基于深度学习的蒙古语合成语音检测研究

张锦华#, 1　　梁凯麟#, 1　　刘瑞*, 1

1.内蒙古大学计算机（软件）学院, 呼和浩特, 010000, 中国

摘　要：合成语音检测是用于识别和鉴别计算机生成的合成语音与真实人类语音之间的差异，以侦测可能的虚假信息或欺骗。这种技术通常用于声纹识别和反欺诈应用中，以提高安全性和身份验证的可靠性。近年来，针对英语、汉语等主流语言的合成语音检测工作发展迅速，但是针对蒙古语等小语种的合成语音检测工作还处于空白阶段。鉴于蒙古语语音合成近几年的发展成果，为了填补这一空白，我们基于实验室强大的蒙古语语音合成模型构建了蒙古语合成语音检测数据集，并在该数据集上对主流的合成语音检测框架进行试验比较，同时我们在https://github. com/ssmlkl/NCMMSC2023 开源了相关数据集和基线模型。本文是首次对蒙古语合成语音检测展开深入研究，旨在推动蒙古语的合成语音检测领域的发展，为小语种合成语音检测的研究做出一定贡献。

关键词：合成语音检测；蒙古语；深度学习

中图分类号：请查阅《中国图书馆分类法》　　**文献标识码：**　　**DOI:**10.16798/j.issn.1003-0530. ****.**.***

Research on Mongolian synthetic speech detection based on deep learning

Jinhua Zhang#,1　　Kailin Liang#,1　　Rui Liu*,1

1.School of Computer (Software), Inner Mongolia University, Hohhot, 010000, China

Abstract: Synthetic speech detection is used to identify and discriminate the differences between computer-generated synthetic speech and real human speech in order to detect possible false information or deception. This technology is commonly used in voiceprint recognition and anti-fraud applications to improve security and the reliability of authentication. In recent years, synthetic speech detection for English, Chinese and other mainstream languages has developed rapidly. However, the synthetic speech detection for minority languages such as Mongolian is still in the blank stage. In view of the development achievements of Mongolian speech synthesis in recent years, in order to fill this gap, we built Mongolian synthetic speech detection dataset based on the powerful Mongolian speech synthesis model in the laboratory, and conducted experiments and comparisons with mainstream synthetic speech detection frameworks on this dataset. We also open sourced the dataset and baseline models at https://github.com /ssmlkl/NCMMSC2023. This paper is the first in-depth study of synthetic speech detection in Mongolian, aiming to promote the development of synthetic speech detection in Mongolian and make a contribution to the research of synthetic speech detection in minority languages.

