计算机科学
COMPUTER SCIENCE

基于LFBank与FBank混合特征的声纹识别研究

崔琳，王芷悦

引用本文

崔琳，王芷悦. 基于LFBank与FBank混合特征的声纹识别研究[J]. 计算机科学, 2022, 49(11A): 211000194-5.

CUI Lin, WANG Zhi-yue. Study on Voiceprint Recognition Based on Mixed Features of LFBank and FBank [J]. Computer Science, 2022, 49(11A): 211000194-5.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.211000194

基于 LFBank 与 FBank 混合特征的声纹识别研究

崔　琳　　王芷悦
西安工程大学电子信息学院　西安 710699

摘　要　语音特征提取是声纹识别过程中的重要步骤，对于声音频率的分布男性与女性差距较大，但现有的特征提取算法并没有针对不同性别声音频率特性做出相应改进。针对上述问题，提出了为女性声纹识别所设计的语音特征提取算法 LFBank，将线性滤波器组用于特征提取过程，利用其线性分布的特点弥补传统梅尔滤波器组提取高频区域信息时的不足。另一方面，为了突破单一性别局限，拓宽应用场景，综合线性滤波器组与梅尔滤波器组的优势，将 LFBank 与 FBank 特征结合得到混合特征向量进行声纹识别。将 LFBank 和常用特征 FBank 与 MFCC 进行实验对比，实验结果表明，基于线性滤波器组的特征向量在识别女性声音时更有优势。对于混合特征而言，在与单一特征的对比实验中，混合特征能够达到比单一特征更好的识别效果，具有更广泛的应用场景。

关键词：声纹识别；特征提取；声音频率；线性滤波器组；梅尔滤波器组；混合特征
中图法分类号　TN912

Study on Voiceprint Recognition Based on Mixed Features of LFBank and FBank

CUI Lin and WANG Zhi-yue
School of Electronic Information，Xi'an Polytechnic University，Xi'an 710699，China

Abstract　Speech feature extraction is an important step in the process of voiceprint recognition. There is a large gap between men and women in the distribution of sound frequency，but the existing feature extraction algorithms have not made corresponding improvements for the sound frequency characteristics of different genders. To solve the above problems，a speech feature extraction algorithm LFBank designed for female voiceprint recognition is proposed. The linear filter banks is introduced into the feature extraction process，and its linear distribution is used to make up for the deficiency of the traditional Mel filter banks in extracting high-frequency region information. On the other hand，in order to break through the limitation of single gender and broaden the application scenarios，combining the advantages of linear filter banks and Mel filter banks，LFBank and FBank features are combined to obtain mixed feature vectors for voiceprint recognition. The LFBank is compared with the commonly used feature FBank and MFCC，and experimental results show that the feature vector based on linear filter bank has more advantages in recognizing female voice. For mixed features，in the comparison experiment with single features，they can achieve better recognition effect than single features and have a wider range of application scenarios.

Keywords　Voiceprint recognition，Feature extraction，Sound frequency，Linear filter banks，Mel filter banks，Mixed feature

1　引言

声纹识别又称说话人识别，是生物识别技术的一种，它是通过对比不同语音的深度特征来达到区分说话人的目的。相比其他需要接触采集的生物识别方式，声纹识别获取样本方式更多样，被采集者接受度更高。声纹识别在刑事侦查、国防监听、金融证券、生活加密等领域都有重要的应用[1]。

语音是人类特有的生理行为，对于每个人来说，其语音都有各自的生物特性，如基因差异带来的生理结构不同、后天形成的发声差异以及性别的不同等[2]。可以从不同说话人的语音中提取出表现其说话特点的语音表现特征，之后通过分析语音特征并利用说话人语音的识别模型来对说话人进行判别[3]。因此，在声纹识别模型中语音特征提取是十分重要的部分。

