计算机科学
COMPUTER SCIENCE

构音障碍说话人自适应研究进展及展望

康新晨, 董雪燕, 姚登峰, 钟经华

引用本文

康新晨, 董雪燕, 姚登峰, 钟经华. 构音障碍说话人自适应研究进展及展望[J]. 计算机科学, 2024, 51(8): 11-19.

KANG Xinchen, DONG Xueyan, YAO Dengfeng, ZHONG Jinghua. Advancements and Prospects in Dysarthria Speaker Adaptation [J]. Computer Science, 2024, 51(8): 11-19.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230700161

构音障碍说话人自适应研究进展及展望

康新晨[1]　董雪燕[1]　姚登峰[1,2,3]　钟经华[1]
1 北京联合大学北京市信息服务工程重点实验室　北京 100101
2 清华大学人文学院计算语言学实验室　北京 100084
3 清华大学心理学与认知科学研究中心　北京 100084
(kxc4088@163.com)

摘　要　自动化语音识别工具让构音障碍者和正常人的沟通变得顺畅,因此,近年来构音障碍语音识别成为了一项热门研究。构音障碍语音识别的研究包括:收集构音障碍者和正常人的发音数据,对构音障碍者和正常人的语音进行声学特征表示,利用机器学习模型比较和识别发音的内容并定位出差异性,以帮助构音障碍者改善发音。然而,由于收集构音障碍者的大量语音数据非常困难,且构音障碍者存在发音的强变异性,导致通用语音识别模型的效果往往不佳。为了解决这一问题,许多研究提出将说话人自适应方法引入构音障碍语音识别。对大量相关文献进行调研发现,当前此类研究主要围绕特征域和模型域对构音障碍语音进行分析。文中重点分析特征变换和辅助特征如何解决语音特征的差异性表示,以及声学模型的线性变换、微调声学模型参数和基于数据选择的域自适应方法如何提高模型识别的准确率。最后总结出构音障碍说话人自适应研究当前遇到的问题,并指出未来的研究可以从语音变异性的分析、多特征多模态数据的融合以及基于小数量的自适应方法的角度,提升构音障碍语音识别模型的有效性。

关键词:构音障碍;说话人自适应;辅助特征;变换;微调;域自适应
中图分类号　TP183

Advancements and Prospects in Dysarthria Speaker Adaptation

KANG Xinchen[1],DONG Xueyan[1],YAO Dengfeng[1,2,3] and ZHONG Jinghua[1]
1 Beijing Key Laboratory of Information Service Engineering,Beijing Union University,Beijing 100101,China
2 Lab of Computational Linguistics,School of Humanities,Tsinghua University,Beijing 100084,China
3 Center for Psychology and Cognitive Science,Tsinghua University,Beijing 100084,China

Abstract　Automatic speech recognition tools make communication between dysarthria and normal individuals smoother,therefore,dysarthric speech recognition has become a hot research topic in recent years. The research on dysarthric speech recognition includes:collecting pronunciation data from dysarthria and normal individuals,representing acoustic features of dysarthria speech and normal speech,comparing and recognizing the content of pronunciation by machine learning model,and locating differences,so as to help dysarthria to improve their pronunciation. However,due to the significant difficulties in collecting a large amount of speech data from dysarthria,and the strong variability of their pronunciation,the performance of universal speech recognition models is often poor. To address this issue,many studies have proposed to introduce speaker adaptation methods into dysarthric speech recognition. Through extensive research on relevant literature,it has been found that current research mainly focuses on analyzing dysarthria speech in the feature domain and model domain. This paper focuses on analyzing how feature transformation and auxiliary features solve the differential representation of speech features,how linear transformation of acoustic models,fine-tuning of acoustic model parameters,and domain adaptation methods based on data selection improve the accuracy of model recognition. Finally,the current problems encountered in the research of dysarthria speaker adaptation are summarized,and it is pointed out that future research can improve the effectiveness of dysarthric speech recognition models from the perspectives of analyzing speech variability,fusing multi-feature and multi-modal data,and using a small number of speaker adaptation methods.

到稿日期:2023-07-20　返修日期:2023-12-02
基金项目:北京市自然科学基金(4202028);国家语言文字工作委员会项目(YB145-25);国家自然科学基金(62036001);国家社会科学基金(21BYY106,21&ZD292);2019 年度北京市教育委员会科技一般项目(KM201911417005)
This work was supported by the Natural Science Foundation of Beijing,China(4202028),General Project of the National Language Committee(YB145-25),National Natural Science Foundation of China (62036001),National Social Science Foundation of China(21BYY106,21&ZD292) and 2019 Science and Technology Plan of Beijing Municipal Education Commission(KM201911417005).
通信作者:董雪燕(tjtxueyan@buu.edu.cn)

Keywords Dysarthria, Speaker adaptation, Auxiliary features, Transformation, Fine-tuning, Domain adaptation

1 引言

构音障碍是因与言语相关的神经病变引起的一种言语表达障碍。构音障碍者在口语表达过程中通常存在发音不准确、语言不流畅、语速缓慢以及音量、清晰度较低等情况，其语音具有强变异性的特点，与正常人的语音之间存在较大差异。因此，构音障碍语音识别成为信息无障碍方向的一大难点。

在语音识别中，说话人之间的差异导致语音数据与声学模型不匹配。说话人自适应[1]能够利用说话人数据对说话人无关模型进行调整，解决由说话人差异带来的不匹配问题，从而提高语音识别系统的准确率。构音障碍说话人自适应的研究对于构音障碍语音识别研究具有重要意义。说话人自适应是将一系列语音声学特征向量 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_t,\cdots,x_T\}$ 利用说话人数据得到特征参数 θ^E，从而将声学特征映射到单词序列 W 上[2]。可将模型定义为：

$$y_t=f(x_t;\theta;\theta^E) \tag{1}$$

其中，$f(x_t;\theta;\theta^E)$ 是具有参数 θ 和参数 θ^E 的模型，y_t 是帧 t 的输出标签。

与通用说话人自适应相比，构音障碍说话人自适应更具挑战性，主要体现在以下 3 个方面：

1）构音障碍语音因发音特点与正常语音差异较大，故两者在声学特征向量的概率分布上也不一致，这导致构音障碍语音与说话人无关模型无法很好地匹配。

2）对于构音障碍者而言，发音不准确、语音不流畅等情况都会引起其语音的变异性，并且其语音变异性比正常人更为显著，导致同一构音障碍者在不同情况下针对同一词汇的发音差异更为明显。另外，同一构音障碍者同一发音的语音特征也不相同，从而加剧了自适应声学模型与语音之间的不匹配问题。

