (前 5%)
- 信息科技英才班，辅修工商管理专业

实习经历

➢ **阿里巴巴通义实验室语音团队** *2025.03 – 今*

研究型实习生

课题简介：构建可变帧率的单码本语音编解码器

研究内容：

- 基于目前最先进的单码本语音编解码器 WavTokenizer
- 通过对隐藏特征采用峰值密度聚类算法，得到帧率可变的语音编解码器

产出成果：所提出的帧率可变的语音编解码器在下游语义基准取得更好的结果，并在下游 TTS 任务取得更好的自然度。

研究经历

➢ **语音编解码器**

■ **通过码本内与码本间优化方法提升基于残差矢量量化的语音编解码器性能** *2024.05 - 2024.08*

课题简介：基于残差矢量量化的语音编解码器存在码本崩溃问题，从而减少有效码本大小并导致性能不佳。

研究内容：

- 码本内优化：结合在线聚类策略和码字均衡损失，以确保码本利用率的平衡和高效
- 码本间优化：通过最小化连续量化之间的相似性来提高量化特征的多样性

产出成果：作为针对不同结构的语音编解码器的通用方法大幅提高了语音编解码器的编解码质量，同时对下游基于语音编解码器表示的语音-文本大模型的性能有所改善。相关论文已被***TASLP***(语音处理领域顶刊)接收。

■ **增强语音编解码器的噪声鲁棒性** *2024. 11 - 2025.02*

课题简介：噪声鲁棒性是神经语音编解码器开发中的一个关键挑战，特别是对于现实世界的语音通信场景。

研究内容：

- 在码字级别模拟噪声扰动, 与选择最接近的码本向量的传统量化不同, 从前 K 个最接近的候选向量中进行概率采样
- 从 RVQ 中的最后一个 VQ 到第一个 VQ 逐步引入噪声鲁棒性

产出成果：在仅需干净数据训练的情况下提高了语音编解码器的噪声鲁棒性，同时保留对干净语音的编解码能力。相关论文正在IEEE Signal Processing Letters审稿中。

➢ **结合大语言模型的智能语音对话系统** *2024.09 – 2025.02*

课题简介：基于开源大语言模型(LLM)构建资源高效的语音交互模型.

研究内容：

- 使用 CosyVoice-1.0 模型对开源语音对话数据集 VoiceAssistant-400K 进行答案语音的合成
- 利用 HuBERT 预训练模型和 KMeans 聚类提取语音的离散化特征
- 在实验室条件下(2 张 A100)，在 LLaMA-3.1-8B 和 LLaMA-3.2-1B 模型上复现该领域文章 LLaMA-Omni

产出成果：基本实现LLaMA-Omni原文效果，仍在做进一步改进。

➢ **结合多模态的语音生成方法**

■ **从静默模式下的唇部视频与舌部超声图像中生成语音** *2022.03 - 2022.10, 2023.09 - 2024.03*

课题简介：静默发声模式下说话人只激活发音器官而不发出声音, 无法像传统唇形视频到语音生成一样使用监督训练。

研究内容：

- 为静默模式下的唇部视频和舌部超声图像生成伪造声学特征用于监督训练：通过动态时间扭曲(DTW)对齐不同域的发音特征，或通过构造基于唇形视频的配音模型
- 对编码器进行静默域与标准有声域的对抗训练，提高编码器得到的发音特征表示的鲁棒性

产出成果：大幅提高从静默模式下的唇部视频与舌部超声图像中生成的语音的可懂度。两篇会议论文分别被***MM 2024***(CCF A类会议)和***ICASSP 2023***(语音处理领域顶会)录用。

■ **结合舌部超声图像的视听语音增强方法** *2022.11 - 2023.08*

课题简介：结合超声舌像来提升基于唇部的视听语音增强系统的性能，并解决在推理过程中无法获取超声舌像的挑战

研究内容：

· 构建具有语音-唇形视频-舌部超声图像三种输入的多模态语音增强网络

· 通过在训练过程中应用知识蒸馏和记忆存储器，使模型具备在推理时无舌部图像输入进行三模态语音增强的能力

产出成果：大幅提高结合唇形视频的视听语音增强质量。一篇期刊论文和一篇会议论文分别被***TASLP***(语音处理领域顶刊)和***INTERSPEECH 2023***(语音处理领域顶会)录用。

论文发表

➢ 结合多模态的语音生成方法

[1]. **Rui-Chen Zheng**, Yang Ai, Zhen-Hua Ling, "Speech Reconstruction from Silent Lip and Tongue Articulation by Diffusion Models and Text-Guided Pseudo Target Generation", in *Proc. ACM Multimedia 2024*, pages 6559-6568. (**多模态处理领域顶级会议, CCF-A类会议**)