对于语音特征提取算法，梅尔倒谱系数(Mel-scale Frequency Cepstral Coefficients，MFCC)[4]是其中的经典算法之一。在 MFCC 特征提取时，语音预处理将语音信号转换成时域信号，然后进行快速傅里叶变换将时域信号转换成频域信号，利用三角滤波器将语音的频率转换成人耳能够接收的频率，这就是最常用的 MFCC 语音特征[5]。在深度学习[6]流行之前，进行声纹特征识别的大部分是机器学习[7]模型。MFCC 特征具有应用离散余弦变换对滤波器组系数去相关的步骤，对于机器学习算法而言，这一步是十分必要的，因此传统的声纹识别模型[8-10]使用的大多为 MFCC 特征。后来，FBank(Filter Banks)特征[11-13]被提出，FBbank 特征没有应用离散余弦变换进行去相关处理，与 MFCC 相比，其计算量更小且特征相关度更高，包含更多的信息。目前，随着深度学习

基金项目：国家自然科学基金青年项目(61901347)
This work was supported by the National Natural Science Foundation of China(61901347).
通信作者：崔琳(864940295@qq.com)

技术的不断发展，神经网络[14-15]开始被应用于越来越多的领域。在基于深度学习的声纹识别模型[16-18]中，更需要 FBank 这种更符合声音信号本质的语音特征。

在声音信号的频率分布上，成年女性与成年男性有很大不同。男性的声带较长且宽厚，发声时震动频率低，女性声带较短且窄薄，发声时震动频率高。因此，对男性和女性的声音进行分别研究是提升声纹识别准确率的有效途径。FBank 特征所使用的 Mel 滤波器组强调了语音的低频信息，却忽视了高频区域，对女性声音进行特征提取时会损失重要信息。线性滤波器组是提取线性频率倒谱系数特征[19]所使用的滤波器组，它对于不同频率的声音敏感度是线性的，因此能够在高频区域捕获更多的频谱细节。对于女性声音的识别，使用线性滤波器组会比 Mel 滤波器组更有优势。

综合上述研究，本文提出基于线性滤波器系数的特征提取算法 LFBank(Linear Filter Banks)，分别对男性声音与女性声音进行研究，引入线性滤波器组捕获更多高频区域信息，弥补 Mel 三角滤波器组的不足。同时，为了获得准确度高且应用场景丰富的语音特征表示，本文将 LFBank 与 FBank 两种特征进行结合。在 LSTM[20]为主干网络的基础上，将男性与女性语音分开进行识别实验，比较不同特征提取算法在不同性别情况下的识别准确率。最后，将两种特征混合，用于识别模型的输入数据，比较单一特征与混合特征的识别表现。

2 特征提取研究背景

特征提取是将原始的声音文件经过计算转化得到特征参数表示的过程，这个过程是声纹识别模型中的关键步骤之一。声纹特征带有每个说话人不同的个性信息特征，主要着眼于语音信号的频谱结构。从语音波形中提取出能够反映说话人特性的特征参数，不仅可以去除语音信号中的冗余信息，还可以减少运算量和所需的存储空间[21]。MFCC 和 FBank 特征是常用的特征提取算法，它们使用的都是 Mel 滤波器组，不同之处在于 FBank 特征没有进行离散余弦变换。离散余弦变换去除了各维信号之间的相关性，这也使得没有进行此步骤的 FBank 特征更具有声音的本质信息且比 MFCC 计算量更小。声纹识别的传统方法高斯混合模型-通用背景模型(Gaussian Mixture Model-universal Background Model, GMM-UBM)广泛使用时，使用 MFCC 语音特征是主流，基于深度学习的方法流行之后，在实验中发现 FBank 的表现更好。

Mel 滤波器组是将功率谱通过一组 Mel 刻度的三角滤波器来提取频带。Mel 滤波器组的工作原理是仿照人耳感知声音的方式而设计的，人耳对不同频率的声音有着不同的灵敏度，且这种变化是非线性的，在较低的频率下更具辨别力，在较高的频率下辨别力变低。因此，Mel 滤波器组的分布也是非线性的，在低频区域滤波器分布比较密集，在高频区域滤波器分布比较稀疏。式(1)表示了声音真实频率与 Mel 频率之间的对应关系。

$$Mel(f)=2595\times\log_{10}\left(1+\frac{f}{700}\right) \tag{1}$$

FBank 特征向量是语音信号经过 Mel 滤波器组形成的频谱再进行对数能量处理得到对数频谱。计算过程如式(2)所示：

$$S_M(m)=\ln\left(\sum_{k=0}^{N-1}|X(k)|^2H_m(k)\right),0\leqslant m\leqslant M \tag{2}$$

