DOI:10.19651/j.cnki.emt.2416108

基于 LMD 改进特征提取的三路病理语音识别*

张 楠 陈媛媛 陈鑫钰 侯懿桃
(中北大学信息与通信工程学院 太原 030024)

摘 要:针对发音障碍患者发音不够清晰准确,导致病理语音识别率低的问题,提出一种基于 LMD 改进的 Gammatone 滤波器组图谱特征提取算法进行三路病理语音识别,首先,该算法采用 LMD 分解语音信号,对分解后的各语音分量做短时傅里叶变换后进行频率合成,提取滤波器组特征及其一阶、二阶差分特征,构成能获取病理语音有效局部特征的 LMD-GFbank 图谱特征;其次,为了进一步优化网络模型在训练过程中遗漏掉部分有效特征信息,提出一种三路病理语音识别模型;最后,结合语音特征信息进行病理语音识别模型训练和测试。实验结果表明,LMD-GFbank 图谱特征在三路病理语音识别模型上的识别率达到了 93.36%,优于传统 MFCC、GFCC、Fbank 特征的语音识别效果,验证了所提算法及识别模型能提升病理语音识别准确率。

关键词:发音障碍;局部均值分解;病理语音识别;特征提取
中图分类号:TN912.34;R741 **文献标识码**:A **国家标准学科分类代码**:510.4040

Three channel pathological speech recognition based on LMD improved feature extraction

Zhang Nan Chen Yuanyuan Chen Xinyu Hou Yitao
(School of Information and Communication Engineering, North University of China, Taiyuan 030024, China)

Abstract: Aiming at the problem that patients with dysphonia lack clear and accurate pronunciation, which leads to low pathological speech recognition rate, an improved Gammatone Filter Bank map feature extraction algorithm based on LMD is proposed for three channel pathological speech recognition. Firstly, the algorithm uses LMD to decompose speech signals, performs short-time Fourier transform on each decomposed speech component, and synthesizes frequency to extract filter bank features and their first-order and second-order differential features, forming LMD-GFbank map features that can obtain effective local features of pathological speech. Secondly, in order to further improve the problem that the network model will miss some effective feature information during the training process, a three-way pathological speech recognition model is proposed. Finally, the pathological speech recognition model is trained and tested by combining the speech feature information. The experimental results show that the recognition rate of LMD-GFbank map features on the three channel pathological speech recognition model reaches 93.36%, which is better than the speech recognition performance of traditional MFCC, GFCC, and Fbank features, and verified that the proposed algorithm and recognition model can improve the accuracy of pathological speech recognition.

Keywords: pronunciation disorders; local mean decomposition; pathological speech recognition; feature extraction

0 引 言

发音障碍患者发出的病理语音是由发音相关器官或神经病变所导致的语音失真,近年来,随着科学的进步和社会的发展,人们对发音障碍患者的关注度在逐步提高,针对病理语音识别相应的研究正在逐年增加[1-4]。寻找适用于不同发音障碍语音特征参数,再结合神经网络模型良好的特征学习能力,进一步提升患者病理语音识别准确率。Zhang 等[5]提出一种采用 Bark 滤波器组来提取病理语音信号非线性特征信息的识别方法,但该特征的计算相关维数较复杂、消耗时间。Ye 等[6]提出一种混合病理语音识别模型,提取语音信号的梅尔倒谱系数(mel frequency cepstral coefficents,MFCC)作为输入特征,但该模型对语音图谱特征的识别效果不佳,对提取更深层次的语音特征

收稿日期:2024-05-23
*基金项目:山西省基础研究计划项目(202203021221103)资助

信息还有待提高。Mariya 等[7]采用多分辨率提取病理语音信号特征参数和虚拟话筒阵列合成,进一步提升了语音识别效果。石宇[8]采用深度学习方法搭建神经网络识别模型,来学习病理语音信号特征信息,但未能很有效地提取语音信号深层次的特征信息。季薇等[9]采用部分遮掩的方法提取病理语音信号的语谱图特征,来根据语音信号判别是否具有帕森金病症,但未能识别出患者的语音。国内外研究者针对患者病理语音识别相关技术研究还存在一定的问题,不能提取患者病理语音信号差异性的有效特征信息及搭建性能好的病理语音识别模型,导致病理语音识别效果不佳。

为了进一步提升发音障碍患者的病理语音识别准确率,首先,采用局部均值分解(local mean decomposition,LMD)方法优化语音特征参数,提出一种基于 LMD 的 Gammatone 滤波器组(local mean decomposition gammatone filter banks,LMD-GFbank)图谱特征提取算法,该算法能够有效地提取语音特征信息,更加精确的分析语音信号;其次,为了进一步优化神经网络训练模型,提出一种基于深度可分离卷积(depthwise separable convolution,DSC)的三路病理语音识别模型,该模型能够充分地学习语音信号的深层特征信息;最后,设计不同的病理语音特征在不同的病理语音识别模型上进行训练和测试语音识别效果的对比实验,验证本文所提算法及模型能够进一步提高病理语音识别率,能够进一步将病理语音识别技术与现代医疗技术进行相互融合、协同发展。