Key words: Synthetic Speech Detection; Mongolian; Deep learning

---

收稿日期：2023-10-29；修回日期：****-**-**

基金项目：国家自然科学基金青年科学基金项目（62206136）；广东省数字孪生人重点实验室（华南理工大学 ）开放课题（2022B1212010004）

共同第一作者

* 通讯作者

1 引言

合成语音检测[1]旨在辨别真实人类语音和计算机生成的虚假语音之间的区别，以应对日益增多的欺诈和虚假信息传播。在当今数字化社会中，合成语音技术的快速发展使得欺诈分子能够通过模仿真实声音来制造虚假信息，从而加剧了信息安全的风险，例如伪造具有社会影响力的人物的声音发布假新闻，或是对熟人声音进行伪造实施诈骗、获取他人信息等[2]，这将会对我国的社会信任、新闻真实性、监控和司法取证等方面带来巨大挑战，因此对于深度语音合成的内容需要进行有效的检测。目前已经有一些工作在进行合成语音检测的研究[3,4]，但仍然存在许多挑战。例如，如何准确地区分真实和合成的音频内容，以及如何防止攻击者通过提高合成技术的复杂度来绕过检测算法等，因此需要继续进行深入的研究以应对不断发展的深度语音合成技术带来的挑战。现有的合成语音检测方法大致分为两个方向，分别是基于统计模型的检测方法与基于深度学习的检测方法。基于统计模型的合成语音检测方法主要是利用语音信号的统计特征，建立分类器来区分真实语音和合成语音。其中，最常用的统计模型包括高斯混合模型（Gaussian Mixture Model，GMM）和支持向量机（Support Vector Machine，SVM）。高斯混合模型是一种基于概率密度函数的模型，主要用于语音信号的分类和识别。在合成语音检测领域，研究者们利用 GMM 来提取语音信号的倒谱特征，并使用这些特征来区分真实语音和合成语音。例如，S. Chetlur[5]等研究者提出了一种基于 GMM 的检测方法，该方法能够有效地识别出利用 Mel-cepstral 系数生成的合成语音。支持向量机是一种基于结构风险最小化的机器学习算法，具有较好的泛化性能。在合成语音检测领域，研究者们将 SVM 应用于语音信号的特征分类和识别。例如，Lu H[1]等研究者提出了一种基于对数似然比（LLR）的支持向量机（SVM）分类器，用于自动检测语音合成系统中的错误单元选择。Huang T[6]等人使用 GRU-SVM 模型进行合成语音检测的方法，通过结合 GRU 和 SVM 的优点，提高了语音信号的分类准确率。基于深度学习的合成语音检测方法，研究人员采用卷积神经网络[7]、深度神经网络[8]、递归神经网络[9]等来构建音频合成检测架构，基于深度学习的合成语音检测模型往往提取音频的不同特征，如通过对音频信号进行傅里叶变换，得到频谱图，再对频谱图进行逆变换，得到倒谱特征。此外，还可以提取音频的声学特征，例如梅尔频率倒谱系数（MFCC）、线性预测系数（LPC）等，以及语音信号的波形特征、共振峰特征[10,11]等。不难看出，现有的深度学习模型在合成语音检测工作中取得了很好的效果，主要是因为它们能够有效地提取音频的不同特征来进行鉴别。这些特征包括但不限于音频的倒谱特征、声学特征、波形特征和共振峰特征等。同时，在深度学习模型的帮助下，可以实现更高效、准确和自动化的检测工作。这些模型可以自动提取音频特征并进行比较，从而得到更加客观和准确的结果。

以上所介绍的合成语音检测工作均基于主流语言英语与汉语，然而全国范围内对于蒙古语等小语种的合成语音检测研究却相对较少，但针对于这些小语种的研究同样具有重要的意义。例如在蒙古语语音合成语音的研究中，Liu R 等人提出一种将蒙古语单词分割为词干和词缀的新方法，大大提高了蒙古语韵律短语预测的性能[12]，紧接着又提出了基于 DNN 的蒙古语语音合成系统，其性能优于传统的 HMM [13]。此外，他还引入了双向长期记忆（BiLSTM）模型，以改进蒙古语语音合成系统中的断句预测步骤，使其更适用于蒙古语系统[14]。之后，Liu R 提出了完全非自回归的，韵律自然度高的蒙古语语音合成模型 MonTTS[15]。由于小语种语言的复杂性和独特性，使得对于小语种的合成语音检测研究更具挑战性，为了推动小语种合成语音检测领域的发展，需要更多的学者和研究机构关注并投入此领域的研究工作。本工作是合成语音检测在蒙古语领域的首次尝试，我们依托于实验室强大的蒙古语语音合成模型等成果与先进的录音室设备构建了一个高质量的蒙古语合成语音检测数据集，基于此数据集搭建了三个基于时域信号与 LFCC（Line Spectral Frequency Cepstral Coefficients，线谱频率倒谱系数）信号的基线模型，并比较了三个基线模型在此数据集的效果，其中 LFCC 特征作为输入的 ResNet50 达到了最好的模型效果，EER 为 5.377，min t-DCF 为 0.183。

本工作的贡献如下：

1. 我们填补了蒙古语合成语音检测方面的空白，为小语种的语音合成检测研究做出了贡献。

2. 我们在实验室已有强大的蒙古语语音合成模型等前期研究成果的支撑下，构建了蒙古语合成语音检测数据集。

3. 我们针对蒙古语合成语音检测数据集，训练了三个主流合成语音检测模型作为基线模型，并进一步比较分析了实验结果。

本文第二章介绍蒙古语鉴伪数据集的构成，第三章介绍基于此数据集的三个基线模型，并展示比较了数据集在基线模型训练的结果，最后对本工作进行总结与展望。

2 蒙古语合成语音检测数据集

我们在实验室现有蒙古语语音合成模型和蒙古语语音合成数据集的支撑下，构建了蒙古语合成语音检测数据集，并在该数据集的基础上对主流的合成语音检测框架进行实验对比。在先前的工作中，我们公布了一个高质量的多说话人蒙古语语音合成数据集 MnTTS2[16]，随后我们在 MnTTS2 数据集的基础之上构建了蒙古语合成语音检测数据集。在第二部分内容中，我们简要介绍基于 MnTTS2 数据集和最先进的语音合成模型 VITS[17]构建的蒙古语合成语音检测数据集。

