基于隐空间拼接与对比学习的语音匿名方法研究

张旭龙[1]，瞿晓阳[1]，倪晓俊[2]，田晖[3]，王健宗[1]
1. 平安科技（深圳）有限公司，广东 深圳 518000；
2. 中国科学院计算技术研究所，北京 100190；
3. 华侨大学计算机科学与技术学院，福建 泉州 362021

摘要

随着语音技术的广泛应用，语音隐私保护面临严峻挑战。基于此，提出一种基于隐空间拼接与对比学习的语音匿名方法kLCAnoy++，旨在有效去除说话人身份信息，同时保持语音内容可懂度与自然度。该方法首先利用预训练的WavLM模型提取自监督语音特征；其次，通过k近邻算法在多样化参考音频特征库中匹配目标特征，实现隐空间特征拼接，生成初步匿名语音；最后，引入对比学习机制，以自然拼接为正样本、k近邻拼接为负样本，优化声码器解码过程，提升匿名语音的连贯性与自然度。在VCTK数据集上的实验表明，所提方法显著优于基线方法，可视化分析进一步验证了该方法能有效解耦说话人身份信息。

关键词

语音匿名化；说话人去标识；自监督语音表示；隐空间特征拼接；对比学习

中图分类号：TN912.3；TP391.4　　文献标志码：A　　doi：10.11959/j.issn.2096-6652.2026059

Research on speaker anonymization method via latent space splicing and contrastive learning

Zhang Xulong[1], Qu Xiaoyang[1], Ni Xiaojun[2], Tian Hui[3], Wang Jianzong[1]
1. Ping An Technology (Shenzhen) Co. , Ltd. , Shenzhen 518000, China
2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
3. College of Computer Science and Technology, Huaqiao University, Quanzhou 362021, China

Abstract

With the widespread application of speech technology, voice privacy protection faces severe challenges. A speech anonymization method (kLCAnoy++) based on latent space splicing and contrastive learning was proposed, aiming to effectively remove speaker identity information while preserving speech content intelligibility and naturalness. The method first employed the pretrained WavLM model to extract self-supervised speech features. Subsequently, it matched target features in a diverse reference audio feature library using the *k*-nearest neighbors algorithm, achieving latent space feature splicing to generate preliminary anonymized speech. Finally, a contrastive learning mechanism was introduced, treating natural splicing as positive samples and *k*-nearest neighbor splicing as negative samples, to optimize the vocoder decoding process and enhance the coherence and naturalness of anonymized speech. Experiments

on the VCTK dataset demonstrated that the proposed method significantly outperformed baseline approaches, with visualization analysis further confirming its effectiveness in disentangling speaker identity information.

Key words
speech anonymization, speaker de-identification, self-supervised speech representations, latent space splicing, contrastive learning

0 引言

随着语音技术在日常生活中的广泛应用，如语音助手、身份认证、语音通信等，大量语音数据被生成和存储。然而，语音数据中蕴含着丰富的个人隐私信息，包括年龄、性别、民族、宗教信仰等。这些隐私信息一旦泄露，可能会给个人带来如身份被盗、被监视等安全隐患[1-2]。因此，语音匿名化技术[3]应运而生，其主要任务是在保护语音内容的同时，去除语音中的说话人身份信息，以实现语音数据的隐私保护。语音匿名化技术不仅能够保障个人隐私，还能使语音数据在多种场景下得到安全应用，如提升客服中心和语音助手功能、协助执法机构保护证人隐私等[4]。

早期的语音匿名化方法主要基于信号处理技术[5]，通过直接修改语音信号的特性来实现匿名化，如对语音信号进行时域尺度变换、音高变换或频谱包络变换[6-7]。这些方法简单易行，不需要训练数据，能快速对语音信号进行处理。然而，语音信号的物理变换范围有限，攻击者可能通过多次尝试恢复原始语音，导致匿名化效果不佳，且这些方法可能导致语音质量下降，影响语音的可懂度和自然度。