1　基于 LMD 的 Gammatone 滤波器组图谱提取算法

LMD 方法能够自适应地分解非平稳的病理语音信号,得到若干个乘积函数(product function, PF),从而得到原始病理语音信号完整的时频特征,它在语音信号分解及冗余消除方面表现出一定的优越性,为了能够有效地提取病理语音信号的有效局部特征,从而更加精确的分析语音信号,因此,采用 LMD 对声学特征进行改进,提出一种基于 LMD 的 LMD-GFbank 图谱特征提取算法,该算法采用 LMD 方法对病理语音信号进行分解,对分解后的 K 个 ***PF*** 分量分别进行短时傅里叶变换,将变换后的结果进行频率合成,送入 Gammatone 滤波器组滤波后进行非线性响度变换后,分别提取其一阶、二阶差分,再将各帧特征拼接在一起构成新特征图谱。LMD-GFbank 图谱提取过程如图 1 所示。

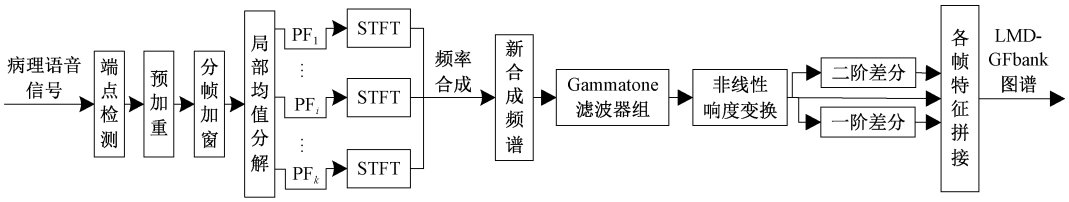

图 1　LMD-GFbank 图谱提取流程图

Fig. 1　LMD-GFbank graph extraction flowchart

LMD 方法[10]在分解开始时,首先找出语音信号的所有局部极大、极小值点,再通过滑动平均的方式来求取语音信号对应的局部均值和包络估计函数,然后去除掉语音信号的局部均值函数,一直重复进行包络估计函数的解调,获取纯调频函数之后就终止整个循环迭代过程,将整个过程中获得的包络估计函数的乘积当成包络函数,再与纯调频函数相乘后获取第一阶 ***PF*** 分量,不断重复迭代此过程,以此获取分解语音信号的其他阶 ***PF*** 分量。

LMD-GFbank 图谱具体提取流程如下:

1)发音障碍患者的病理语音信号进行前期预处理操作,端点检测剔除语音信号的空白和噪音,从而获取语音信号的有效内容;预加重可以在一定程度上增强语音信号的高频信息,获取平缓的语音信号频谱;分帧加窗操作在 10～30 ms 内分割语音信号,获取到近似为平稳的每一帧语音信息,使得整体的病理语音信号更加连贯。

2)经过预处理后,采用 LMD 方法对病理语音信号进行分解,假设原始语音信号为 $x(t)$ 可以分解为 k 个 $\boldsymbol{PF}_p(t)$ 和残余信号 $u_k(t)$,可表示如下:

$$x(t)=\sum_{p=1}^{k}\boldsymbol{PF}_p(t)+u_k(t) \tag{1}$$

其中,$\boldsymbol{PF}_p(t)$ 为第 p 个 ***PF*** 分量,t 代表帧同步时间,k 代表分解循环次数。

3)对分解后的 ***PF*** 分量进行短时傅里叶变换(short timefourier transform,STFT)得到病理语音信号频域信息 $P(n)$,再对 $P(n)$ 进行平方计算得到频域上能量信息 $|P_p(n)|^2$。

4)对每个 ***PF*** 分量在频域上能量信息进行合成,得到新的合成病理语音信号频谱信息 $E(n)$,可表示如下:

$$E(n)=\sum_{p=1}^{k}|P_p(n)|^2 \tag{2}$$

5)采用 Gammatone 滤波器组[11]对 $S(n)$ 进行滤波后,得到频域上的能量谱,滤波器组的时域脉冲响应 $g(t)$ 表示如下:

$$g_i(t)=At^{n-1}\cos(2\pi f_i t+\phi_i)u(t)\times e^{\frac{-49.4\pi(4.37f_i+1)t}{1\,000}},\quad t\geqslant 0,\ 1\leqslant i\leqslant N \tag{3}$$