3）由于构音障碍者的数量较少，因此构音障碍语音数据的收集相对困难，为构音障碍者构建的说话人自适应模型难以获得足够的训练数据。

构音障碍说话人自适应的关键是利用构音障碍相关语音信息，对声学模型进行修正和调节，以解决其语音差异性与变异性引起的不匹配问题。其自适应思路分为两种：1）加工构音障碍说话人声学特征参数，使之能适用于原有模型；2）利用构音障碍语音进行模型训练，通过改变模型参数来适应构音障碍语音的变异性特征。根据自适应思路中参数调整方式的不同，构音障碍说话人自适应分为基于特征域的自适应和基于模型域的自适应。本文按照两类自适应方法展开。

2 基于特征域的自适应

构音障碍者的发音特点包括发音不准确、语音不流畅、语速慢以及声音的音量和清晰度较低。不同严重程度的构音障碍者在发音上存在差异，即便是同一构音障碍者，对同一词汇的发音也不相同。因此，构音障碍者的发音具有多样性，其多样性表现在构音障碍语音的强变异性特征中。基于特征域的自适应是将构音障碍语音特征进行加工，使得与构音障碍相关的语音变异性特征能够匹配模型。

2.1 特征变换

特征变换的自适应是对特征进行线性变换，将构音障碍语音间的差异进行归一化，使得特征更能体现构音障碍的相关特点。为解决不同构音障碍说话人之间音高变化的问题，Hahm 等[3]通过研究声学及发音空间中说话人归一化方法，来证明特征空间最大似然线性回归（Feature Space Maximum Likelihood Linear Regression，F-MLLR）[4]可以归一化构音障碍说话人的声学及发音空间。在 Hahm 等研究的基础上，Blat 等[5]在构音障碍的自适应训练解码时，将 F-MLLR 方法用在输入特征上，对构音障碍说话人的特征空间进行归一化。其方法在基于 GMM-HMM 与 DNN-HMM 的系统上的词错误率（Word Error Rate，WER）分别为 34.07% 和 32.69%，同时证明基于 DNN-HMM 的系统在识别性能上的优势。

此外，在早期的神经网络模型中，输入向量以及隐藏层输出向量都可以用于进行线性变换。特征变换的参数数量通常较少，变换结构也较为简单。然而，由于构音障碍语音具有强变异性，仅归一化语音差异不足以实现构音障碍的自适应，即使自适应数据增多，自适应效果也无法得到改善。

2.2 辅助特征

辅助特征的自适应通过刻画构音障碍者的语音特征，利用其特征辅助声学模型训练。该方法需将构音障碍语音进行加工处理，分别得到两种特征，即声学特征和辅助特征，声学特征对应构音障碍者的声学信息，辅助特征则对应构音障碍者的身份信息。与说话人身份信息相关的特征以身份向量 i-vector[6] 和瓶颈特征（Bottleneck Features）为代表，说话人相关信息被概括成固定长度的向量，这些特征作为辅助特征辅助对应的声学特征训练模型。Wang 等[7]证明由时延神经网络（Time Delay Neural Network，TDNN）训练的说话人特征向量可以作为辅助特征，以减少构音障碍说话人的可变性。该方法通过压缩语音特征维度得到 x-vectors 来消除高维特征的细微差异，使得辅助特征能够表示构音障碍语音的共性特征。此外，构音障碍语音特征经 F-MLLR 变换归一化后得到 F-MLLR 特征，该特征仅针对语音特征空间变换，也可与其他辅助特征结合进行模型训练。Yilmaz 等[8]通过训练 DNN、CNN 以及时频卷积神经网络（Time-Frequency CNN，TFCNN），将声学特征转换为更为紧凑的瓶颈特征，然后与 F-MLLR 特征相结合形成辅助特征。这种瓶颈特征表示可以捕捉语音的产生空间，有助于解释声学空间中的可变性，使得特征组合具有鲁棒性。

辅助特征的自适应不能仅通过刻画构音障碍语音的共用特征来区分构音障碍语音与正常语音，而应在更加详细的粒度上针对构音障碍的强变异性特征进行建模，从而能够区分不同程度的构音障碍语音变异性特征。因此，所获得的辅助特征需要表示构音障碍语音的强变异性，以便辅助声学特征进行声学模型的训练。通过分析构音障碍语音特征提取与构音障碍相关的说话人信息，对与构音障碍相关的说话人信息进行编码。获得辅助特征的自适应方法有两种：1）说话人