近年来，语音转换（voice conversion，VC）[8-11]技术逐渐成为语音匿名化的重要手段。这类方法通过将说话人的语音转换为另一个人的语音来实现匿名化。传统的语音转换方法通常依赖于对语音信号的声学特征建模，如频谱特征、基频特征等，再通过映射函数将源说话人的特征转换为目标说话人的特征。例如，基于高斯混合模型（Gaussian mixture model，GMM）的语音转换方法[12]，通过学习源说话人和目标说话人语音特征的概率分布，并建立两者之间的映射关系来实现语音转换。然而，传统方法在语音质量、说话人相似度和语音自然度等方面存在一定的局限性，而且需要大量的平行语音数据进行训练，这在实际应用场景中难以满足。

随着深度学习的发展，基于神经网络的方法在语音匿名化领域取得了显著进展。这类方法通常通过学习语音信号的深层特征表示，将语音内容和说话人身份特征进行解耦[13]。例如，利用自编码器结构，将语音信号编码为低维特征向量，然后通过解码器重建语音信号，同时在编码过程中引入对抗训练或正则化项，以抑制说话人身份信息的编码保留[14]。此外，预训练的语音表示模型（如HuBERT[15]、Wav2Vec 2.0[16]、WavLM[17]等）也为语音匿名化提供了强大的特征提取能力。这些预训练模型能够在大规模无监督语音数据上学习到通用的语音特征表示，通过微调这些模型或基于其特征进行进一步的处理，可以实现更有效的语音匿名化。然而，这些方法仍然存在一些挑战：如何在去除说话人信息的同时保持语音内容的完整性、如何提高匿名化语音的自然度和可懂度，以及如何应对更强大的攻击场景等。

为了解决现有语音匿名化方法在去除

说话人信息和保持语音质量之间的平衡问题，本文提出了一种基于隐空间拼接与对比学习的语音匿名方法kLCAnoy。在隐空间拼接方面，该方法借鉴了ContentVec[18]的思想，对预训练的语音表示进行处理，以获得与说话人无关的语音内容表示。具体来说，本文使用WavLM模型[19]学习通用语音表示，并通过设计k近邻（k-nearest neighbor，kNN）目标语音表示拼接机制将目标匿名语音内容特征进行组合，从而得到匿名化语音的隐变量表示。为了保障声码器从隐变量直接解码生成语音的效果，本文进一步提出了kLCAnoy++，引入对比学习使自然拼接作为正样本，kNN匹配拼接作为负样本，使得声码器的解码能够从隐特征空间中学习到连贯的自然拼接，保障拼接特征合成的语音连贯、自然。通过这种方式，所提方法在保持语音内容可懂度的同时，有效地去除说话人身份信息，提高了语音匿名化的性能。

1 方法

基于隐空间拼接与对比学习的语音匿名方法，旨在有效去除语音中的说话人身份特征，同时保留语音的可懂度和自然度。基于隐空间拼接解码优化的语音匿名框架如**图1**所示，其包含三大模块，即自监督特征提取模块、特征匹配与拼接模块以及对比学习声码器优化模块。

1.1 自监督特征提取

自监督学习（self-supervised learning，SSL）模型因其在语音处理任务中的强大特征提取能力而被广泛采用。本文选用预训练的WavLM模型作为特征提取器。WavLM模型通过联合学习掩蔽语音预测和去噪任务，能够处理包含说话人验证、语音分离和语音识别在内的全栈下游语音任务，其特定层的特征能够将相同音素的帧在特征空间中拉近，同时保留说话人的音色信息，这对于语音转换任务具有重要价值。在提取过程中，原始语音信号输入WavLM模型，获取固定维度的SSL特征。这些特征包含了语言内容和说话人音色的综合信息，为后续的特征拼接和转换提供了基础。