其中,A 是滤波器增益,$u(t)$ 是阶跃函数,f_i 是滤波器中心频率,ϕ_i 是相位,n 是滤波器阶数,N 是滤波器个

数，i 表示第几个滤波器，t 是帧同步时间。

6)对经过步骤 5)得到的能量谱，再进行非线性响度变换[12]，模拟人耳听觉特性在不同频率段下计算不同的幂指数，计算公式如下：

$$G(m)=\left(\sum_{k=0}^{N-1}H_G(k)E(k)\right)^{\beta},0\leqslant m<N \tag{4}$$

其中，$H_G(k)$ 表示经过第 k 个 Gammatone 滤波器对语音信号滤波，N 表示滤波器个数，m 表示第几个滤波器，β 是幂指数值，在不同频率段下的计算公式如下：

$$\beta=\begin{cases}(f-29\ 000)/10\ 000, & 0\leqslant f<1\times10^3\\(f-58\ 000)/20\ 000, & 1\times10^3\leqslant f<2\times10^3\\(f-97\ 000)/30\ 000, & 2\times10^3\leqslant f<3\times10^3\\(f-116\ 000)/40\ 000, & 3\times10^3\leqslant f<4\times10^3\\(f-145\ 000)/50\ 000, & 4\times10^3\leqslant f<5\times10^3\\(f-184\ 000)/60\ 000, & 5\times10^3\leqslant f<6\times10^3\\(f-203\ 000)/70\ 000, & 6\times10^3\leqslant f<7\times10^3\\(f-232\ 000)/80\ 000, & 7\times10^3\leqslant f<8\times10^3\\1/3, & f\geqslant8\times10^3\end{cases} \tag{5}$$

其中，f 是 Gammatone 滤波器组的频率。

7)对经过步骤 6)非线性响度变换后的能量谱，进行一阶差分[13]处理，对一阶差分特征再进行差分处理得到二阶差分特征，引入一阶、二阶差分方法能够有效地分析语音信号各帧之间的动态连续性，提取到语音信号各帧之间动态特征信息，同时能有效地提取语音信号各帧之间时变信息以及相邻三帧之间特征信息的关联性，差分特征 ΔG 可表示如下：

$$\Delta G_1(j,d)=\frac{\sum_{z=1}^{2}z(G(j+z,d)-G(j-z,d))}{\sqrt{2\sum_{z=1}^{2}z^2}} \tag{6}$$

$$\Delta G_2(j,d)=\frac{\sum_{z=1}^{2}z(\Delta G_1(j+z,d)-\Delta G_1(j-z,d))}{\sqrt{2\sum_{z=1}^{2}z^2}} \tag{7}$$

其中，$G(j,d)$ 表示非线性响度变换后的能量谱，$\Delta G_1(j,d)$ 表示非线性响度变换后的能量谱的一阶差分特征，$\Delta G_2(j,d)$ 非线性响度变换后的能量谱的二阶差分特征，j 表示语音信号经过加窗分帧后第 j 帧，d 表示语音特征维度。

8)将得到第 j 帧非线性响度变换后的能量谱特征及其一阶差分特征、二阶差分特征进行拼接，得到语音信号第 j 帧组合特征 $C(j,3d)$，可表示如下：

$$C(j,3d)=\{G(j,d);\Delta G_1(j,d);\Delta G_2(j,d)\} \tag{8}$$

9)将经过步骤 8)所得每帧组合特征再进行拼接，最终得到病理语音信号的 LMD-GFbank 图谱特征信息。

2　基于 DSC 的三路病理语音识别网络模型

近年来，针对单路神经网络模型在深度方向进行不断地优化，但在学习特征过程中还是不够充分，遗漏掉一些有效特征信息，由此，为了改善单路语音识别网络模型面临的问题，提出一种基于 DSC 的三路病理语音识别模型，该模型采用三条支路分别对病理语音特征信息进行学习，为了解决搭建的网络层数过深时，最终的识别效果反而变差，由此添加改进的残差单元块在模型中以解决此问题；再采用 DSC 对网络模型进行优化，能够有效降低网络训练模型的参数和卷积所需的乘法运算量。该模型其中一条支路为单路 DSC，采用卷积层、DSC 层、池化层进行串联对病理语音特征信息进行不断学习，另两条支路分别为单路残差网络＋DSC，在前部分先使用卷积层、池化层进行串联，然后在后部分添加两个改进后的残差单元块结合三个 DSC 层不断学习语音特征信息，再串联池化层降低特征维度；采用 Concat()函数将三条支路学习的语音特征信息进行合并后，送入全连接层继续学习，最后，使用联结时序分类(connectionist temporal classification，CTC)算法构建基于 DSC 的三路病理语音识别模型的结构，如图 2 所示。

DSC[14]要经过两个独立内核进行处理，首先，采用尺寸 3＊3 深度卷积核对不同通道上语音特征特征信息执行卷积运算以提取空间维度上的特征参数，再采用尺寸 1＊1 传统卷积核对深度卷积核的输出迭代每个点，沿着不同通道维度上加权合并特征参数。DSC 层可以有效降低语音