[2]. **Rui-Chen Zheng**, Yang Ai, Zhen-Hua Ling, "Incorporating Ultrasound Tongue Images for Audio-Visual Speech Enhancement", *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (Volume: 32), pages 1430-1444. (**语音处理领域顶级期刊, 声学学科SCI-1区期刊, CCF-B类期刊**)

[3]. **Rui-Chen Zheng**, Yang Ai, Zhen-Hua Ling, "Incorporating Ultrasound Tongue Images for Audio-Visual Speech Enhancement Through Knowledge Distillation", in *Proc. INTERSPEECH 2023*, pages 844-848. (**语音处理领域顶级会议, CCF-C类会议**)

[4]. **Rui-Chen Zheng**, Yang Ai, Zhen-Hua Ling, "Speech Reconstruction from Silent Lip and Tongue Articulation by Pseudo Target Generation and Domain Adversarial Training", in *Proc. ICASSP 2023*, pages 1-5. (**语音处理领域顶级会议, CCF-B类会议**)

➢ 语音编解码及语音合成

[1]. **Rui-Chen Zheng**, Hui-Peng Du, Xiao-Hang Jiang, Yang Ai, Zhen-Hua Ling, "ERVQ: Enhanced Residual Vector Quantization with Intra-and-Inter-Codebook Optimization for Neural Audio Codecs", 已被TASLP(**语音处理领域顶级期刊, 声学学科SCI-1区期刊, CCF-B类期刊**)接收。

[2]. Yao Guo, Yang Ai, **Rui-Chen Zheng**, Hui-Peng Du, Xiao-Hang Jiang, Zhen-Hua Ling, "Vision-Integrated High-Quality Neural Speech Coding", accepted by *INTERSPEECH 2025*.

[3]. Xiao-Hang Jiang, Yang Ai, **Rui-Chen Zheng**, Zhen-Hua Ling, "A Streamable Neural Audio Codec With Residual Scalar-Vector Quantization for Real-Time Communication", *IEEE Signal Processing Letters* (Volume: 32), pages 1645-1649.

[4]. Hui-Peng Du, Yang Ai, **Rui-Chen Zheng**, Zhen-Hua Ling, "APCodec+: A Spectrum-Coding-Based High-Fidelity and High-Compression-Rate Neural Audio Codec with Staged Training Paradigm", in *Proc. ISCSLP 2024*, pages 676-680.

[5].Xiao-Hang Jiang, Yang Ai, **Rui-Chen Zheng**, Hui-Peng Du, Zhen-Hua Ling, "MDCTCodec: A Lightweight MDCT-based Neural Audio Codec towards High Sampling Rate and Low Bitrate Scenarios", in *Proc. SLT 2024*, pages 540-547.

[6].Fei-Liu, Yang Ai, Hui-Peng Du, Ye-Xin Lu, **Rui-Chen Zheng**, Zhen-Hua Ling, "Stage-Wise and Prior-Aware Neural Speech Phase Prediction", in *Proc. SLT 2024*, pages 638-644.

[7].Yang Ai, Ye-Xin Lu, Xiao-Hang Jiang, Zheng-Yan Sheng, **Rui-Chen Zheng**, Zhen-Hua Ling, "A Low-Bitrate Neural Audio Codec Framework with Bandwidth Reduction and Recovery for High-Sampling-Rate Waveforms", in *Proc. Interspeech 2024*, pages 1765-1769.

荣誉奖项

➢ 江淮未来汽车奖学金 *2024.12*

➢ 中国科学技术大学博士一等、硕士一等(两次)、二等学业奖学金 *2024.09 & 2023.09 & 2022.09 & 2021.09*

➢ 中国科学技术大学本科毕业生荣誉等级(Honor Rank，前5%) *2021.06*

➢ 华为奖学金 *2020.12*

➢ 中国科学技术大学优秀学生奖学金金奖(两次) *2019.12 & 2018.12*

➢ 拔尖计划奖学金(两次) *2019.12 & 2018.12*

技能

➢ **英语能力**

· TOEFL iBT: 106 (Reading: 28, Listening: 29, Speaking: 26, Writing: 23) *2023.11*

其他经历

➢ **审稿**

■ IEEE/ACM Transactions on Audio, Speech, and Language Processing, IEEE Signal Processing Letters, Speech Communication 受邀审稿人。

➢ **助教经历**

■ 语音信号处理基础(中国科学技术大学，凌震华教授) *2022 & 2021秋季学期*

■ 计算机程序语言设计A(中国科学技术大学，司虎讲师) *2020 秋季学期*