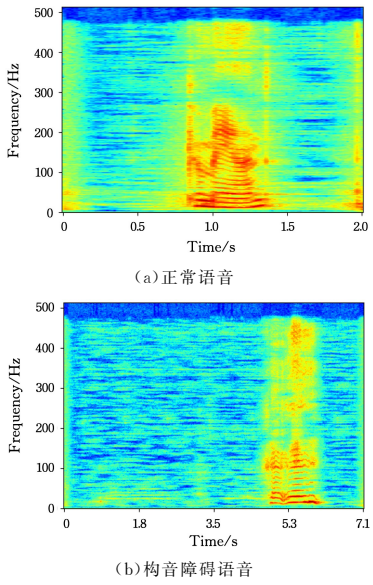
信息提取；2）说话人信息编码方法。

2.2.1 提取说话人信息作为辅助特征

说话人信息提取方法通过特征分析提取得到带有说话人信息的特征表示，将其作为辅助特征，进而实现说话人自适应。早期构音障碍语音特征研究依赖机器学习算法，提取常用的声学特征并进行特征选择。Liang 等[9]提出了基于多特征组合的构音障碍语音识别方法。该方法从语音的韵律特征、频谱特征、人耳听觉特征、嗓音质量特征和声道模型特征中构建特征的组合。通过遗传算法选择出能够获得最高分类准确率的特征子集，然后采用支持向量机分类器对所选特征进行识别。Al-Qatab 等[10]将声学特征、特征选择方法及分类器相组合，旨在探究各种组合方式对区分构音障碍不同严重程度的影响。该研究表明，相比其他组合，脉冲编码调制（Pulse Code Modulation，PCM）编码的频谱特征、Relief 特征选择方法及随机森林分类器的组合对构音障碍语音间的区分精度最高。PCM 编码的频谱特征可以提供较为丰富的信息，有助于应对不同声音条件的变化，对于构音障碍语音的可变性而言更具鲁棒性。Relief 特征选择方法可以有效减少特征的冗余，从而提高分类器的效率。然而，特征选择的组合方式在很大程度上依赖于训练数据，构音障碍语音数据不足会影响此类方法的表现。

尽管常用的声学特征具有一定的优势，但在人工提取构音障碍语音特征的过程中，异常语音中抽象但关键的信号可能被忽略[11]。对于构音障碍语音，特征提取的关键是对构音障碍语音变异性特征进行建模，充分捕捉其特征的细粒度表示。利用深度学习方法可以有效提取构音障碍语音中的异常特征[12]。Yao 等[13]使用鲸鱼优化算法（Whale Optimization Algorithm，WOA）自动选择 DCNN 的最佳结构，并训练可变长度 DCNN 模型以获得与构音障碍相关的语音特征。该方法在互联网协议地址（Internet Protocol Address，IPA）编码的基础上，利用鲸鱼向量对 DCNN 进行编码，并利用特定鲸鱼向量维度的衰减层设计可变长度的 DCNN，对构音障碍语音的分类准确率高达 95.77%。

构音障碍语音与正常语音的频谱-时间（Spectro-Temporal）存在差异[14]，如图 1 所示。受此启发，Chandrashekar 等[15]证明频谱-时间特征有利于区分构音障碍语音间的差异，在 UA-Speech[16]及 Torgo[17]数据库上的平均分类准确率分别为 92%和 97.9%，其通过研究构音障碍语音中不同时频的表示评估语音在单词级别和句子级别上的可懂度，结果表明常数 Q 变换谱图（Constant-Q Transform Spectrograms）和经感知增强及 mel 扭曲的短时傅里叶变换谱图（Short Time Fourier Transform Spectrograms）在分类任务中表现更好，在单词级别及句子级别上的可懂度分别为 72.6%和78.5%[18]。Geng 等[19]采用提取时间频谱子空间特征的方法，以实现对不同受损程度的构音障碍语音特征的细粒度表示。这些特征被作为辅助特征，成功实现了构音障碍语音的自适应。此外，Geng 等[20]利用多任务训练同时完成构音障碍的语音识别和严重程度预测，通过调整训练损失的权重，使得两个任务的训练过程得以平衡。由构音障碍严重程度预测所得的频谱子空间特征有助于说话人自适应的训练，从而可以细粒度地适应不同的构音障碍说话人。



（a）正常语音

（b）构音障碍语音

图 1 发音 command 的语谱图

Fig. 1 Spectrogram of the pronunciation of command

鉴于构音障碍的标注语料数据稀疏，如何利用未标注数据进行说话人信息提取成为了亟待解决的问题。近年来，以 Wav2vec 预训练模型为代表的方法备受关注，用于对未标注数据进行预训练，这也成为构音障碍语音特征研究的热点。Hernandez 等[21]将 Wav2vec 的自监督语音表示作为训练构音障碍语音识别系统的语音特征。由于构音障碍语音具有极强的变异性，自监督学习虽具有较好的泛化能力，但并不能完全捕捉到构音障碍语音的所有变异性特征，仍需通过参数微调适应特定数据。Baskar 等[22]提出了自适应网络，将第二代预训练模型 Wav2vec2 与构音障碍说话人自适应结合。该方法首先自适应数据集中每个说话人的声学特征，将说话人信息作为辅助特征，同时通过参数更新，将特征与编码器相连接，从而有效地微调 Wav2vec2 模型。

然而，保留说话人信息的特征训练需要一定的时间，从而导致了自适应过程的延时。Geng 等[23]提出了基于说话人级别方差正则化频谱基嵌入（Speaker-Level Variance Regularized Spectral Basis Embedding）特征的方法，用于在构音障碍语音识别中减小自适应过程的延时。该方法通过在训练基于频谱嵌入特征的过程中引入额外的正则化项，以保证构音障碍语音特征在说话人级别上的同质性，从而在测试时间内达到即时适应的效果。正则化项可以限制特征空间，以减小其特征在说话人间的差异，从而降低由语音变化引起的识别错误，提高识别的稳定性。

说话人信息提取方法需在语音中提取与构音障碍相关的特征表示，因此该方法的重点是构音障碍语音特征分析。构音障碍语音特征分析方法如表 1 所列。构音障碍语音通常表现出语音特征的多样性和复杂性，仅使用语音的单一视图往往难以捕捉和利用多源数据中的潜在信息。Wu[24]采用多视图学习方法，旨在显式使用数据的多个不同表示，并建模这些表示之间的关系。该方法利用多个不同视图的音频数据，通过自注意力和交叉注意力模块在编码器阶段进行特征编码，然后在解码器中整合这些特征。该方法在各个视图学习得到的隐层表示中引入了交叉注意力信息约束，有助于更好地捕捉