为了实现语音匿名化，需要构建一个参考音频特征库。首先，收集大量具有多

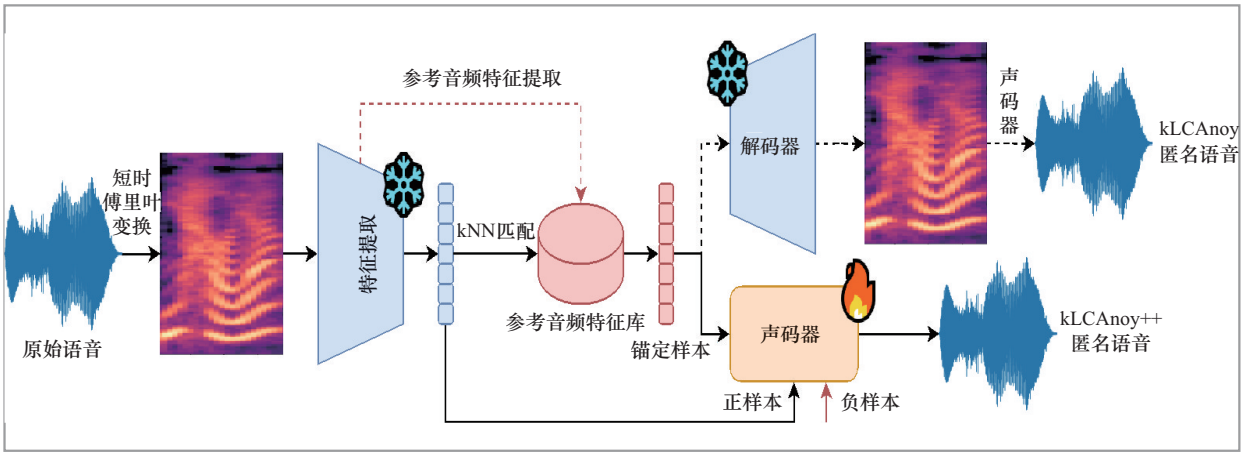

图1 基于隐空间特征拼接解码优化的语音匿名框架

样化说话人特征的参考音频样本，这些样本要确保能够涵盖不同性别、年龄以及地域口音等各个维度，从而为后续可选择的音色特征提供丰富的来源；在收集到足够的样本后，对每个参考音频样本运用与原始语音相同的 WavLM 模型来提取 SSL 特征，以此得到参考音频特征集；随后，对该参考音频特征集进行预处理以及归一化操作，目的是使其与原始语音特征具备可比性和兼容性。经过以上过程，构建完成的参考音频特征库便能为后续的特征匹配与拼接环节提供丰富的候选特征资源。

1.2 特征匹配与拼接

特征匹配与拼接是隐空间拼接的核心操作。对于原始语音的每一帧特征，均运用 kNN 算法[20]在参考音频特征库中搜寻与之最为相近的特征。kNN 算法凭借其简洁且高效的特性，在特征匹配任务中能够迅速精准地锁定与原始特征相似度较高的参考特征。在匹配环节，为了提升匹配的准确性以及鲁棒性，综合考量 WavLM 模型最后几层的特征用于匹配操作，而针对后续合成步骤，则选取特定层的特征。采用这一策略的原因在于，模型最后几层所蕴含的特征具备更强的区分度，所包含的内容信息更为丰富，从而有助于实现对语义相关特征更为精准的匹配。一旦确定了原始特征在参考特征库中的近邻集合，便将原始特征替换为这些近邻特征的加权平均值，由此达成特征层面的拼接效果。加权平均的权重依据距离进行确定，其目的在于确保拼接后的特征在最大限度保留原始语音内容的基础上，成功融入新的说话人音色特征。采用这种特征匹配与拼接方式，本文将原始语音的特征转化为融合了参考说话人音色特征的全新特征。

完成特征拼接操作后，所得到的初步匿名语音特征在连贯性以及与原始语音内容的匹配度等方面仍存在不足。基于此，必须对拼接后的特征开展融合与优化处理。在特征融合阶段，通过非线性变换，将拼接后的特征与原始语音特征进行融合，在保留大部分参考音色特征的同时，进一步强化匿名语音与原始语音在内容表达方面的一致性，有效规避因特征替换而可能引发的语音语义偏差问题。在特征优化阶段，将对比学习的损失函数作为优化目标，促使匿名语音特征在满足隐私保护要求的前提下，尽可能提升语音的自然度、可懂度等质量指标。

1.3 对比学习解码优化

对比学习的核心在于通过对样本之间相似性和差异性的学习，生成更具区分性的特征表示。在本文所提的方法中，对比学习模块的关键目标是让模型能够精准区分不同说话人的语音特征，优化拼接特征的合成效果。拼接后的特征通过投影头将原始特征映射到一个更适于对比学习的对比空间，这一映射过程采用全连接层来实现，通过投影头的转换，在对比空间中衡量样本之间的相似性和差异性。对比损失函数衡量正样本对和负样本对之间的相似性差异，并指导模型的训练过程，表示为：

$$\mathcal{L}=-\ln\frac{\exp\left(\mathrm{sim}\left(a,p\right)/\tau\right)}{\exp\left(\mathrm{sim}\left(a,p\right)/\tau\right)+\sum_{i}\exp\left(\mathrm{sim}\left(a,n_i\right)/\tau\right)} \tag{1}$$