2.1 蒙古语语音合成数据集

MnTTS2 是一个多说话人的蒙古语语音合成数据集。MnTTS2 数据集有三个说话人，我们选择其中两个说话人：F1 和 F2 来构建蒙古语合成语音检测数据集，数据集的文件结构如图 1 所示。每位演讲者的录音文件和相应的文本集保存在以演讲者姓名命名的文件夹中。所有音频以 WAV 格式文件存储，采样率为 44.1 kHz，采样精度为 16 位。所有文本保存在以 UTF-8 编码的 TXT 文件中。音频文件名与相应的文本文件名相同，每个文件的名称由说话人 ID、文档 ID 和文本内容 ID 组成。

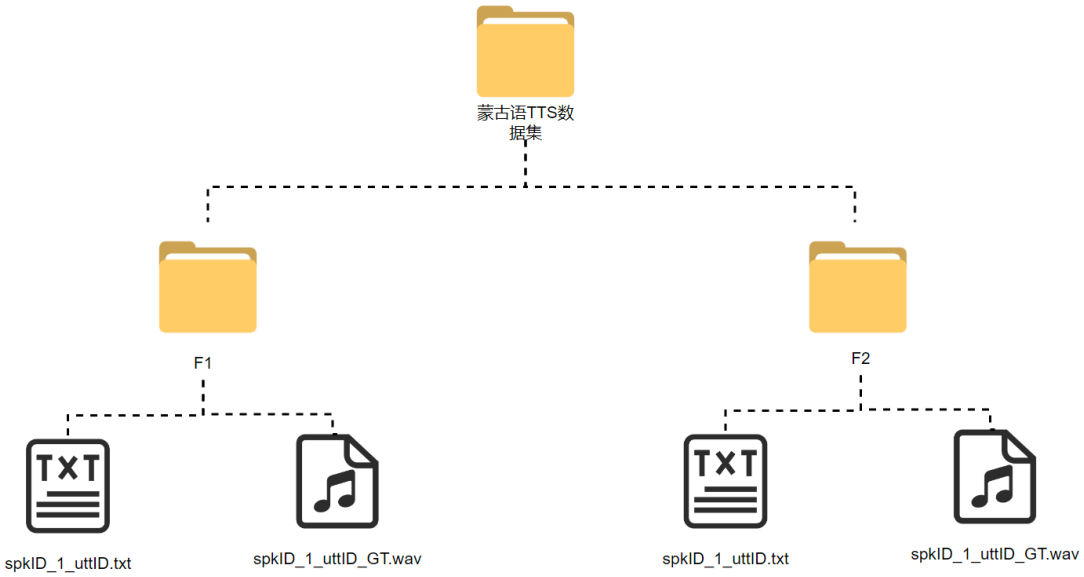

图 1 蒙古语 TTS 数据集的文件结构

Fig.1 The folder structure of the Mongolian TTS corpus.

2.2 VITS 语音合成模型

VITS 是一种并行的、完全端到端的 TTS 模型。该模型能够生成比两阶段 TTS 系统更加自然和高质量的语音。该模型通过归一化流程和对抗性训练过程增强的变分推导，极大程度上提高了生成模型的表达能力。提出的随机持续时间预测器能够从输入文本合成不同节奏的语音。这使得输入的文本能够以多种不同

的音调和节奏讲出。VITS 是目前最为先进的 TTS 模型。

VITS 具有易于训练、快速合成、对长文本的稳定性和丰富的语音多样性等特点，这是我们选择 VITS 的主要原因。数据集中每个说话者的数据按照 93:5:2 的比例被手动划分为训练集、测试集和验证集。整个 VITS 模型在两个 Tesla P100 GPU 上进行了训练。