不同视图之间的信息关联，以促进信息的交叉融合。Zhao等[25]提出了一种多尺度梅尔域特征图谱提取算法。该算法采用经验模态分解法对语音信号进行分解，针对3个有效分量分别提取Fbank特征以及其一阶差分，再将各帧特征进行拼接构成新的特征图谱，以捕捉频域细节信息。通过经验模态分解得到有效表达语音信息的分量，弥补漏掉的语音低频细节特征；考虑到人耳的结构特性，对Fbank特征及其差分特征的提取有助于保留语音信号数据之间的相关性，以捕获语音信号的时变信息以及相邻帧信息之间的联系。

表1 构音障碍语音特征分析方法

Table 1 Methods of speech feature analysis for dysarthria

名称	方法	优势	劣势	评价指标	时间
基于注意力机制[26]	注意力机制 LSTM	可懂度评估无需参考正常语音	听觉显著性特征没有纳入注意力机制计算	可懂度评估准确率 76.97%±0.28%	2020
基于音素后验特征[27]	在音素后验特征空间与广义语音后验特征空间中匹配构音障碍说话人语音	不依赖对照语音数量	特征空间中构音障碍语音变化的捕捉能力欠佳	主观和客观可懂度：皮尔森系数(Pearson) 0.961 斯皮尔曼系数(Spearman) 0.931	2021
基于时间包络和精细结构[28]	计算时间包络和精细结构的表示 CNN	捕捉构音障碍语音的感知线索	模型较为简单	检测准确率85.72%	2021
基于非负矩阵分解的语音信号时频特征[29]	时频矩阵表示、时频矩阵分解、特征提取和分类	捕捉语音信号的突变及不连续	不适用于连续语音信号	单词分类准确率97%	2021
基于频谱时间子空间[30]	奇异值分解提取频谱时间子空间	捕捉构音障碍语音中不精确发音，考虑分析时间子空间	对参考说话人的选择不太敏感	检测准确率96.3%	2021
前端语音参数化[31]	短时傅里叶变换幅度谱累加器 对数压缩频谱动态计算	捕捉精细程度上与构音障碍相关的鉴别信息	过分强调高频区域，准确率随滤波器的增加而降低	基于概率线性判别分析的评分方案检测准确率 87.262%	2022
原始声源和滤波器组件的声学建模[32]	声源和滤波器组件融合	归一化构音障碍说话人的相关属性	得到构音障碍说话人相关属性与i-vector相似，两者结合性能变差	词错误率： UASpeech 30.3% Torgo 35.3%	2022
自监督预训练[33]	声学特征与wav2wec2.0的特征融合 TDNN联合译码 使用语音表示的多遍译码	解决过拟合和泛化较差的问题	没有考虑说话人个性化问题	词错误率： UASpeech 22.56%	2023

大多数构音障碍语音分析研究使用原始语音时域信号或者其频谱训练模型。而原始语音在声音传播过程中的声道共振可能干扰构音障碍语音特征的分析。为了排除声道共振信号对构音障碍语音的干扰，一些研究提出使用声门源(Glottal Source)波形，即气流通过声门时的波动形式，作为另一种时域信号分析构音障碍语音特征。Narendra等[34]将分别由原始语音波形与声门源波形训练所得的特征进行对比，结果表明从声门源波形中提取的特征更有利于识别构音障碍语音。声门源波形的提取和分析相对于原始语音波形更为复杂，需要额外的计算和处理步骤以处理声门源波形。此外，该团队[35]利用准闭合相位分析(Quasi-Closed Phase Analysis)[36]提取声门源特征，并评估其在区分构音障碍语音可懂度中的有效性。与原始语音波形的基线相比，两类声门源特征组合在语音可懂度评估中始终表现最佳。

构音障碍语音分析研究主要侧重于声学特征，而对发音运动空间的研究则较为有限。通过对发音器官(如舌头、喉部等)的生理运动分析，可以更全面地捕捉构音障碍患者的语音问题。Duan等[37]通过比较分析构音障碍者与正常人以及不同病情程度患者之间的发音运动空间差异，深入研究构音障碍者的语音特征。研究采用3D散点发音轨迹和空间位移分析，以揭示构音障碍者的发音运动特征。结果表明，构音障碍者的舌部发音运动更趋向于口腔后部、左侧和下方，而且病情越严重，舌部的抬起运动愈加困难。此外，研究使用K-means算法计算了发音运动空间的质心，通过显著性分析，揭示了不同病情程度的构音障碍者在上下方向上存在显著差异。最后研究以质心和位移中值为特征，采用随机森林分类器进行病情分级，分类准确率高达98%，较J48决策树提高了6.45%。

上述研究将构音障碍语音变异性视为说话人的差异，并试图通过使用说话人相关的信息处理语音变异性。然而，构音障碍语音变异性是由不同的条件引起的，即使对同一个人而言，其变异性也可能不确定。因此，说话人信息编码方法的出现旨在处理构音障碍语音中的这种变异性。

2.2.2 编码说话人信息作为辅助特征

编码说话人信息的方法通过显式编码构音障碍语音的变异性，并将其用作声学模型的辅助特征，从而将每一个说话人的特征映射到一个说话人无关的特征空间上。Xie等[38]使用基于变分自动编码器的可变性编码器(Variational Auto-Encoder Based Variability Encoder，VAEVE)编码构音障碍的语音变异性，通过应用随机梯度变分贝叶斯算法(Stochastic Gradient Variational Bayes)建模其语音的不确定性，用于生成可变性编码。VAEVE结合音素信息和低维潜在变量，以重新构建构音障碍语音特征，并利用潜在变量编码与音素无关的变异性。结果表明，使用可变性编码的DNN系统始终优于基线系统，同时VAEVE的引入有助于增强说话人自适应训练的效果。该方法在潜在空间中生成连续表示，有助于更好地捕捉构音障碍语音的变异性。然而，对于变分自动编码