其中，$X(k)$为信号频谱，$H_m(k)$为 Mel 滤波器组的传递函数。

将 $S_M(m)$经离散余弦变换得到倒谱频域，即可得到 MFCC 特征，计算式如下：

$$c(n)=\sum_{m=1}^{M-1}S(m)\cos\left(\frac{\pi n(m+12)}{M}\right) \tag{3}$$

3 基于 LFBank 特征的声纹识别

3.1 LFBank 特征

声纹识别的过程是通过计算语音的深度特征找到对应说话人，语音数据是最根本的识别依据。人类的发声方式在生理层面可以简单描述为：声带振动发出一定频率的声音，之后经过声道，最后到达嘴部将声音辐射出去。通过改变舌头的位置、嘴部张开的大小等可以帮助我们发出各种各样的声音，进而形成语言。声带是发声器官的主要部分，声带结构不同发出的声音也会有所不同。成年男性声带长而宽，长度为 18～24 mm，成年女性声带形状较男性的来说正好相反，是短而窄的，长度为 14～18 mm，这就导致成年女性声音共振峰频率相对较高。

针对男女声音频率分布不同的特点，本文提出基于线性滤波器组的特征提取算法 LFBank。LFBank 特征提取算法与 FBank 特征提取过程的不同之处在于 LFBank 使用线性滤波器组。线性滤波器组同样是由一组三角带通滤波器组成，但它们不是按照 Mel 频率分布，而是按照线性排列，它可以得到声音信号的线性频谱表示。使用线性滤波器组计算经过快速傅里叶变换后得到的功率谱，会在高频率处获得更好的分辨率。LFBank 的具体提取过程如图 1 所示。

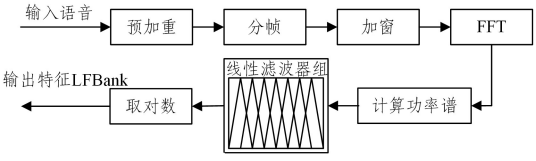

图 1 LFBank 特征提取过程

Fig. 1 LFBank feature extraction process

预加重的过程是将语音信号输入传递函数为式(4)的高通滤波器[22]进行处理，可以增强语音信号中的高频部分，同时使整个频段能有相同的信噪比。预加重的目的是为了补偿在发声过程中语音信号被抑制的高频部分，消除发声过程中声带和嘴唇的效应。

$$y(n)=x(n)-a*x(n-1) \tag{4}$$

其中，a 为预加重系数。

分帧是将语音信号按照给定的时间长度进行切割，切割后的小段称为帧。语音信号是短时平稳的，因此将整段非平稳的信号分割为短时帧后才能进行傅里叶变换。为了避免分帧后的相邻帧变化过大，一般会保留一些两帧间的重叠区域。

加窗是在信号分割成短帧后对每一帧乘上一个不断移动的有限长窗函数的过程。加窗是为了增加帧左端和右端的连续性，以减少频谱泄露。式(5)是本文使用汉明窗[23]的函数表达式，N 为窗口长度。

$$w(n,a)=\begin{cases}(1-a)-a*\cos\left(\dfrac{2\pi n}{n-1}\right), & 0\leqslant n\leqslant N-1\\ 0, & \text{其他}\end{cases} \tag{5}$$

语音信号经过预加重、分帧、加窗的预处理过程后，要对每一帧数据进行快速傅里叶变换。快速傅里叶变换是为了将信号从时域转换到频域，在时域上很难观察信号的特性，对于语音信号的计算识别等过程都是在频域中进行的，计算式[24]如下：

$$X(k)=\sum_{n=0}^{N-1}x(n)e^{\frac{-j2\pi k}{N}},0\leqslant k\leqslant N-1 \tag{6}$$