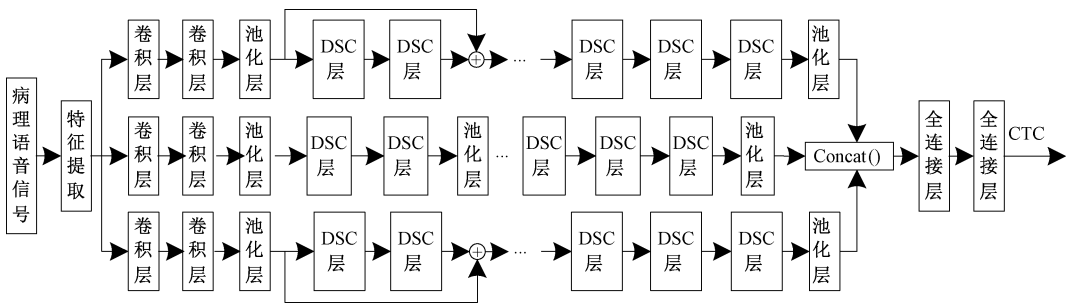

图 2　基于 DSC 的三路病理语音识别网络模型

Fig. 2　A three channel pathological speech recognition network model based on DSC

识别训练模型的参数和卷积所需的乘法运算量,而且输入数据量越大,使得参数量和计算成本降低得越多。深度可分离卷积结构如图 3 所示。

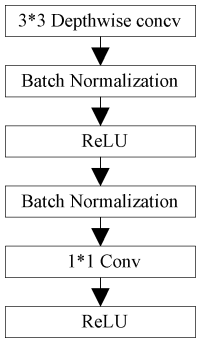

图 3　深度可分离卷积结构

Fig. 3　Depthwise separable convolutional structure

在残差单元块结构[15-16]中,一是添加了一条跳跃连接,使得模型单元可以不用学习整个模型单元输出特征,而是直接学习上个单元输出特征,能够减少特征信息在各单元间传递造成有效特征参数丢失、梯度消失与爆炸的问题;二是添加了两个 DSC 层,能够降低训练模型中语音特征的参数量和卷积运算过程的复杂程度,这样会大幅度降低训练模型的计算成本,在一定程度上加快了网络模型的训练速度,提升病理语音识别模型的泛化能力,从而进一步提高模型的语音识别准确率。引入 DSC 层来改进残差单元块结构,其中含有 DSC 层、BN 层、ReLU 激活函数、池化层以及一个跳跃连接构成,如图 4 所示。

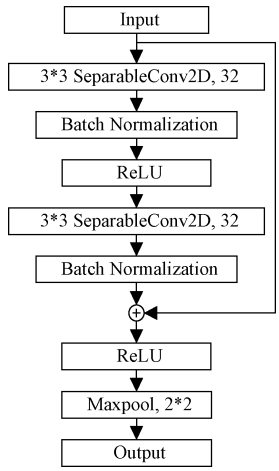

图 4　改进的残差单元块结构

Fig. 4　Improved residual unit block structure

3　病理语音数据集

本实验采用公开 UA-Speech 语料库[17]中 26 种孤立单词的发音样本如表 1 所示,15 名患者的语音清晰度如表 2 所示,语音清晰等级分为 4 种(非常低、低、中等、高)。实验总共使用了 5 694 个病理语音样本,进行识别 26 种孤立单词语音,每一种包含 219 个语音样本。其中 4 640 个语音样本用于基于 DSC 的三路病理语音识别模型训练,其中男性患者语音样本 910 个,女性患者语音样本 220 个;1 130 个语音样本用于网络模型测试;572 个语音样本用于网络模型验证。实验过程中使用的每个语音样本只能出现在划分的训练、测试、验证数据集其中的一个。

表 1　病理语音患者发音语料表

Table 1　Pronunciation corpus of pathological speech patients

语料名称	发音项	合计
国际广播字母表	ALPHA,BRAVO,CHARLIE,DELTA,ECHO,FOXTROT,GOLF,HOTEL,INDIA,JULIET,KILO,LIMA,MIKE,NOVEMBER,OSCAR,PAPA,QUEBEC,ROMEO,SIERRA,TANGO,UNIFORM,VICTOR,WHISKEY,X-RAY,YANKEE	26

表 2　不同患者病理语音清晰度

Table 2　The clarity of pathological speech in different patients

病理语音患者	M4	F3	M12	M1	M7	F2	M6	M16	M5	F4	M11	M9	M14	M8	M10
语音清晰度/%	2	6	7	15	28	29	39	43	58	62	62	86	90	93	93

4　实验结果与分析

4.1　实验准备

在病理语音数据集上进行病理语音识别实验,分为训练和测试两个阶段,训练阶段对训练数据集中病理语音信号提取 LMD-GFbank 图谱特征,然后送入三路病理语音识别模型中进行训练,获取最佳的病理语音识别模型参数;测试阶段对测试数据集中病理语音信号提取 LMD-GFbank 图谱特征,然后在训练好的病理语音识别模型上进行语音识别,得到模型测试的语音识别效果,以此验证本文所提特征提取算法的有效性及搭建的三路病理语音识别模型具有一定的泛化性。