器而言，潜在空间的复杂性通常难以解释，尤其是空间中各个维度难以直接映射到可解释的语音学属性或声学特征。因此，在编码说话人信息时需考虑如何增强潜在空间的可解释性。

辅助特征的自适应通过将构音障碍语音加工得到的构音障碍相关特征嵌入到模型之中，以完成自适应过程。然而，该类方法并未充分利用构音障碍者的身份信息进行模型训练。与基于模型域的自适应相比，构音障碍说话人自适应数据量相对较少，自适应效果的提升幅度相对有限。

3 基于模型域的自适应

基于模型域的自适应直接利用自适应数据进行声学模型训练，通过改变模型参数模拟构音障碍语音的变异性，从而构建构音障碍相关模型。早期的声学模型以 GMM-HMM 最为典型，其中 HMM 负责建立状态间的转移概率分布，用于描述信号的动态特性，信号的静态特性则由 GMM 描述，主要负责生成 HMM 的观察值概率。由于 HMM 的每个状态都由 GMM 建模，以确定在该状态下观察的可能性，基于 GMM-HMM 的自适应方法一般是对 GMM 模型的调整。随着深度神经网络的崛起，早期声学模型的弊端逐步显现，例如，GMM 既不能利用帧的上下文信息，也无法学习到深层非线性特征变换，再加上 GMM 是生成性模型，而 DNN 是鉴别性模型，基于 GMM 的自适应无法直接对基于 DNN 的系统进行自适应。因此，基于模型域的自适应集中在深度神经网络模型上。

3.1 线性变换声学模型

线性变换的自适应指对模型参数进行线性变换，从而将说话人无关模型转换为相关模型。基于 GMM 的自适应主要是最大似然线性回归（Maximum Likelihood Linear Regression，MLLR）[39]，该算法利用自适应数据，对原模型的高斯均值进行线性变换，从而得到转换矩阵，该矩阵可以由所有高斯函数共用。通过共用自适应参数，MLLR 可以为 GMM 的每个高斯函数作出相应的自适应训练。

基于 DNN 的自适应指在说话人无关模型中添加线性变换层，利用变换关系更新变换层，对模型原有参数进行调整，从而将说话人无关模型变换为相关模型。然而，加入线性变换层后，变换时所有层的输出都会随之改变，从而导致自适应模型参数急剧增加。对于相对有限的构音障碍自适应数据来说，大量自适应参数的存在极易导致自适应训练的过拟合问题。为了减少自适应参数，Yu 等将学习性隐藏单元分布（Learning Hidden Unit Contributions，LHUC）[40]方法用于构音障碍的自适应[41-42]。具体思路是，通过给定说话人信息学习构音障碍者相关的隐藏层，利用自适应数据改变神经元间的组合模式，为隐藏层中所有的神经元赋予新的权重，通过无监督自适应过程对权重不断更新直至收敛，从而使网络能够匹配目标说话人。

此外，隐藏单元偏差（Hidden Unit Bias，HUB）[43]使自适应参数作为特定 DNN 层中的向量偏差，将说话人级别向量偏差添加到隐藏单元输出，同样可以减少自适应参数。参数化激活函数（Parameterized Activation Functions，Pact）[44]在馈入 ReLU（Rectified Linear Unit）激活函数前对输入特征进行说话人级别向量缩放或偏差的变换。激活函数为语音数据在 DNN 模型中的最后阶段，激活函数参数可以直接影响该层的输出。此外，与数量多的权重相比，该隐层节点相对较少，对有限的构音障碍语音进行自适应的效果较好。基于线性变换的自适应方法如图 2 所示。Liu 等[45]在自动神经结构搜索（Neural Architecture Search）自动配置的 DNN 上采用上述 3 种自适应方法对构音障碍说话人进行自适应，并将方法与 i-vector 的自适应方法进行对比。结果表明，使用 LHUC 自适应方法的模型性能最优，3 种基于线性变换的自适应方法始终优于基于 i-vector 的自适应方法。

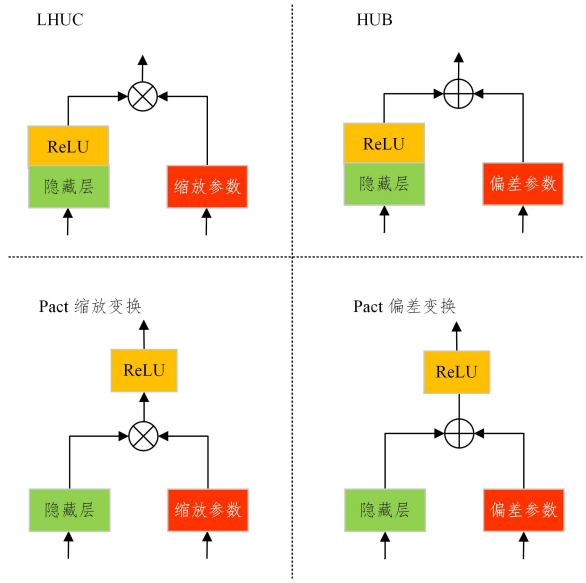


图 2 线性变换声学模型的自适应方法

Fig. 2 Adaptation methods of linear transformation acoustic model

尽管采用 LHUC 自适应的模型性能较为出色，但仍然会受到自适应数据稀疏问题的影响。因此，构音障碍自适应研究往往将辅助特征的自适应与 LHUC 自适应相结合，自适应后的网络能够更加匹配特定说话人并取得更好的性能。Geng 等[23]提出了基于说话人级别谱特征的即时 LHUC 变换，将辅助特征嵌入和自适应训练相结合，从而更好地实现构音障碍说话人自适应。即时 LHUC 变换通过实时预测声学特征，直接在适应过程中生成和应用参数，解决了多次解码导致的延时问题。

线性变换的自适应训练在消除构音障碍说话人之间的差异性的同时，通过模拟构音障碍说话人内部变异性线性变换说话人无关模型的模型参数，使得变换后的说话人相关模型与构音障碍语音相匹配。此外，该方法通过减少自适应参数，在一定程度上减小了构音障碍语音数据稀疏带来的自适应训练过拟合影响。