2.3 语料库结构与统计

我们使用训练得到的 VITS 模型合成了蒙古语语音合成数据集中的每条音频。并将合成的数据集和真实数据集组合得到蒙古语合成语音检测数据集。数据集包含两个说话人：F1 和 F2。每个说话人数据包括若干条文本、与文本对应的真实语音和通过 VITS 合成的语音。数据集结构如图 2 所示，spk_ID_1_uttID.txt 代表说话人 ID 为 spkID，文档 ID 为 1，文本内容 ID 为 uttID 的语音文本。spkID_1_uttID_GT.wav 和 spkID_1_uttID_FK.wav 分别代表文本内容所对应的真实音频（Ground-Truth）和通过 VITS 合成的虚假音频（Fake）。所以音频以 wav 格式的文件存储，采样率为 22.05kHz,采样精度为 16 位。所有文本以 UTF-8 编码存储在 TXT 文件中。

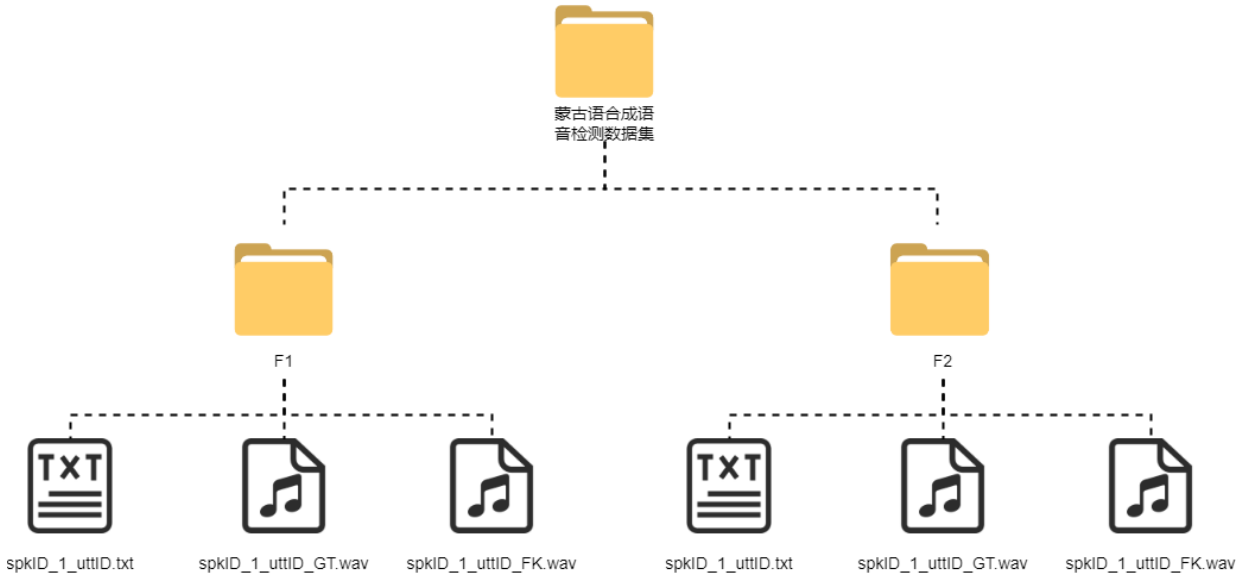

图 2　蒙古语音频鉴伪数据集的文件结构

Fig.2　The folder structure of the Mongolian audio fake detection dataset

我们用真实音频和合成音频及其所对应的文本共同构成蒙古语合成语音检测数据集。并将数据集按照 8:2:2 的比例分割为训练集、开发集和验证集。对于数据集中真实和合成音频的统计信息如表 1 所示，数据集中说话人均为女性，其中真实语音和合成语音的个数相等。训练集中包含 10907 条真实音频和 VITS 合成的音频，开发集包含 2767 条真实音频与合成音频，最后验证集中包含了 2606 条真实音频与合成音频。

表 1　训练、开发和评估数据集中非重叠目标说话人和语句的数量

Tab.1　Number of non-overlapping target speakers and utterances in the training, development and evaluation datasets.