其中，a 为锚样本，p 为正样本，n_i 为负样本（i 为负样本的索引），τ 为温度超参数，用于缩放样本间的相似度，调节模型对困难样本的区分度，$\mathrm{sim}\left(a,p\right)$ 和 $\mathrm{sim}\left(a,n_i\right)$ 为相似性函数，用于计算锚样本与正样本、

锚样本与负样本之间的相似性。相似性函数可以是点积、余弦相似性等。该方法采用InfoNCE损失函数，在给定锚样本的情况下，计算其与正样本的相似性以及与多个负样本的相似性，再通过softmax函数将这些相似性值转换为概率分布，使正样本对的概率最大化，同时将负样本对的概率最小化，从而优化拼接特征，以提升其解码为合成语音的效果。

2 实验

为了验证所提出的基于隐空间拼接与对比学习的语音匿名方法的有效性和优越性，本文设计了一系列实验来评估该方法在去除说话人身份信息方面的性能，同时考察其对语音质量和可懂度的影响。通过与现有先进方法的比较，本文验证了所提方法在语音匿名化任务中的优势。

2.1 数据集

实验采用了公共语音数据集VCTK[21]，该数据集包含110名母语为英语的说话人朗读的语音，每人提供大约400个句子。模型使用VCTK数据集中100个说话人的语音进行训练，使用VCTK剩余10位说话人进行评估。VCTK数据集中用于测试的说话人信息见**表1**，其中包含了5位男性和5位女性说话人。

2.2 基线方法

为了对所提方法进行全面比较，实验考虑了几种典型的语音匿名化基线方法。其中，基于McAdams系数的方法[22]通过改变语音信号短时功率谱的极点角度实现说话人匿名化；基于Praat工具包的方法[23]则通过随机改变语音的基频和共振峰实现匿名化。此外，MAIN-VC方法[24]是一种基于自编码器的模型，它采用激活瓶颈对内容嵌入进行约束，以此改善转换语音的合成质量与说话人相似度之间的平衡。StableVC方法[25]通过将语音分解为语言内容、音色和风格这3种特征，利用条件流匹配模块基于这些特征重建高质量的梅尔频谱，将源说话者的音色和风格独立地、

表1 VCTK数据集中用于测试的说话人信息

说话人	年龄	性别	口音	地区
p225	23	女	英语	南英格兰(Southern)
p226	22	男	英语	萨里(Surrey)
p227	38	男	英语	坎布里亚(Cumbria)
p228	22	女	英语	南英格兰(Southern)
p229	23	女	英语	南英格兰(Southern)
p230	22	女	英语	蒂斯河畔斯托克顿(Stockton-on-tees)
p231	23	女	英语	南英格兰(Southern)
p232	23	男	英语	南英格兰(Southern)
p237	22	男	苏格兰语	法夫(Fife)
p241	21	男	苏格兰语	珀斯(Perth)

可控地转换到不同的未见过的目标说话者。本文提出的方法（使用kNN与WavLM-SSL进行直接匹配匿名）在实验中记为kLCAnoy，在此基础上增加对比学习以优化解码过程的方法记为kLCAnoy++。

2.3 实验指标

实验的主要评估指标包括说话人识别/验证的等错误率（equal error rate，EER）和语音识别的词错误率（word error rate，WER）。其中，EER用于衡量隐私保护效果，该值越高表示匿名化效果越好；WER则用于衡量语音的可懂度，该值越低表示语音失真越小，可懂度越高。

2.4 实验设置

在自监督特征提取阶段，实验对比使用预训练的WavLM、Wav2Vec 2.0以及HuBERT模型作为特征提取器。在构建参考音频特征库时，收集具有多样化说话人特征的参考音频样本，利用预训练模型提取SSL特征，并进行预处理和归一化。在特征匹配与拼接环节，采用kNN算法在参考音频特征库中搜索与原始语音特征相近的特征，完成特征拼接。在对比学习解码优化阶段，引入余弦相似性损失函数，对拼接后的特征进行优化，使原始语音及其变换后的版本在特征空间中更加相似，从而减少说话人相关信息的差异。

