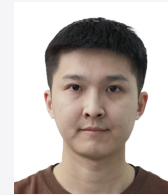


王天锐

18381077123 | wangtianrui@tju.edu.cn

<https://scholar.google.com/citations?hl=en&user=eGqo7CYAAAAJ>

28岁 | 男 | 172cm | 58kg | 四川成都 | 汉族 | 共青团员



教育背景

天津大学	2023.09 - 至今
电子信息 博士 智能与计算学部 导师: 王龙标&党建武	天津
研究方向为 语音合成、语音理解生成统一建模 中国科协青年科技人才培养工程博士生专项计划	一等学业奖学金
南洋理工大学	2025.02 - 至今
计算机科学 博士 计算与数据科学学院 联培导师: Eng Siong Chng	新加坡
研究方向为 语音合成中细粒度情感控制、基于LLM的情感可控语音合成	国家公派留学奖学金
北京交通大学	2020.09 - 2023.07
信息与通信工程 硕士 计算机与信息技术学院 导师: 朱维彬	北京
研究方向为语音增强、语音识别、自监督学习、深度学习、听觉建模	一等学业奖学金 校级优秀毕业论文
中北大学	2016.09 - 2020.07
物联网工程 学士 大数据学院	太原
绩点专业排名 1/137 AI+移动互联校级创新实验室负责人(110+人规模)	多项计算机相关赛事国家一等奖

科研交流与实习经历

上海交通大学	2022.05 - 至今
联培学习 跨媒体语言智能实验室 联培导师: 陈谐	
探索语音自监督预训练、语音合成、鲁棒语音识别、音频理解生成统一相关模型	
腾讯集团	2026.04 - 至今
实习生 语音大模型算法中心	
音频理解生成统一相关模型研究	
华为技术有限公司	2026.01 - 至今
校企合作项目学生负责人	
细粒度情感可控端到端对话模型研究	
微软亚洲研究院	2022.12 - 2023.07
语音算法实习生 自然语言计算组	
探索理解生成统一的语音大模型(语音识别、机器翻译、语音到文本的翻译和语音合成)	
中国移动研究院	2021.05 - 2022.11
语音算法实习生 人工智能与智慧运营部	
主要负责流式语音增强和鲁棒语音识别系统的研发,并成功评选“超星计划”(实习生人才计划)	
慧言科技有限公司	2023.10 - 2024.02
语音合成算法实习生	
主导中英双语语音合成系统的研发:数据爬虫、海量数据预处理(15W小时数据)、多语言语音合成系统的搭建和训练	

学术科研情况（仅列出一作论文）

Evaluating the Expressive Appropriateness of Speech in Rich Contexts

第一作者（Annual Meeting of the Association for Computational Linguistics, ACL 2026收录, Oral, Top5%）

- 首次提出了在丰富上下文设定中评估语音表现力的恰当性模型
- 评估模型从情感、韵律、音调、音效、节奏、非语言表达等多个维度评估目标语音在指定上下文中的表现力
- 公开了14小时人工精细化标注的具有丰富上下文的表现力真实语音数据以及每句话的表现力恰当性评分
- 公开了2500小时的具有丰富上下文的表现力真实语音数据以及qwen3-omni-captioner的弱标注

Word-Level Emotional Expression Control in Zero-Shot Text-to-Speech Synthesis

第一作者（Annual Conference on Neural Information Processing Systems, NeurIPS2025 Spotlight 收录）

- 提出了一种创新的多轮推理与自训练框架，在无需情感标注数据的情况下，实现了TTS中情感与语速的词级控制
- 通过引入动态情感注意力偏置与过渡平滑机制，生成自然且富有表现力的语音
- 该方法在保持原始zero-shot能力的同时，在词级情感表达上达到当前最优性能

Progressive Residual Extraction based Pre-training for Speech Representation Learning

第一作者（IEEE Transactions on Audio, Speech, and Language Processing收录，语音领域顶刊）

- 将语音解耦的思想引入自监督预训练，提出渐进式解耦自监督语音预训练策略
- 所提方法能在不引入标签的基础上，逐步地从语音中提取出音高、说话人和内容信息
- 同时地提升了预训练模型在多种任务上的性能，并超越WavLM

VIOLA: Conditional Language Models for Speech Recognition, Synthesis, and Translation

第一作者（IEEE Transactions on Audio, Speech, and Language Processing收录，语音领域顶刊）

- 基于Decoder-only的框架，提出跨语言和跨模态的语音文本相互转换模型
- 模型能兼顾语音识别、机器翻译、语音合成、语音到文本翻译、语音到语音翻译五种任务
- 兼顾五种任务的同时，在每个子任务上都实现了较好的性能表现

Harmonic Attention for Monaural Speech Enhancement

第一作者（IEEE Transactions on Audio, Speech, and Language Processing收录，语音领域顶刊）

- 提出Harmonic Attention机制，借助谐波的梳状结构和高能量特性对谐波进行建模和补偿
- 使用渐进式增强策略，在处理过程中尽可能地保留声学结构，提出基于谐波注意力的渐进式语音增强网络(HAPNet)
- 提出基于响度压缩的听觉损失函数，在提升增强结果与人耳听觉一致性的同时减少听觉效应带来的失真