Subset	#Speakers		#Utterances	
	Male	Female	Genuine	Spoofed
Training	0	2	10907	10907
Development	0	2	2767	2767
Evaluation	0	2	2606	2606

3 蒙古语合成语音检测实验

为了验证蒙古语合成语音检测数据集的可用性，我们分别以时域信号与音频的 LFCC 信号作为模型的输入搭建了三个基线模型，并分别比较了实验结果。其中，RawNet2 与 RawGAT 是以时域信号作为模型输入的基线模型，ResNet50 是以 LFCC 信号作为模型输入的基线模型。

在第三部分内容中，3.1 介绍了三个基线模型的模型结构，3.2 介绍了评估基线模型所用到的评价指标，3.3 展示并分析了基线模型在此数据集上的实验结果。

3.1 基线模型

RawNet2[18]：RawNet2 是基于深度学习的端到端合成语音检测模型，其直接将原始波形作为系统输入，进而提取合成语音的特征。并通过正交卷积减少滤波器之间的相关性，优化 Sinc-conv 的参数，从而提高辨别能力；同时引入时间卷积网络（TCN）来捕捉语音信号中的长时依赖关系，以进一步优化模型。相较于传统方法，RawNet2 避免了复杂的特征提取步骤，提高了检测准确性和鲁棒性。其通过引入残差块和批量归一化层，有效缓解了梯度消失问题，增强了模型表达能力。

Raw-GAT[19]：Raw-GAT 是基于图注意力网络（Graph Attention Network，GAT）的一种新型的端到端合成语音检测模型。该模型将语音信号转化为图结构，并利用 GAT 进行特征提取和分类。此框架的核心是 GAT 模块，它通过多头注意力机制，将节点的特征聚合为全局特征。具体来说，GAT 模块将输入的节点特征矩阵作为查询（Q）、键（K）和值（V），通过矩阵乘法和 softmax 函数计算节点之间的注意力权重，然后将节点特征矩阵与对应的权重矩阵相乘，得到全局特征。在 Raw-GAT 框架中，输入是语音信号的原始波形，经过预处理后，转化为节点特征矩阵。然后，通过 GAT 模块进行特征提取和分类。该模型可以直接处理原始波形作为系统输入，避免了传统方法中需要先进行特征提取的步骤。

ResNet50[7]：ResNet50 模型使用了残差学习的方法，它的主要特点是深度，有 50 层卷积神经网络，这个模型的深度使得它能够学习更复杂的特征，从而提高了准确率。ResNet50 的另一个特点是使用了全局平均池化层，这个层将每个特征图的所有像素的平均值作为该特征图的输出，这个层的作用是减少模型的参数数量，从而减少过拟合的风险。我们的基线模型将提音频的 LFCC 特征作为 ResNet50 模型的输入，以最终的二分类器来进行合成语音的判别。

3.2 评价指标

我们使用了以下两种评价指标来对基线模型的效果进行直观的比较。

EER（Equal Error Rate，等错误率）：在二元分类任务中，EER[20]表示在不同分类系统的阈值设置下，使误拒绝率（False Rejection Rate，FRR）等于误接受率（False Acceptance Rate，FAR）的阈值。EER 是一种衡量分类系统性能的方式，其中 FRR 和 FAR 是系统的两个关键错误率。如下式所示：

$$EER = \frac{FAR+FRR}{2} \quad (1)$$

其中，FAR 是指将实际为负的样本错误地判定为正类的概率，FRR 是指将实际为正的样本错误地判定为负类的概率。当 FAR = FRR 时，EER 的值最小，此时的分类器性能最佳。EER 是分类器性能评估中常用的指标之一，特别是在声纹识别、说话人识别等应用领域中。

min t-DCF（最小化 t-DCF）：在语音领域，min t-DCF[21]是一项重要的性能指标，t-DCF 代表了目标检测任务中的性能评估，其中系统旨在区分真实语音信号和合成的语音信号，以确保系统不易受到欺诈性攻击。首先，t-DCF 考虑了误拒绝和误接受的代价，误拒绝指的是将真实语音信号错误地标记为伪造，而误接受则是将合成语音信号错误地标记为真实。min t-DCF 的目标是找到一个最优的系统配置，以最小化总体代价。最小化的代价通过调整鉴别阈值来实现，调整阈值可以在降低误接受和误拒绝之间进行权衡，以满足特定应用需求。