4.2　实验参数及评价指标

实验过程采用 Python 中的深度学习框架来搭建基于

DSC 的三路病理语音识别模型，其中单路 DSC 支路包含 2 个尺寸 3 * 3 卷积层、5 个尺寸 3 * 3DSC 层、3 个尺寸 2 * 2 池化层；单路残差网络＋DSC 包含 2 个尺寸 3 * 3 卷积层，3 个尺寸 3 * 3DSC 层，2 个改进的残差单元块，2 个尺寸 2 * 2 池化层；其中残差单元块结构含有 2 个尺寸 3 * 3DSC 层，1 个尺寸 2 * 2 池化层及 1 个跳跃连接，同时使用 ReLU 激活函数、Adam 优化、softmax 函数及 BN 层[18]来搭建三路病理语音识别模型。

实验中使用单词识别准确率[19]（word recognition accuracy，WRA）对发音障碍患者病理语音识别效果进行评价，可表示如下：

$$WRA = \frac{\text{正确识别病理语音样本个数}}{\text{实验总共使用病理语音样本个数}} \times 100\% \tag{9}$$

4.3　实验结果及分析

在病理语音数据集上进行语音识别实验，使用不同的病理语音特征参数在基于 DSC 的三路病理语音识别模型上进行训练和测试，不同输入特征的识别效果对比如表 3 所示；使用 LMD-GFbank 图谱特征参数在不同的病理语音识别网络模型上进行训练和测试，不同语音识别网络模型的识别效果对比如表 4 所示。

表 3　不同输入特征下的识别效果

Table 3　Recognition performance under different input features

输入特征	WRA/%		
	男患者	女患者	总体
MFCC 参数	79.67	75.00	78.76
GFCC 参数	81.32	75.45	80.18
传统 Fbank 特征	87.25	83.18	86.46
LMD-Gfbank 剔除差分	92.09	90.45	91.77
LMD-GFbank 图谱特征	94.18	90.00	93.36

表 4　不同语音识别模型下的识别效果

Table 4　Recognition performance under different speech recognition models

语音识别模型	WRA/%		
	男患者	女患者	总体
单路 DSC 识别模型	87.80	84.55	87.17
单路残差网络＋DSC 识别模型	90.77	87.73	90.17
双路残差网络＋DSC 识别模型	92.53	90.91	92.21
基于 DSC 的三路病理语音识别模型	94.18	90.00	93.36

观察表 3 可知，LMD-GFbank 图谱特征在基于 DSC 的三路病理语音识别模型上的病理语音总体识别率已经达到 93.36%，针对男性、女性发音障碍患者的识别率达到 94.18%、90.00%。LMD-GFbank 图谱特征的语音识别效果最好，相比 MFCC 参数、伽玛通频率倒谱系数（gammatone filter cepstral coefficient，GFCC）参数、传统滤波器组（filter banks，Fbank）特征和 LMD-GFbank 剔除差分特征的总体语音识别率分别提升了 14.60%、13.18%、6.90%和 1.59%，针对男性患者的识别率分别提升了 14.51%、12.86%、6.93%和 2.09%。LMD-GFbank 剔除差分部分特征相比 GFCC 参数、传统 Fbank 特征的总体语音识别率分别提升了 11.59%、13.18%，针对男性患者的识别率分别提升了 14.60%、5.31%，针对女性患者的识别率达到 15.00%、7.27%，由于引入 LMD 方法分析病理语音不同频带上的语音特征信息，能够有效地提取病理语音信号的有效局部特征，从而更加精确的分析语音信号。LMD-GFbank 图谱特征相比 LMD-GFbank 剔除差分特征的总体语音识别率提升了 1.86%，针对男性患者的识别率提升了 2.09%，由于引入一阶、二阶差分方法分析语音信号各帧之间的动态连续性，能有效地提取语音信号各帧之间时变信息以及相邻三帧之间信息的关联性。

观察表 4 可知，LMD-GFbank 图谱特征在基于 DSC 的三路病理语音识别模型、双路残差网络＋DSC 识别模型、单路残差网络＋DSC 识别模型和单路 DSC 识别模型上的病理语音总体识别率分别达到 93.36%、92.21%、90.17%和 87.17%，针对男性患者的识别率分别达到 94.18%、92.53%、90.77%和 87.80%，针对女性患者的识别率分别达到 90.00%、90.91%、87.73%和 84.55%。基于 DSC 的三路病理语音识别模型的语音识别效果达到了最佳，相比双路残差网络＋DSC 识别模型、单路残差网络＋DSC 识别模型和单路 DSC 识别模型上的总体语音识别率分别提升了 1.15%、3.19%和 6.19%，针对男性患者的识别率分别提升了 1.65%、3.41%和 6.38%，由于引入 DSC 能降低模型中的病理语音特征参数量和运算过程的复杂程度；引入残差单元块能够通过残差学习去解决神经网络退化问题，解决搭建深层网络导致语音识别模型效果变差的问题；引入三条支路来学习语音特征信息，能够更加充分地学习语音信号的深层特征，弥补在学习特征信息中遗漏掉部分有效的特征，表现出最佳的语音识别率。