3.2 微调声学模型参数

构音障碍的模型域自适应是对说话人无关的基本模型进行自适应，基本模型通常由正常语音预训练得到。然而，正常语音与构音障碍语音在特征向量的概率分布上存在不匹配问题，从而限制了基本模型的泛化能力。模型微调的自适应通过模拟构音障碍语音的强变异性，更新基本模型的部分参数，将基本模型转化为说话人相关的模型，使得模型与构音障碍语音特征相匹配。

对GMM模型的均值参数而言，最大后验概率(Maximum A Posteriori，MAP)算法[46]基于原有均值μ和自适应数据$\boldsymbol{X}=\{x_1,x_2,\cdots,x_t,\cdots,x_T\}$，对模型进行训练，得到某状态自适应后的均值$\hat{\mu}$，从而使原有模型参数和自适应数据的估计达到平衡。然而，MAP方法一般要求大量数据，因此一些研究将MLLR与MAP方法结合[47-48]，以缓解数据稀疏问题。Sehgal等[49]使用基于MLLR-MAP的混合方法，仅利用构音障碍语音对说话人无关系统进行说话人自适应训练(Speaker Adaptive Training)，从而减小不同程度构音障碍语音之间的差异性。与单一的MAP方法相比，其准确率提高了11.05%。虽然该方法在一定程度上能够缓解构音障碍语音中数据规模有限的问题，但其带来了更高的复杂度。

大多数研究只更新声学模型的部分参数，如无词图最大互信息(Lattice Free Maximum Mutual Information)训练的TDNN[50-51]系统和RNN-T[52-53]系统。Takashima等[50]首先在说话人无关的TDNN上适应多个构音障碍者的一般说话风格，构建构音障碍的适应模型，然后利用目标说话人信息对适应模型进行微调，从而进一步适应目标说话人。此外，该研究利用多个构音障碍说话人语音分别从头训练说话人无关和说话人相关的构音障碍模型，以及利用目标构音障碍说话人语音直接微调基本模型进行适应。结果表明，两步适应方法比利用目标说话人数据直接微调基本模型的适应方法效果更好。通过在多个构音障碍者语音上进行第一步训练，模型能够学习到通用的构音障碍特征。由于模型在源域上得到了充分训练，在第二步的目标说话人适应中进行参数微调变得容易。

然而，自适应数据在数据稀疏的情况下更新模型参数会破坏基本模型的信息。为解决数据稀疏所导致的模型参数不确定问题，Deng等[54]提出贝叶斯参数和神经结构域自适应方法及贝叶斯可微结构搜索(Differentiable Architectural Search)超网络模型。其中，神经结构域自适应地根据源域和目标域数据动态调整神经网络的结构，旨在应对不同域数据间的分布差异。该模型既可以在域自适应过程中对不同的TDNN结构进行有效搜索，又可以在模型参数不确定的情况下稳健建模。Wang等[55]将超参量的自适应转换为可微结构搜索任务。在可微结构搜索的超网络中，候选结构与仅使用源域数据的编码块和解码块的变化超参数设置相关，随后该结构被适应到构音障碍目标域之中。在这个过程中，大量的标准Conformer超网络参数被继承并适应有限的目标域数据，而相对较少的超参数选择权重在自适应过程中被微调。为防止微调后模型参数偏离原模型参数，Kim等[56]提出了基于L2正则化自适应方法，反映了构音障碍说话人的语音变化。其中，L2正则化方法通过向目标函数添加惩罚项，使得模型可收敛到一个局部最优解，确保自适应参数对应等高线离中心点距离适当，避免了过拟合，在保留说话人特定信息的同时，约束了自适应模型与说话人无关模型之间的距离。正则化方法基于一种假设，即目标域的概率分布与源域的概率分布没有太大的不同，但就构音障碍语音间的差异而言，上述假设不一定成立。针对过拟合问题，Qi等[57]将适配器融合应用于构音障碍说话人自适应中。该方法首先利用构音障碍源说话人数据训练适配器，再使用目标说话人数据训练融合层，通过融合层的注意力机制将来自源适配器的表示进行组合，计算注意力分数并将其分配给线性层的输出值。此外，该方法通过减少query层和key层的大小，并利用Household变换对线性层进行重新参数化，进一步提高了融合层的参数效率。该融合层仅用三分之一的参数就能达到与微调方法相当的识别效果，从而避免了自适应过程中的过拟合。

3.3 基于数据选择的域自适应

由于构音障碍语音间也存在差异，因此模型参数微调的关键是选择合适的训练数据，使得来自源域的数据(训练集)概率分布接近于来自目标域的数据(测试集)概率分布。通常采用数据选择以及权重调整来增强构音障碍语音的相似性，主要做法是利用多个源域数据进行模型训练，并利用未标注的目标域数据实现自适应。

Wang等[7]将与模型无关的元学习(Model-Agnostic Meta Learning)和Reptile算法进行扩展，通过反复模拟适应不同的构音障碍说话人元更新基本模型，证明了元学习和其他算法相结合可以解决少样本条件下的构音障碍自适应任务。此外，该研究尝试通过改变源模型的参数模拟目标说话人的声学特征。Christensen等[58]提出了基于说话人的数据选择策略，通过测量目标说话人与构音障碍说话人池之间的相似度，根据说话人与目标说话人的匹配程度选择可用说话人池的子集，从而利用严重程度相似的构音障碍语音进行训练。由于不同构音障碍者的语音特征也都不同，因此，严重程度相似的其他说话人的语音数据不一定能提高系统性能。Xiong等[51]提出了基于话语的数据选择策略，旨在提高迁移效率。该方法通过分析特定说话人相关模型的后验概率熵，从可用的源域数据中选择数据，并将数据添加到特定目标说话人的训练数据中。研究仅利用源域数据训练声学模型，通过神经网络权重自适应，将声学模型与选择出的数据同时适应到目标域上。