3.3 实验结果

我们将三个基线模型在蒙古语合成语音检测数据集上分别进行了训练，模型的 EER 与 min t-DCF 评估结果如表 2 所示。从表中不难看出，目前基于汉语、英语等主流语言所使用的深度学习模型在蒙古语合成语音检测方面的表现效果欠佳，原因可能是蒙古语语言的复杂性和独特性，使得对于小语种的合成语音检测研究更具挑战性。在三个基线模型中，以音频的 LFCC 特征作为模型输入的 ResNet50 在 EER 与 min t-DCF 的结果均优于以原始音频信号作为模型输入的另外两个基线模型 RawNet2 与 Raw-GAT，原因可能是 LFCC 捕获了声音的谱线条纹和共振频率等重要信息，有助于区分蒙古语语音中较为复杂的音素、音节和语音特征，且 LFCC 的计算包括了对音频信号的预加重、分帧、快速傅立叶变换等预处理步骤，这些步骤也有助于突出蒙古语语音特征，并使模型更容易捕捉和分析。在今后的研究工作中，我们将针对于蒙古语特殊的韵律等特征，构建适用于蒙古语合成语音检测的更加强大的模型，并继续扩充和完善蒙古语合成语音检测数据集。

表 2　三个基线模型在蒙古语合成语音检测数据集上针对于评价指标 EER 与 min t-DCF 上的表现。

Tab.2　Three baseline models were used to evaluate the performance of EER and mint-DCF on the Mongolian synthetic speech detection dataset.

Baselines	Input-feature	EER	min t-DCF
RawNet2[14]	Raw-audio	8.323	0.278
Raw-GAT[16]	Raw-audio	7.368	0.229
ResNet50[10]	**LFCC**	**5.377**	**0.183**

4　结论

小语种的合成语音检测研究具有重要意义，但目前对于蒙古语等小语种的研究相对较少，为了推动小语种合成语音检测领域的发展，需要更多学者和研究机构关注并投入此领域的研究工作。本工作是合成语音检测在蒙古语领域的首次尝试，我们通过实验室的蒙古语语音合成模型和先进的录音室设备，构建了高质量的蒙古语合成语音检测数据集，并基于该数据集搭建了三个基于时域信号与 LFCC 信号的基线模型，并比较了三个基线模型在此数据集的效果，由于蒙古语语言韵律的复杂性和独特性，现有的基于汉语与英语的主流深度学习模型对于蒙古语并不十分适用，其为蒙古语合成语音检测的工作增加了研究的挑战性。在未来的工作中，我们将致力于蒙古语语言自身的特点搭建更适合于蒙古语语言特点的合成语音检测模型，为小语种语言的合成语音检测研究作出新的贡献。

参考文献

[1] Lu H , Ling Z H , Wei S ,et al.Automatic Error Detection for Unit Selection Speech Synthesis Using Log Likelihood Ratio based SVM Classifier[C]//Annual conference of the International Speech Communication Association;INTERSPEECH 2010.2011.

[2] J. Mallet, L. Pryor, R. Dave, and M. Vanamala, Deepfake detection analyzing hybrid dataset utilizing cnn and svm[J].ArXiv, vol. abs/2302.10280, 2023.

[3] Ferrer, A., & Font, M. (2019). Detection of Voice Conversion Using Temporal, Spectral and Prosodic Measures[J].Speech Communication, 2019, 107166.

[4] Zhang, Q., Wang, Z., & Wu, X. (2018). Automatic Detection of Voice conversion using machine learning[C]//Proc. of the 5th Int. Conf. on Signal Processing and Communication Engineering (ICSPCE). IEEE, 2018: 1-5.

[5] S. Chetlur, V. Navot, and H.S. Seung. Voice Forgery Detection Using Gaussian Mixture Model (GMM) Based on Mel-Cepstral Coefficients[J].EURASIP Journal on Audio, Speech, and Music Processing, vol. 2014, no. 1, 2014.

[6] [1] Huang T , Wang H , Chen Y ,et al.GRU-SVM Model for Synthetic Speech Detection[C]//International Workshop on Digital Forensics and Watermarking.Springer, Cham, 2019.DOI:10.1007/978-3-030-43575-2_9.

[7] Z. Lei, Y. Yang, C. Liu, and J. Ye.Siamese convolutional neural network using gaussian probability feature for spoofing speech detection[C]//Proc. Interspeech 2020, pp. 1116–1120, 2020.

[8] Li J , Sun M , Zhang X .Multi-task learning of deep neural networks for joint automatic speaker verification and spoofing detection[C]//Asia-Pacific Signal and Information Processing Association Annual Summit and Conference.IEEE, 2019.DOI:10.1109/apsipaasc47483.2019.9023289.