图 5 表示不同发音障碍患者的 LMD-GFbank 图谱特征在双路残差网络＋DSC 识别模型上的语音识别效果对比，观察图 5 可知，男性 M10 患者的语音清晰度（93%）最高的语音识别率达到了 97.22%，M04 患者的语音清晰度（2%）最低的语音识别率达到了 85.53%，M10 患者相比 M04 患者的语音识别率高出了 11.69%；针对女性 F04 患者的语音清晰度（62%）最高的语音识别率达到了 92.42%，F03 患者的语音清晰度（6%）最低的语音识别率达到了 87.34%，F04 患者相比 F03 患者的语音识别率高出了 5.08%，不同发音障碍患者的语音清晰度越高得到语音识别率也越高。

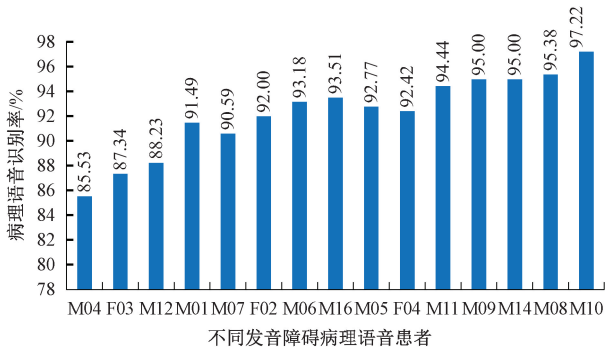

图 5　不同患者 LMD-GFbank 图谱特征在双路残差网络+DSC 模型上的识别效果

Fig. 5　Recognition performance of LMD-GFbank map features of different patients on dual residual network+DSC model

图 6 表示不同发音障碍患者的 LMD-GFbank 剔除差分特征在基于 DSC 的三路病理语音识别模型上的语音识别效果对比,观察图 6 可知,男性、女性发音障碍患者的语音清晰度越高得到语音识别率也越高,针对男性 M10 患者的语音清晰度(93%)最高的语音识别率达到了 95.83%,M04 患者的语音清晰度(2%)最低的语音识别率达到了 84.21%,M10 患者相比 M04 患者的语音识别率高出了 11.62%;针对女性 F04 患者的语音清晰度(62%)最高的语音识别率达到了 93.94%,F03 患者的语音清晰度(6%)最低的语音识别率达到了 88.61%,F04 患者相比 F03 患者的语音识别率高出了 5.33%。

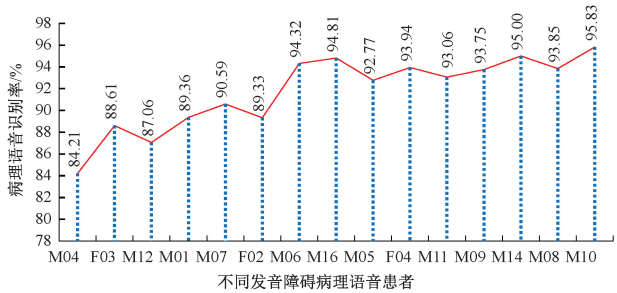

图 6　不同患者在 LMD-GFbank 剔除差分特征下的识别效果

Fig. 6　Recognition performance of different patients in LMD-GFbank with the removal of differential features

图 7 表示不同发音障碍患者的 LMD-GFbank 图谱特征在基于 DSC 的三路病理语音识别模型上的语音识别效果对比,观察图 7 可知,男性、女性发音障碍患者的语音清晰度越高得到语音识别率也越高,针对男性 M10 患者的语音清晰度(93%)最高的语音识别率达到了 100.00%,M04 患者的语音清晰度(2%)最低的语音识别率达到了 85.53%,M10 患者相比 M04 患者的语音识别率高出了 14.47%;针对女性 F04 患者的语音清晰度(62%)最高的语音识别率达到了 92.42%,F03 患者的语音清晰度(6%)最低的语音识别率达到了 87.34%,F04 患者相比 F03 患者的语音识别率高出了 5.08%。

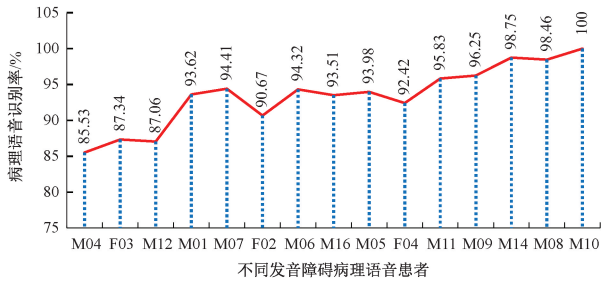

图 7　不同发音障碍患者在 LMD-GFbank 图谱特征下的识别效果

Fig. 7　Recognition performance of patients with different speech disorders under LMD-GFbank map features