接下来对信号的频谱取模的平方以计算语音信号的功率谱[25]，计算式如下：

$$p(k)=\frac{1}{N}|x(k)|^2 \tag{7}$$

将获得的信号频谱送入线性滤波器组，式(8)是单独一个滤波器的数学表达式，每个滤波器都是一个带通滤波器，多个带通滤波器线性排列组成线性滤波器组。经过快速傅里叶变换的信号分别与每个滤波器进行频率相乘累加，得到的值即为该帧数据在该滤波器对应频段的能量值。

$$L_m(k)=\begin{cases}0, & k<f(m-1)\\ \dfrac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1)\leqslant k\leqslant f(m)\\ \dfrac{f(m+1)-k}{f(m+1)-f(m)}, & f(m)<k\leqslant f(m+1)\\ 0, & k>f(m+1)\end{cases} \tag{8}$$

其中，m 为带通滤波器的中心频率。

最后将经过滤波器组形成的频谱再进行对数能量处理后得到 LFBank 的计算式，如下：

$$S_L(m)=\ln\left(\sum_{k=0}^{N-1}|X(k)|^2L_m(k)\right),0\leqslant m\leqslant M \tag{9}$$

3.2　基于混合特征的端到端声纹识别模型

LFBank 特征在表示高频区域信息时更有优势，应用此特征来弥补 FBank 在高频处提取特征信息不足的问题。同时 FBank 作为常用特征表示，由于 Mel 滤波器组的分布特点，使其在处理男性声音时更有优势。使用单一特征作为声纹识别模型的输入具有相应的局限性，为了综合两种特征的优势，得到准确率高且应用场景广泛的特征向量，本文将 LFBank 与 FBank 两种特征进行结合。具体混合特征方式如式(10)所示：

$$S_{Mix}=[(S_M),(S_L)] \tag{10}$$

对于端到端的声纹识别模型来说，神经网络是计算语音特征深度嵌入的工具，主干网络的选择非常重要。本文使用长短时记忆网络(Long Short-term Memory，LSTM)做为主干网络，LSTM 是改进的循环神经网络(Recurrent Neural Network，RNN)，它改变了传统 RNN 网络的内部结构，避免梯度在传递过程中的大量连乘，使网络只对有价值的信息进行记忆。LSTM 擅于处理序列信号，具有信息持久性，因此适合用来处理具有时间连续性的语音信号。

本文建立基于 LFBank 与 FBank 混合特征的端到端声纹识别模型，模型的主要流程为：首先将原始语音文件进行预处理；然后信号经快速傅里叶变换后再经过 Mel 滤波器组与线性滤波器组得到混合特征向量；最后将特征向量输入 LSTM 网络进行深度嵌入，计算每条语音深度特征间的差距，通过 softmax 分类器进行分类，输出识别结果。图 2 为本文建立模型的整体流程图。

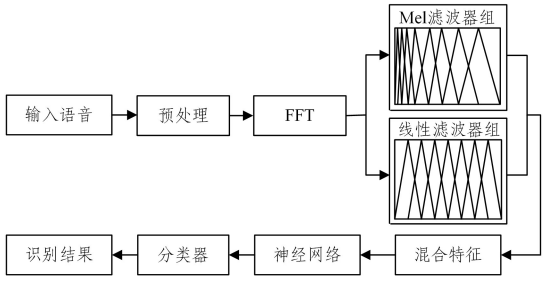

图 2　模型的整体流程图

Fig. 2　Overall flow chart of model

4　实验

4.1　实验数据集

本文实验是基于 AISHELL 数据集进行男性声音数据集、女性声音数据集及男女各半数据集的制作。AISHELL 数据集是由希尔贝壳公司制作，用于非商业的语音识别或声纹识别。说话人的身份背景广泛，来自中国不同地区，说话人在室内环境下使用中文普通话，语气平缓地读出给定语句。一共 1 991 位说话人，总时长 1 000 h，约有 100 万条短句。本文选取其中音频文件采样率为 16 kHz，16 bit，单声道，WAV 格式的语音文件。将男性声音与女性声音分开进行实验，分别制作 100 名、300 名、500 名男性说话人数据集，100 名、300 名、500 名女性说话人数据集，100 名、300 名、500 名男女各半说话人数据集。其中，每人 10 句话，每句话的时长为 1～3 s。所有数据集都按照 9∶1 划分训练集和测试集。