然而，构音障碍语音的标注难度较大，与源域数据概率分布不一致的目标域数据通常是不可见的，即出现构音障碍语料库之外的数据。当目标域数据不可见时，仅利用源域数据训练的模型性能仍会显著下降。因此，一些研究利用已标注的源域数据和未标注的目标域数据构建模型。如Wang等[59]通过域对抗训练将参数域鉴别器与标记编码器交替更新，同时利用互信息最小化方法，减少生物标志嵌入与域相关信息间的依赖性，对构音障碍语音检测进行无监督域自适应。为了增强说话人相似性和可比较的语音自然度，该团队提出对抗性说话人自适应的多任务学习策略[60]，使用目标说话人语音对说话人编码器进行微调，并有效捕捉身份相关信息。该团队所提出的多任务学习策略包括3个任务，即构音障碍存在分类(Dysarthria Presence Classification，DPC)、域对抗训练(Domain Adversarial Training，DAT)和互信息最小化(Mutual Information Minimization，MIM)，以解决域不匹配的问题。DPC有助于生物标志嵌入捕获构音障碍的重要指标，DAT强制生物标志嵌入在源域和目标域中无法区分，MIM进一步降低了生物标志嵌入与领域相关线索之间的相关性。通过充分利用大量未标记的目标域数据，该方法能够获取包含

构音障碍存在的关键指标的域不变生物标志嵌入，以实现准确和稳健的构音障碍检测。然而，单一源域数据毕竟有限，与多源域数据相比，在单一源域内选择与目标域数据相似的数据范围相对较小。多源域数据的利用可以最大程度地增加源域与目标域中构音障碍语音的相似性。Turrisi 等[61]提出了基于最优传输的无监督多源域自适应（Multi-source Domain Adaptation）算法，旨在克服单一源域的局限。该算法利用源域概率分布的多样性，通过调整目标任务的权重选择与目标域最相似的源域，并通过更新目标分类器进一步增加源域与目标域中说话人在语音特征方面的相似性，从而使源域与目标域数据的概率分布在最大程度上保持一致。该方法采用最优传输原理，以最低成本将源域转化为目标域，确保源域与目标域数据分布实现最大相似。最优传输原理能够准确地捕获和量化两个域之间的分布差异，从而提高自适应能力。

4 结束语

4.1 当前现状

针对构音障碍语音与模型间的不匹配问题，本文基于特征域和模型域两个角度，从自适应基本方法出发，对构音障碍自适应方法的现有研究进行回顾和分析，其重点是解决构音障碍带来的语音变异性及差异性。

特征变换的自适应通过对特征进行线性变换，消除构音障碍语音间的差异性。然而，由于构音障碍语音的变异性，线性变换不足以归一化说话人语音特征，特征变换的自适应效果较为有限。

辅助特征的自适应通过对构音障碍语音变异性进行建模，捕捉与构音障碍相关的语音特征，利用其特征辅助对应的声学特征进行模型训练，从而完成自适应过程。而构音障碍语音特征分析集中于构音障碍语音的频谱-时间，通过分析构音障碍语音间的频谱-时间差异，捕捉构音障碍语音特征。大部分研究利用深度学习方法提取特征，将特征作为辅助特征进行自适应。已有研究尝试将说话人的声门源特征用作区分构音障碍语音的特征并取得了一定的效果。与此同时，构音障碍语音特征分析方法不能仅对构音障碍语音的共有特征进行刻画，还要在更加详细的粒度上建模构音障碍语音变异性。此外，针对构音障碍语音变异的不确定，可以利用自动编码器显式编码构音障碍语音的变异性。

线性变换的自适应是对模型参数进行线性变换，通过模拟构音障碍语音变异性，从而使原模型转换为说话人相关模型。尽管该方法在一定程度上可以减少自适应训练的过拟合问题，但仍会受到构音障碍自适应数据稀疏的影响。因此，线性变换的自适应往往与辅助特征的自适应结合训练，以缓解构音障碍语音的数据稀疏。

模型微调的自适应通过改变基本模型的模型参数，将声学模型变为与构音障碍相关的模型，从而使模型与构音障碍语音相匹配。基本模型通常由正常语音训练得到。

由于构音障碍语音与正常语音间的差异较大，因此两者的特征概率分布存在不一致的问题，域自适应的研究重点在于增强语音间的相似性。主要思路是通过数据选择及权重调整策略使得训练数据的概率分布接近目标说话人语音的概率分布，从而实现对基本模型参数的微调。为进一步增加语音相似性，可以从多个源域数据中利用数据概率分布的多样性增强语音特征的相似性。

4.2 挑战与趋势

基于对大量文献的调研发现，构音障碍自适应目前的挑战体现在：1）不同程度构音障碍语音的变异性以及其变异不确定性的建模；2）模型自适应中源域与目标域概率分布的平衡性问题；3）数据稀疏下的构音障碍识别的自适应问题。

未来构音障碍自适应将呈现以下发展趋势：

1）语音变异性的分析：构音障碍语音的变异性分析成为了构音障碍自适应的一大重点。可以借鉴通用语音识别中深度学习模型的特征学习代替人工设定的声学特征，对构音障碍语音变异性进行特征表示；此外，利用编码器结构编码构音障碍语音的变异性，捕捉其变异的不确定性。

2）多特征多模态数据的融合：采用特征融合，将构音障碍语音的生物特征信息（如声门运动、声带震动等）与声学特征结合；整合多模态数据，如音频、视频、语谱图，有助于提供更丰富的信息，以增强构音障碍的自适应。

3）小数据量下的自适应：数据是构音障碍自适应研究的最大挑战。面对规模有限的构音障碍语音数据，小样本学习可能是未来解决构音障碍自适应研究瓶颈的潜在方法，包括元学习、迁移学习和生成网络。其中，迁移学习一般利用预训练模型，针对特定数据进行参数微调。另外，考虑构音障碍语音的差异性，利用对抗训练及权重调整策略增强语音相似性，平衡源域与目标域数据的概率分布，可将预训练模型融入构音障碍的自适应中。

参 考 文 献

[1] RIGOLL G. Speaker adaptation for large vocabulary speech recognition systems using speaker Markov models[C]// International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1989: 5-8.