A Progressive Generation Framework with Speech Pre-trained Model for Expressive Voice Conversion

第一作者（ICME2025 收录，CCFB会议）

- 提出了一种三阶段表现型声音转换框架，利用预训练语音模型的特征解耦能力，逐步整合文本、情感、说话人和声学信息以生成高质量语音
- 相较于非渐进方法，渐进式的框架更有效地利用了解耦信息间的关系，实现更自然的情感与说话人转换
- 构建了一个可用于评估预训练模型语音解耦能力的实用框架

Characteristic-Specific Partial Fine-Tuning for Efficient Emotion and Speaker Adaptation in Codec Language Text-to-Speech Models

第一作者（投稿至Speech Communication，CCFB类期刊）

- 提出了特征导向的部分微调方法，用于高效适配通用 Codec TTS 模型，实现特定情感与说话人特征的语音合成
- 通过分析 Transformer 各层对情感与说话人控制的贡献，显著提升了微调效果并减少知识遗忘风险
- 方法在多个开源模型上验证了高效性与适应性，并进一步探索了将 TTS 模型用作语音情感识别与说话人识别编码器的潜力

An Adapter Based Multi-label Pre-training for Speech Separation and Enhancement

第一作者 (ICASSP2023收录, 语音领域顶会)

- 提出基于适配器的多标签预训练策略
- 将增强和分离思想引入到语音预训练任务, 来缓解预训练任务与分离任务之间的数据域不匹配和训练任务不匹配问题
- 引入基于适配器的预训练策略, 在模型学习新任务的同时保留住原始任务的性能 (如ASR)

HGCN: Harmonic Gated Compensation Network For Speech Enhancement

第一作者 (ICASSP2022 收录, 语音领域顶会)

- 提出高分辨率谐波积分矩阵, 能在有限频域分辨率下较为准确地预测谐波的位置
- 利用深度学习模型在谐波位置的指导下对谐波进行幅度补偿
- 所提方案具有良好实时性的同时显著提升了处理效果

Harmonic Gated Compensation Network Plus For ICASSP 2022 DNS CHALLENGE

第一作者 (ICASSP2022 收录, 语音领域顶会)

- 基于HGCN做了全带信号处理、谐波积分谱优化、粗处理和补偿模型的改进等尝试, 提出了HGCN+全带信号处理模型
- HGCN+能兼顾全带和宽带信号的处理 (不对信号resample的前提下)
- 在公开的宽带和全带测试集上, HGCN+处理结果优于参考模型, 并在ICASSP 2022 DNS Challenge上取得第 5 名的成绩

基于梅尔倒谱系数、深层卷积和Bagging的环境音分类方法

第一作者 (计算机应用 正刊收录) (本科时发表)

- 基于Bagging集成多个弱分类模型, 以提升模型在小规模数据集上的泛化能力, 在小规模音频分类任务上实现SOTA效果

个性化发音风格描述系统

第一作者 (本科毕业设计)

- 从生理特性、发音调控以及宏观评价等角度对人的发音风格进行分析, 提出了具有层次结构的个性化发音风格描述系统
- 通过知识总结以及问卷调查, 从生理特性、发音调控以及宏观评价的角度筛选出了31个描述维度

轻耳-基于机器学习的智能手机音量调节软件

项目负责人 (省级大学生创新创业计划项目) (本科时期)

- 负责环境音分类与检测模型搭建、根据环境类型与环境嘈杂程度以及用户运动状态对播放音量进行回归预测

专业竞赛

ICASSP 2025 LIMMITS'25 Challenge 低资源语音合成比赛第2名 (双赛道均第2)	2024.11
ISCSLP 2024 ICAGC Challenge 中文情感可控语音合成比赛第2名	2024.06
Inter Neuromorphic DNS Challenge 第2名	2023.11
ICASSP 2022 Deep Noise Suppression Challenge 第5名	2022.02
入选中国移动研究院“超星计划”(设置以来首例)	2021.09
Kaggle房价预测赛全球前10% (此条及以下为本科期间完成)	2019.03
华北五省计算机应用大赛国家一等奖 (作品名: 营养大师-基于机器学习的膳食管家)	2018.11
“互联网+”大学生创新创业大赛国家铜奖 (作品名: 吾乡-红色智慧旅游项目)	2018.09
全国计算机设计大赛国家一等奖 (作品名: 识菜帮-基于深度学习的厨房助手)	2018.08
“创青春”浙大双创杯全国大学生创业大赛省级金奖 (作品名: 智电科技)	2018.05
全国大学生Google安卓挑战赛优秀奖	2017.04

其他

- 技能: Latex、Python、Pytorch、Tensorflow、Java、Android
- 语言: CET-6