4.2　实验环境设置

本文实验在 windows64 位操作系统上进行，采用 pytorch 深度学习框架。进行模型训练时使用 SGD 优化器，学习率设置为 0.01，整个训练过程共经历 500 轮迭代。本文建立端到端声纹识别模型，使用主干网络为 LSTM，LSTM 设置为 3 层，每层的隐藏节点数为 256。在特征提取过程中，设置预加重系数为 0.97，帧长和帧移分别为 512，160，汉明窗大小为 400，预处理后的信号进行长度为 512 的快速傅里叶变换，实验使用的两种滤波器组均由 40 个滤波器排列而成。

本文采用模型测试时的等错误率(Equal Error Rate，EER)作为评价指标。等错误率是错误拒绝率(False Rejection Rate，FRR)与错误接受率(False Acceptance Rate，FAR)相等时的值，等错误率越小说明两种错误率同时越低，即声纹识别结果越准确。

4.3　实验结果分析

4.3.1　LFBank 特征

本文针对频谱分布较高的女性声音，提出了一种新的语音特征提取算法 LFBank。为了验证所提算法的有效性，本文制作全女性语音数据集、全男性语音数据集与男女混合数据集，分别进行不同特征提取算法的对比实验。首先进行全女性数据集实验，将 4 种不同特征向量作为相同神经网络的输入，分别是 LPCC，MFCC，FBank，LFBank，实验结果如图 3 所示。从图 3 可以看出，LPCC 和 MFCC 在基于神经网络的识别模型中表现最差，这是因为它们进行离散余弦变换后，将计算得到数据之间的关联性去除，不适合神经网络的运算方式。在说话人数为 100 时，LFBank 与 FBank 特征等错误率

都较高。这是因为神经网络需要大量的数据支持才能发挥出较好的效果。LFBank 在说话人数为 500 时优势最明显，相比 FBank 特征的等错误率下降比例约为 19%，在说话人数为 300 时下降比例约为 6%。这表明在数据充足时，针对女性语音，LFBank 用其采集高频信息的优势在实验中获得了更好的识别结果，同时也验证了本文方法的有效性。

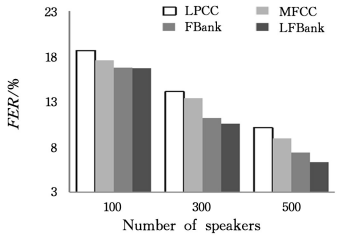

图 3 女性数据集识别结果

Fig. 3 Female dataset recognition results

对于男性语音，本文同样使用 LPCC，MFCC，FBank，LFBank 对进行实验，实验结果如图 4 所示。从图 4 可以看出，不管数据量多大，FBank 特征提取算法在本文的模型中等错误率在四者之间保持最低。这是因为，男性语音有效信息大多分布在低频处，Mel 滤波器组在低频区域分布最密集，采集到的信息最多，进而在男性语音识别上更有优势。另一方面，FBank 具有一定的鲁棒性，在数据较少时，也能够将等错误率上升数值保持在一定水平内。

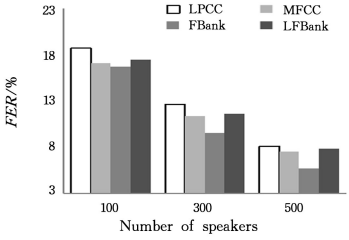

图 4 男性数据集识别结果

Fig. 4 Male dataset recognition results

在男女各半数据集上的实验结果如图 5 所示，从图中可以看出作为经典特征提取算法的 FBank 依然发挥了自身的优势，在四者中保持最低的等错误率。但是相对于全男性数据集的实验，LFBank 与 FBank 的差距变小了，这是因为 LFBank 在女性声音中的优势。同时，LFBank 与 LPCC 和 MFCC 特征相比保持了明显优势。