表2 不同预训练模型及特征表示的匿名化性能

模型	EER↑	WER↓
原始语音	1.20%	5.29%
HuBERT-SSL	44.30%	9.47%
Wav2Vec 2.0-SSL	44.60%	9.39%
WavLM-SSL	44.80%	7.50%

表3 不同语音转换方法在VCTK测试集上的性能

方法	EER↑	WER↓
原始语音	1.20%	5.29%
McAdams	5.60%	11.74%
基于Praat的基频/共振峰调整	35.20%	10.27%
MAIN-VC	44.40%	44.37%
StableVC	46.10%	12.93%
WavLM-SSL(消融)	44.80%	7.50%
kLCAnoy(消融)	44.90%	6.24%
kLCAnoy++	49.40%	6.56%

2.5 实验结果

为对比预训练模型的特征表示效果，实验分别采用HuBERT、Wav2Vec 2.0以及WavLM作为特征提取器，不同预训练模型及特征表示的匿名化性能见**表2**。从**表2**可以看出，使用不同预训练模型对语音匿名化性能有显著影响。其中，WavLM-SSL相较于HuBERT-SSL和Wav2Vec 2.0-SSL模型，在EER上的表现最佳达到44.80%，表明其在去除说话人身份信息方面的效果更佳，同时WER最低为7.50%，说明语音可懂度更高。

在与基于语音转换技术的说话人匿名方法以及模块消融对比实验中，不同语音转换方法在VCTK测试集上的性能见**表3**。由**表3**可知，原始语音的EER仅为1.20%，而本文提出的kLCAnoy++方法的EER为49.40%，表明其在去除说话人信息方面表现优异。在WER方面，kLCAnoy++方法的WER为6.56%，与原始语音的5.29%较为接近，而且相较于MAIN-VC（WER为44.37%）、StableVC（WER为12.93%）等其他基线方法，失真更小，语音可懂度更高。

为验证语音匿名后的说话人与目标说话人的音色聚类情况，本文对4个目标说话人经匿名转换后的语音进行基于X-vector（extended variable-length vector，延时嵌入提取的说话人特征向量）的t-SNE（t-distributed stochastic neighbor embedding，t分布随机邻域嵌入）聚类可视化分析，结果如**图2**所示。从**图2**可以看出，经过kLCAnoy方法匿名化后的说话人语音的特征向量聚类紧密，且与其他说话人区分明显，与原始说话人的关联性降低，这表明该方法有效去除了说话人的身份信息。

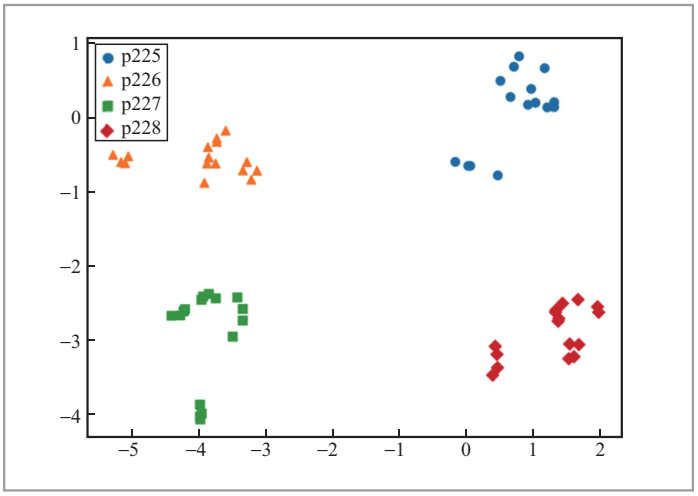

图2 经匿名转换后的语音进行基于X-vector的t-SNE聚类可视化分析结果

本文对目标说话人匿名化过程中，使用自身语料与跨说话人语料生成的语音转换结果进行频谱可视化对比。说话人匿名语音频谱可视化对比结果如**图3**所示。从**图3**中红色框标记位置频谱的对比可以看出，**图3（d）**中p228经过匿名转换后的频谱趋势与**图3（a）**和**图3（c）**更为接近，而与**图3（b）**差异变大。同时，该频谱与p225原始语音的频谱相似度更高。上述结果进一步证明所提方法在保持语音质量和可懂度的同时，能够有效去除说话人身份信息，提升了语音匿名化效果。