[9] Chen Z , Zhang W , Xie Z ,et al.RECURRENT NEURAL NETWORKS FOR AUTOMATIC REPLAY SPOOFING ATTACK DETECTION[C]//2018:2052-2056.DOI:10.1109/ICASSP.2018.8462644.

[10] Yamagishi J , Wang X , Todisco M ,et al.ASVspoof 2021: accelerating progress in spoofed and deepfake speech detection[J]. 2021.DOI:10.48550/arXiv.2109.00537.

[11] ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech[J]. 2019.DOI:10.48550/arXiv.1911.01601.

[12] Liu, R., Bao, F., Gao, G., & Wang, W. Mongolian prosodic phrase prediction using suffix segmentation. In 2016 International Conference on Asian Language Processing (IALP), pages 250 – 253. IEEE, 2016.

[13] Liu, R., Bao, F., Gao, G., & Wang, Y. Mongolian text-to-speech system based on deep neural network. In National Conference on Man-Machine Speech Communication, pages 99–108. Springer, 2017.

[14] Liu, R., Bao, F., Gao, G., Zhang, H., & Wang, Y. Improving Mongolian phrase break prediction by using syllable and morphological embeddings with BiLSTM model. In Interspeech, pages 57–61, 2018.

[15] Liu, R., Kang, S., Gao, G., et al. MonTTS: Fully Non-Autoregressive Real-time, High-fidelity Mongolian Speech Synthesis Model[J]. Journal of Chinese Information Processing, 2022, 36(7): 86-97.

[16] Liang K, Liu B, Hu Y, et al. MnTTS2: An Open-Source Multi-Speaker Mongolian Text-to-Speech Synthesis Dataset[C]//National Conference on Man-Machine Speech Communication. Springer, Singapore, 2023: 318-329.

[17] Kim J, Kong J, Son J. Conditional variational autoencoder with adversarial learning for end-to-end text-to-speech[C]//International Conference on Machine Learning. PMLR, 2021: 5530-5540.

[18] Jung J W , Kim S B , Shim H J ,et al.Improved RawNet with Filter-wise Rescaling for Text-independent Speaker Verification using Raw Waveforms[J]. 2020.DOI:10.48550/arXiv.2004.00526.

[19] Tak H , Jung J W , Patino J ,et al.End-to-End Spectro-Temporal Graph Attention Networks for Speaker Verification Anti-Spoofing and Speech Deepfake Detection[J]. 2021.DOI:10.48550/arXiv.2107.12710.

[20] Khanrajshree N S A .Equal Error Rate and Audio Digitization and Sampling Rate for Speaker Recognition System[J]. Advanced Science Letters, 2014, 20(5a6).

[21] Kinnunen T, Lee K A, Delgado H, et al. t-DCF: a Detection Cost Function for the Tandem Assessment of Spoofing Countermeasures and Automatic Speaker Verification[J].

作者简介

张锦华 女，1999 年出生在内蒙古乌兰察布市，现于内蒙古大学研究生二年级在读，就读于计算机(软件)学院计算机科学与技术专业，主要研究方向为语音鉴伪。E-mail: zjh_imu@163.com。

梁凯麟 男，1998 年出生在甘肃省兰州市，现于内蒙古大学研究生二年级在读，就读于计算机(软件)学院电子信息专业，主要研究方向为跨说话人情感迁移。E-mail:liangkailin98@163.com。

刘瑞(1994--) 博士，研究员，主要研究领域为机器学习、基于深度学习的语音合成和自然语言处理等。

E-mail:liurui_imu@163.com。

创新点说明

本工作首次尝试在蒙古语领域进行蒙古语合成检测的研究。我们利用实验室强大的蒙古语语音合成模型和先进的录音室设备，构建了高质量的蒙古语合成语音检测数据集。在此基础上，我们搭建了三个基于时域信号与 LFCC 信号的基线模型，并比较了它们在数据集上的效果。由于蒙古语语言的韵律复杂性和独特性，现有的基于汉语和英语的深度学习模型并不完全适用于蒙古语合成语音检测的研究，这增加了研究的挑战性。未来，我们将致力于根据蒙古语自身的特点，构建更适合于蒙古语语言特点的合成语检测模型，为小语种语言的合成语检测研究做出新的贡献。