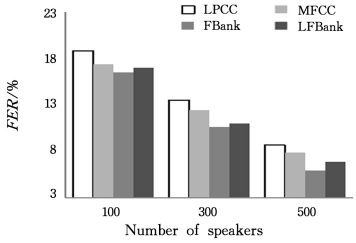

图 5 男女各半数据集识别结果

Fig. 5 Half male and half female dataset recognition results

4.3.2 混合特征

为了突破单一特征在声纹识别时对性别的局限，增加特征向量的应用场景和模型识别准确度，本文将 FBank 和 LFBank 两种特征进行结合，得到混合特征作为神经网络的输入，建立基于混合特征的声纹识别模型并与单一特征进行实验对比。将混合特征与单一特征分别在女性数据集、男性数据集、男女各半数据集上进行实验，实验结果如表 1－3 所列。

表 1 女性数据集上的识别结果

Table 1 Female dataset recognition results

	100	300	500
FBank	16.77	11.22	7.03
LFBank	16.69	10.56	6.31
mix- feature	14.07	8.50	4.74

表 2 男性数据集上的识别结果

Table 2 Male dataset recognition results

	100	300	500
FBank	16.70	9.53	5.69
LFBank	17.49	11.64	7.82
mix- feature	14.51	8.32	4.92

表 3 男女各半数据集上的识别结果

Table 3 Half male and half female dataset recognition results

	100	300	500
FBank	16.58	10.64	5.93
LFBank	17.12	10.98	6.87
mix- feature	14.39	8.18	4.63

从女性数据集的实验结果来看，混合特征在 3 种数据量情况下均达到最小的等错误率。在数据量为 100 人时，混合特征的等错误率相比 FBank 与 LFBank 特征分别下降约 16.1%和 15.7%，在数据量为 300 人时下降比例约为 24.2%和 19.5%，在数据量为 500 人时下降比例约为 36.1%和 24.9%。同样地，在男性数据集和男女各半数据集上，混合特征相比单一特征也得到了最好的识别结果，其中男女各半数据集 500 人时达到最低的等错误率。

从上述实验结果可以看出，混合特征综合了 FBank 与 LFBank 两种特征的优势，在女性、男性或者男女都有的情况下，相比单一特征的识别效果有明显提升，且在数据量不够的情况下依然能够保持相对最低的等错误率。这一实验结果也验证了，本文所提混合特征能够突破对性别的限定，相比单一特征能够在更多识别场景下使用，且能达到更好的识别结果。

结束语 为进一步提高声纹识别准确率，本文分析了不同人群声音特点并提出了专门的特征提取算法。针对女性声音频率较高的特点，提出了基于线性滤波器组的特征提取方法 LFBank，并在实验中将本文算法与 MFCC 和 FBank 进行了对比，验证了 LFBank 在女性声音上的优势。另一方面，为了综合 LFBank 与 FBank 的优势，摆脱单一性别的限制，将两种特征进行混合，分别在不同性别的数据集上进行实验，实验结果表明混合特征能够达到比单一特征更好的识别效果，同时有更大的使用范围。本文针对不同性别的声音提出了改进方法，为声纹识别在实际中的应用提供了新思路。

本文针对成年女性声纹识别提出了一种新的语音特征提取算法，没有对儿童和老人群体进行声音特质研究。在下一步的工作中，将分析更多人群的声音特性，提出针对性的声纹识别算法。

参 考 文 献

[1] PODDAR A, SAHIDULLAH M, SAHA G. Speaker verification

with short utterances:a review of challenges,trends and opportunities[J]. Pattern Recognition,2018,7(2):91-101.

[2] HANSEN J H,HASAN T. Speaker recognition by machines and humans: A Tutorial Review[J]. IEEE Signal Processing Magazine,2015,32(6):74-99.

[3] AI J Q,ZUO Y,LIU J X,et al. A hierarchical clustering approach for speech feature extraction based on cosine similarity [J]. Application Research of Computers,2020,37(S2):147-149.

[4] CHOWDHURY A,ROSS A. Fusing MFCC and LPC features using 1D triplet CNN for speaker recognition in severely degraded audio signals[J]. IEEE Transactions on Information Forensics and Security,2020,15(1):1616-1629.