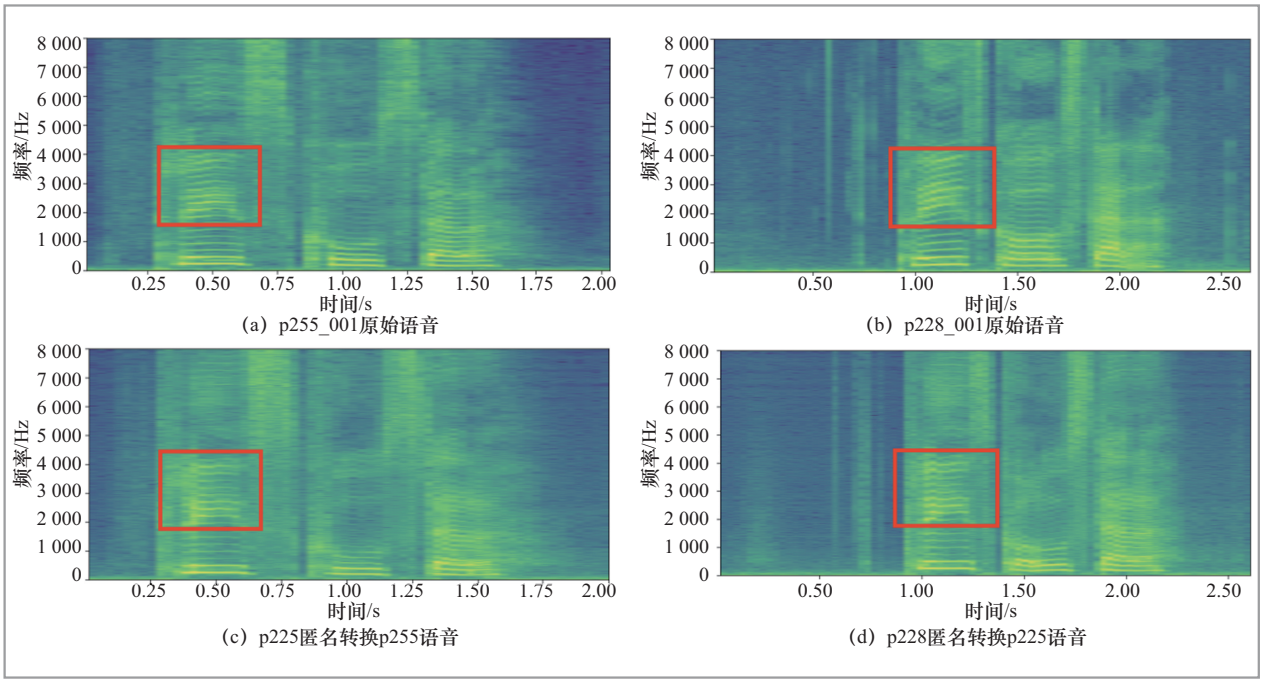

图3 说话人匿名语音频谱可视化对比结果

3 结束语

本文提出了基于隐空间拼接与对比学习的语音匿名方法kLCAnoy++。该方法利用预训练WavLM特征结合kNN拼接有效解耦了说话人音色，并通过对比学习优化解码，显著提升了合成语音的连贯性。实验验证了其在VCTK数据集上达到49.40%的高EER和6.56%的低WER，在掩盖身份与保持语音自然度方面均优于现有基线方法。然而，本研究仍存在一定局限性：一是特征拼接质量依赖参考库的规模与多样性；二是当前评估仅基于录音室数据集，在复杂真实噪声下的鲁棒性有待验证；三是逐帧kNN匹配存在计算开销，限制了其在极低时延场景的应用。未来工作将重点构建多语种及复杂噪声评估基准以提升模型泛化能力；探索高效的近似检索算法以降低特征匹配时延，推动实时流式应用；引入严苛的对抗性攻击评估框架，进一步巩固隐私安全防线。

参考文献：

[1] Erg ü nay S K, Khoury E, Lazaridis A, et al. On the vulnerability of speaker verification to realistic voice spoofing[C]//Proceedings of the 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS). Piscataway: IEEE Press, 2015: 1-6.