表 5 表示不同语音清晰度等级发音障碍患者在不同方法上的语音识别效果对比,观察表 5 可知,方法 1、方法 2 上的发音障碍患者的语音清晰度等级越高得到总体语音识别率也越高,在患者语音清晰度等级高时,方法 1、方法 2 上的总体语音识别率分别达到了 98.31%、94.61%;在患者语音清晰度等级中等时,方法 1、方法 2 上的总体语音识别率分别达到了 94.12%、93.21%,针对男性患者的语音识别率分别达到了 94.84%、92.90%,针对女性患者的语音识别率分别达到了 92.42%、93.94%;在患者语音清晰度等级非常低时,方法 1、方法 2 上的总体语音识别率分别达到了 87.80%、87.11%,针对男性患者的语音识别率分别达到了 87.98%、86.54%,针对女性患者的语音识别率分别达到了 87.34%、88.61%。对于方法 1、方法 2,患者语音清晰度等级高相比患者语音清晰度等级非常低的总体语音识别率分别高出了 10.51%、7.50%,针对男性患者的语音识别率分别高出了 10.51%、7.50%;患者语音清晰度等级中等相比患者语音清晰度等级非常低的总体语音识别率分别高出了 6.32%、6.10%,针对男性患者的语音识别率分别高出了 6.86%、6.36%,针对女性患者的语音识别率分别高出了 5.08%、5.33%。

通过分析表 3～5,图 5～7 可以得出,LMD-GFbank 图谱特征在基于 DSC 的三路病理语音识别模型上表现出最佳的发音障碍患者病理语音识别效果。LMD-GFbank 图谱特征能够有效地提取病理语音信号的有效局部特征,更加精确的分析语音信号,同时有效地提取到语音信号各帧之间动态特征信息以及相邻三帧之间特征信息的关联性;基于 DSC 的三路病理语音识别模型能降低模型中的病理语音特征参数量和运算过程的复杂程度,同时能够更加充分地学习语音信号的深层特征,弥补在学习特征信息中遗漏掉部分有效的特征,表现出最佳的病理语音识别准确率。

表 5　不同语音清晰等级患者在不同方法下的识别效果

Table 5　The recognition performance of patients with different levels of speech clarity under different methods

方法	清晰度等级/%	WRA/%		
		男性患者	女性患者	总体
LMD_GFbank 剔除差分部分＋三路病理语音识别模型(方法 1)	非常低(0～25)	86.54	88.61	87.11
	低 25～50)	93.20	89.33	92.31
	中等(50～75)	92.90	93.94	93.21
	高(75～100)	94.61	—	94.61
LMD_GFbank＋三路病理语音识别模型(方法 2)	非常低(0～25)	87.98	87.34	87.80
	低 25～50)	94.00	90.67	93.23
	中等(50～75)	94.84	92.42	94.12
	高(75～100)	98.31	—	98.31

5　结　　论

针对发音障碍病理语音识别准确率低的问题，从时频特征参数着手，基于 LMD 的 Gammatone 滤波器组图谱特征提取算法，该算法提取的 LMD-GFbank 图谱特征能够有效地提取病理语音信号特征参数，更加充分地分析病理语音信号特征信息；为了提升发音障碍患者病理语音识别率，改善单路神经网络语音识别模型，提出一种三路病理语音识别模型，该模型能够减少特征信息在各模型单元间传递造成有效特征参数的丢失，同时，能够降低训练模型的参数和卷积所需的乘法运算量。在公开 UA-Speech 数据集上设计病理语音识别实验对比，实验过程中使用不同的病理语音特征参数在不同的神经网络模型上进行训练和测试。实验结果表明，文中所提特征提取算法结合语音识别模型，使得发音障碍病理语音识别率达到了 93.36%，优于一些经典特音征参数的语识别效果，更加充分地反映病理语音信号的深层特征，能表现出最佳的语音识别率。在往后的研究中，不断优化发音障碍患者病理语音差异性的有效特征参数提取算法，同时不断改进病理语音识别神经网络模型搭建，学习病理语音信号更加深层次的特征信息，更好地提升病理语音识别准确率。

参考文献

[1] SAMEH S, RIMAH A, BEN S Y. A robust pathological voices recognition system based on DCNN and scattering transform [J]. Applied Acoustics, 2021, 177(6): 107847-107854.

[2] REVATHI A, NAGAKRISHNAN R, SASIKALADEVI N. Comparative analysis of dysarthric speech recognition: Multiple features and robust templates[J]. Multimedia Tools Appl, 2022, 81(22): 31245-31259.

[3] SAGHIRI M C, SAGHIRI A, SAMADI E. A mini-review of pathological voice recognition[J]. Advances in Human Biology, 2023, 13(1): 17-22.