[5] ZHOU P,SHEN H,ZHENG K P. Speaker recognition based on combination of MFCC and GFCC feature parameters[J]. Journal of Applied Sciences,2019,37(1):24-32.

[6] ESTEVA A,ROBICQUET A,RAMSUNDAR B,et al. A guide to deep learning in healthcare[J]. Nature Medicine, 2019, 25(1):24-29.

[7] CARLEO G,CIRAC I,CRANMER K,et al. Machine learning and the physical sciences[J]. Reviews of Modern Physics,2019, 91(4):045002.

[8] ZHAO F,YU Y. Two-level voiceprint recognition algorithm based on VQ and HMM[J]. Journal of Guilin University of Electronic Technology,2017,37(1):8-14.

[9] ZEINALI H,SAMET H,BURGET L. HMM-based phrase-independent i-vector extractor for text-dependent speaker verification[J]. IEEE/ACM Transactions on Audio,Speech,and Language Processing,2017,25(7):1421-1435.

[10] CHEN N,VILLALBAI J,DEHAK N. An Investigation of Nonlinear i-vectors for Speaker Verification[C]// Interspeech. 2018: 87-91.

[11] WANG J,LI L,WANG D,et al. Research on generalization property of time-varying Fbank-weighted MFCC for i-vector based speaker verification[C]// The 9th International Symposium on Chinese Spoken Language Processing. IEEE,2014:423-423.

[12] WANG G B,ZHANG W Q. An RNN and CRNN based approach to robust voice activity detection[C]// 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. IEEE,2019:1347-1350.

[13] CAO J,CAO M,WANG J,et al. Urban noise recognition with convolutional neural network[J]. Multimedia Tools and Applications,2019,78(20):29021-29041.

[14] CHENG T,WANG X,HUANG L,et al. Boundary-preserving mask r-cnn[C]// European Conference on Computer Vision. 2020:660-676.

[15] KARITA S,CHEN N,HAYASHI T,et al. A comparative study on transformer vs rnn in speech applications[C]// 2019 IEEE Automatic Speech Recognition and Understanding Workshop. IEEE,2019:449-456.

[16] RAVANELLI M,BENGIO Y. Speaker recognition from raw waveform with sincnet[C]// 2018 IEEE Spoken Language Technology Workshop. IEEE,2018:1021-1028.

[17] YU L F,LIU Q. Research and application of deep recurrent neural networks based voiceprint recognition[J]. Application Research of Computers,2019,36(1):153-158.

[18] ZHANG C,KOISHIDA K,HANSEN J H. Text-independent speaker verification based on triplet convolutional neural network embeddings [J]. IEEE/ACM Transactions on Audio, Speech,and Language Processing,2018,26(9):1633-1644.

[19] KUMAR A,SHAHNA W S. Robust detection of vowel onset and end points[C]// 2020 International Conference on Signal Processing and Communications(ICSPCC),IEEE,2020:1-5.

[20] YU Y,SI X,HU C,et al. A review of recurrent neural networks:LSTM cells and network architectures[J]. Neural Computation,2019,31(7):1235-1270.

[21] SISMAN B,ZHANG M,LI H Z. Group sparse representation withwavenet vocoder adaptation for spectrum and prosody conversion[J]. IEEE/ACM Trans on Audio,Speech,and Language Processing,2019,27(6):1085-1097.

[22] LUO H T. Pre-processing of speech signal[J]. Journal of Fujian Computer,2018,34(5):91-92.

[23] XIE X J. Research on feature combination method in speaker recognition [D]. Xiangtan:Xiangtan University,2016.

[24] LOU C W,CHAN C K,CHENG P H,et al. FFT-based multirate signal processing for 18-band quasi-ansi sl. 11 1/3-octave filter bank[J]. IEEE Trans on Circuits and Systems II:Express Briefs,2019,66(5):878-882.

[25] SHI L,AHMAD I,HE Y,et al. Hidden Markov model based drone sound recognition using MFCC technique in practical noisy environments[J]. Journal of Communications and Networks, 2018,20(5):509-518.

CUI Lin, born in 1984, lecturer. Her main research interests include speech signal processing, array signal processing and so on.