[2] Vestman V, Kinnunen T, Gonz á lez Hautamäki R, et al. Voice mimicry attacks assisted by automatic speaker verification[J]. Computer Speech & Language, 2020, 59: 36-54.

[3] Tayebi Arasteh S, Arias-Vergara T, P é rez-Toro P A, et al. Addressing challenges in speaker anonymization to maintain utility while ensuring privacy of pathological speech[J]. Communications Medicine, 2024, 4: 182.

[4] Miao X X, McLoughlin I, Wang W C, et al. D-MONA: a dilated mixed-order non-local attention network for speaker and language recognition[J]. Neural Networks, 2021, 139: 201-211.

[5] Hashimoto K, Yamagishi J, Echizen I. Privacy-preserving sound to degrade automatic speaker verification performance[C]//Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2016: 5500-5504.

[6] Qian J W, Du H H, Hou J H, et al. VoiceMask: anonymize and sanitize voice input on mobile devices[PP]. arXiv preprint, 2017, arXiv: 1711.11460.

[7] Mawalim C O, Galajit K, Karnjana J, et al. Speaker anonymization by modifying fundamental frequency and x-vector singular value[J]. Computer Speech & Language, 2022, 73: 101326.

[8] Magariños C, Lopez-Otero P, Docio-Fernandez L, et al. Reversible speaker de-identification using pre-trained transformation functions[J]. Computer Speech & Language, 2017, 46: 36-52.

[9] Jin Q, Toth A R, Schultz T, et al. Speaker de-identification via voice transformation[C]//Proceedings of the 2009 IEEE Workshop on Automatic Speech Recognition & Understanding. Piscataway: IEEE Press, 2009: 529-533.

[10] Qian J W, Du H H, Hou J H, et al. Hidebehind: enjoy voice input with voiceprint unclonability and anonymity

[C]//Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems. New York: ACM, 2018: 82-94.

[11] Huang C Y, Lin Y Y, Lee H Y, et al. Defending your voice: adversarial attack on voice conversion[C]//Proceedings of the 2021 IEEE Spoken Language Technology Workshop (SLT). Piscataway: IEEE Press, 2021: 552-559.

[12] Chen Y N, Chu M, Chang E, et al. Voice conversion with smoothed GMM and MAP adaptation[C]//Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech 2003). ISCA, 2003: 2413-2416.

[13] Pierre C, Larcher A, Jouvet D. Are disentangled representations all you need to build speaker anonymization systems? [C]//Proceedings of the Interspeech 2022. ISCA, 2022: 2793-2797.

[14] Champion P. Anonymizing speech: evaluating and designing speaker anonymization techniques[PP]. arXiv preprint, 2023, arXiv: 2308.04455.

[15] Van Niekerk B, Carbonneau M A, Zaïdi J, et al. A comparison of discrete and soft speech units for improved voice conversion[C]//Proceedings of the ICASSP 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2022: 6562-6566.

[16] Matassoni M, Fong S, Brutti A. Speaker anonymization: disentangling speaker features from pre-trained speech embeddings for voice conversion[J]. Applied Sciences, 2024, 14(9): 3876.

[17] Tan T, Liu S, Duan Y, et al. System description: speaker anonymization system with sentiment transfer and feature interpolation[PP]. arXiv preprint, 2024, arxiv:2409.08913.

[18] Qian K, Zhang Y, Gao H, et al. Contentvec: an improved self-supervised speech representation by disentangling speakers[C]//Proceedings of the International Conference on Machine Learning (ICML). PMLR, 2022: 18003-18017.

[19] Chen S Y, Wang C Y, Chen Z Y, et al. WavLM: large-scale self-supervised pre-training for full stack speech processing[J]. IEEE Journal of Selected Topics in Signal Processing, 2022, 16(6): 1505-1518.

[20] Baas M, van Niekerk B, Kamper H. Voice conversion with just nearest neighbors[PP]. arXiv preprint, 2023, arXiv: 2305.18975.

[21] Yamagishi J, Veaux C, MacDonald K, et al. CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit[R]. 2019-11-13.

[22] Sinha Y, Siegert I. Performance and quality evaluation of a McAdams speaker anonymization for spontaneous German speech[C]//Fortschritte der Akustik-DAGA 2022. Berlin: Deutsche Gesellschaft für Akustik e. V., 2022: 1185-1188.