[4] 宋伟，张杨豪. 构音障碍语音识别算法研究综述[J]. 计算机工程与应用，2024，60(11)：62-74.
SONG W, ZHANG Y H. Survey of specific speech recognition algorithms for dysarthria[J]. Computer Engineering and Applications, 2024, 60(11): 62-74.

[5] ZHANG X J, ZHU X C, WU D, et al. Nonlinear features of bark wavelet sub-band filtering for pathological voice recognition [J]. Engineering Letters, 2021, 29(1): 1-12.

[6] YE W, JIANG Z, LI Q, et al. A hybrid model for pathological voice recognition of post-stroke dysarthria by using 1DCNN and double-LSTM networks [J]. Applied Acoustics, 2022, 197(8): 108924-108934.

[7] MARIYA C T A. VIJAYALAKSHMI P, NAGARAJAN T. Data augmentation techniques for transfer learning-based continuous dysarthric speech recognition [J]. Circuits, Systems, and Signal Processing, 2023, 42(1): 601-622.

[8] 石宇. 不定长病理语音的特征提取与识别研究[D]. 西安：陕西师范大学，2022.
SHI Y. Research on feature extraction and recognition of indefinite-length pathological speech[D]. Xi′an: Shaanxi Normal University, 2022.

[9] 季薇，杨茗淇，李云，等. 基于掩蔽自监督语音特征提取的帕金森病检测方法[J]. 电子与信息学报，2023，45(10)：3502-3510.
JI W, YANG M Q, LI Y, et al. Parkinson′s disease detection method based on masked self-supervised speech feature extraction[J]. Journal of Electronics & Information Technology, 2023, 45(10): 3502-3510.

[10] CHAO L, TING J, SHENG W, et al. Single-Channel speech enhancement based on adaptive low-rank matrix decomposition [J]. IEEE Access, 2020, 8: 37066-37076.

[11] 李炜，刘禹，李立刚，等. 基于自适应降噪的柱塞泵

故障音频特征提取方法[J]. 国外电子测量技术，2023，42(1)：1-6.
LI W，LIU Y，LI L G，et al. Audio feature extraction method for plunger pump fault based on adaptive noise reduction [J]. Foreign Electronic Measurement Technology，2023，42(1)：1-6.

[12] 龙华，黄张衡，邵玉斌，等. 基于改进 CFCC 特征提取的语种识别算法研究[J]. 通信学报，2022，43(12)：211-221.
LONG H，HUANG ZH H，SHAO Y B，et al. Research on language recognition algorithm based on improved CFCC feature extraction[J]. Journal on Communications，2022，43(12)：211-221.

[13] 赵建星，薛珮芸，白静，等. 一种用于构音障碍语音识别的多尺度特征提取算法[J]. 生物医学工程学杂志，2023，40(1)：44-50.
ZHAO J X，XUE P Y，BAI J，et al. A multiscale feature extraction algorithm for dysarthric speech recognition[J]. Journal of Biomedical Engineering，2023，40(1)：44-50.

[14] CHEN Y H，ZHANG S B. A helium speech unscrambling algorithm based on deep learning[J]. Information，2023，14(3)：171-189.

[15] SIDDHANT G，ANKUR T P，MIRALI P，et al. Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments[J]. Neural Networks，2021，139(1)：105-117.

[16] 任健，李鸿燕，张昱，等. 基于 UNet 自适应特征融合的语音增强[J]. 电子测量技术，2022，45(9)：76-81.
REN J，LI H Y，ZHANG Y，et al. Speech enhancement based on UNet adaptive feature fusion[J]. Electronic Measurement Technology，2022，45(9)：76-81.

[17] MOHAMMED S Y，SID-AHMED S，BRAHIM-FARES Z，et al. Improving dysarthric speech recognition using empirical mode decomposition and convolutional neural network[J]. Eurasip J Audio Spee，2020(1)：1-7.

[18] 张小恒，李勇明，王品. 双阶段帕金森病语音聚类包络卷积稀疏迁移学习算法[J]. 仪器仪表学报，2022，43(11)：151-161.
ZHANG X H，LI Y M，WANG P. Two-stage PD speech clustering envelope and convolution sparse transfer learning algorithm[J]. Chinese Journal of Scientific Instrument，2022，43(11)：151-161.

[19] RAJESWARI N，CHANDRAKALA S. Generative model-driven feature learning for dysarthric speech recognition[J]. Biocybern Biomed Eng，2016，36(4)：553-561.

作者简介

张楠，硕士研究生，主要研究方向为语音信号处理、病理语音识别。
E-mail：1412237081@qq.com

陈媛媛(通信作者)，博士，教授，博士生导师，主要研究方向为信号与信息处理、图像多维信息处理。
E-mail：chenyy@nuc.edu.cn

陈鑫钰，硕士研究生，主要研究方向为计算机视觉，图像信息处理。
E-mail：1483652827@qq.com

侯懿桃，硕士研究生，主要研究方向为声学信号处理。
E-mail：2389824055@qq.com