[23] Singh D K, Prajapati G P, Patil H A. Voice privacy using time scale and pitch modification[J]. SN Computer Science, 2024, 5(2): 243.

[24] Li P C, Wang J Z, Zhang X L, et al. MAIN-VC: lightweight speech representation disentanglement for one-shot voice conversion[C]//Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2024: 1-7.

[25] Yao J X, Yang Y G, Pan Y, et al. StableVC: style controllable zero-shot voice conversion with conditional flow matching[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(24): 25669-25677.

作者简介

张旭龙（1988– ），男，博士，平安科技（深圳）有限公司算法研究员，2023年入选上海市东方英才计划青年项目，兼任清华大学深圳国际研究生院及中国科学技术大学先进技术研究院校外导师，以及中国指挥与控制学会具身智能专业委员会委员、中国自动化学会联邦数据与联邦智能委员会委员、中国电子音响行业协会声音与音乐技术专委会常务委员，目前是IEEE、中国自动化学会以及中国计算机学会会员，主要研究方向为大模型、具身智能、跨模态智能计算等。发表学术论文80余篇，获得国家发明专利授权20余项。

瞿晓阳（1988– ），男，博士，平安科技（深圳）有限公司前沿机器学习算法分组负责人、清华大学深圳国际研究生院校外导师、中国科学技术大学校先进技术研究院校外导师、中佛罗里达大学访问学者，主要研究方向为机器学习、大数据、体系结构，以及语音语义分析、自动化机器学习、零样本和小样本学习、高性能计算与存储等。在体系结构方向（如INFOCOM、DAC、IPDPS、TPDS）和人工智能方向（如NeurIPS、IJCAI、ICASSP、Interspeech）的国际顶级会议和顶级期刊发表过近50篇文章，1篇论文荣获会议最佳学生论文奖提名，担任过多个国际顶级期刊的评委，已授权专利70篇，已出版专著两本。

倪晓俊（1973– ），男，中国科学院计算技术研究所助理研究员，主要研究方向为数据库安全、大数据分析、隐私计算。

田晖（1982– ），男，博士，博士生导师，华侨大学计算机科学与技术学院院长、教授，网络与信息安全产业学院院长，厦门市数据安全与区块链技术重点实验室主任，入选福建省青年拔尖创新人才、福建省百千万人才工程省级人选、福建省高等学校新世纪优秀人才、福建省高校杰出青年科研人才培育计划，现为IEEE高级会员、中国计算机学会（CCF）杰出会员、CCF互联网专委会执行委员、CCF区块链技术专委会执行委员、中国中文信息学会大数据安全与隐私保护专业委员会常务委员、福建省计算机学会副理事长、福建省网络空间安全学会理事长，主要研究方向为网络与信息安全、人工智能安全、数据安全、多媒体内容安全等。近年来，主持各类科技项目20余项，其中包括国家自然科学基金项目4项、国家重点研发课题和子课题各1项；在IEEE TDSC、IEEE TIFS、IEEE TASLP、ACM MM等国内外知名期刊和学术会议上发表论文120余篇；获得授权发明专利40余项，软件著作权8项；部分研究成果曾荣获中国电子学会科学技术奖二等奖、海南省自然科学奖二等奖、福建省科学技术进步奖三等奖和福建省自然科学奖三等奖。

王健宗（1983– ），男，博士，正高级工程师，平安科技（深圳）有限公司智能金融前沿技术研究院院长、美国佛罗里达大学人工智能博士后、美国莱斯大学和华中科技大学联合培养博士、CCF杰出会员、CCF大数据专家委员会委员、中国自动化学会联邦数据和联邦智能专业委员会副主任，主要研究方向为具身智能、大模型、联邦学习和深度学习。荣获中国专利优秀奖、广东省科学技术奖二等奖、吴文俊人工智能科学技术奖一等奖、中国人民银行金融科技发展奖一等奖、深圳高层次人才领军人才、福田英才等。

录用日期: 2026-02-03

基金项目: 深港联合基金(A类)项目(No.SGDX20240115103359001)

Foundation Item: The Shenzhen-Hong Kong Joint Funding Project (Category A) (No.SGDX20240115103359001)

通信作者: 倪晓俊，jamesni@sina.com