[2] ZHU F Y, MA Z Q, CHEN Y, et al. A Survey of Speaker Adaptation Methods in Speech Recognition[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(12): 2241-2255.

[3] HAHM S, HEITZMAN D, WANG J. Recognizing dysarthric speech due to amyotrophic lateral sclerosis with across-speaker articulatory normalization[C]// Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies. 2015: 47-54.

[4] GALES M J F. Maximum likelihood linear transformations for HMM-based speech recognition[J]. Computer speech & language, 1998, 12(2): 75-98.

[5] BHAT C, VACHHANI B, KOPPARAPU S K. Recognition of Dysarthric Speech Using Voice Parameters for Speaker Adaptation and Multi-Taper Spectral Estimation[C]// Interspeech. 2016: 228-232.

[6] SAON G, SOLTAU H, NAHAMOO D, et al. Speaker adaptation of neural network acoustic models using i-vectors[C]// 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. IEEE, 2013: 55-59.

[7] WANG D, YU J, WU X, et al. Improved End-to-End Dysarthric

Speech Recognition via Meta-learning Based Model Re-initialization[C]// 2021 12th International Symposium on Chinese Spoken Language Processing(ISCSLP). Hong Kong: IEEE, 2021: 1-5.

[8] YILMAZ E, MITRA V, SIVARAMAN G, et al. Articulatory and bottleneck features for speaker-independent ASR of dysarthric speech[J]. Computer Speech & Language, 2019, 58: 319-334.

[9] LIANG Z Y, LI Y X, SUN Y, et al. Speech recognition of dysarthria based on multi feature combination[J]. Computer Engineering and Design, 2022, 43(2): 567-572.

[10] AL-QATAB B A, MUSTAFA M B. Classification of Dysarthric Speech According to the Severity of Impairment: an Analysis of Acoustic Features[J]. IEEE Access, 2021, 9: 18183-18194.

[11] KONG A P H, TSE C W K, KONG A P H, et al. Clinician survey on speech pathology services for people with aphasia in Hong Kong[J]. Clinical Archives of Communication Disorders, 2018, 3(3): 201-212.

[12] ZHENG W, TIAN X, YANG B, et al. A few shot classification methods based on multiscale relational networks[J]. Applied Sciences, 2022, 12(8): 4059.

[13] YAO D, CHI W, KHISHE M. Parkinson's disease and cleft lip and palate of pathological speech diagnosis using deep convolutional neural networks evolved by IPWOA[J]. Applied Acoustics, 2022, 199: 109003.

[14] KODRASI I, BOURLARD H. Spectro-Temporal Sparsity Characterization for Dysarthric Speech Detection[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 1210-1222.

[15] CHANDRASHEKAR H M, KARJIGI V, SREEDEVI N. Spectro-Temporal Representation of Speech for Intelligibility Assessment of Dysarthria[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(2): 390-399.

[16] KIM H, HASEGAWA-JOHNSON M, PERLMAN A, et al. Dysarthric speech database for universal access research[C]// Ninth Annual Conference of the International Speech Communication Association. 2008.

[17] RUDZICZ F, NAMASIVAYAM A K, WOLFF T. The TORGO database of acoustic and articulatory speech from speakers with dysarthria[J]. Language Resources and Evaluation, 2012, 46(4): 523-541.

[18] CHANDRASHEKAR H M, KARJIGI V, SREEDEVI N. Investigation of different time-frequency representations for intelligibility assessment of dysarthric speech[J]. Ieee transactions on neural systems and rehabilitation engineering, 2020, 28(12): 2880-2889.

[19] GENG M, XIE X, YE Z, et al. Speaker Adaptation Using Spectro-Temporal Deep Features for Dysarthric and Elderly Speech Recognition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30: 2597-2611.

[20] GENG M Z, JIN Z R, WANG T Z, et al. Use of Speech Impairment Severity for Dysarthric Speech Recognition[J]. arXiv: 2305.10659, 2023.

[21] HERNANDEZ A, PÉREZ-TORO P A, NÖTH E, et al. Cross-lingual Self-Supervised Speech Representations for Improved Dysarthric Speech Recognition[C]// Interspeech 2022. ISCA, 2022: 51-55.

[22] KARTHICK BASKAR M, HERZIG T, NGUYEN D, et al. Speaker adaptation for Wav2vec2 based dysarthric ASR[C]// Interspeech 2022. ISCA, 2022: 3403-3407.

[23] GENG M, XIE X, SU R, et al. On-the-fly Feature Based Speaker Adaptation for Dysarthric and Elderly Speech Recognition[J]. arXiv: 2203.14593, 2022.

[24] WU L D. Multi-view dysarthria speech recognition based on deep temporal network [D]. Shanghai: East China Normal University, 2022.

[25] 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.

[26] FERNÁNDEZ-DÍAZ M, GALLARDO-ANTOLÍN A. An attention Long Short-Term Memory based system for automatic classification of speech intelligibility[J]. Engineering Applications of Artificial Intelligence, 2020, 96: 103976.

[27] FRITSCH J, MAGIMAI-DOSS M. Utterance Verification-Based Dysarthric Speech Intelligibility Assessment Using Phonetic Posterior Features[J]. IEEE Signal Processing Letters, 2021, 28: 224-228.

[28] KODRASI I. Temporal Envelope and Fine Structure Cues for Dysarthric Speech Detection Using CNNs[J]. IEEE Signal Processing Letters, 2021, 28: 1853-1857.

[29] KARAN B, SAHU S S, OROZCO-ARROYAVE J R, et al. Nonnegative matrix factorization-based time-frequency feature extraction of voice signal for Parkinson's disease prediction[J]. Computer Speech & Language, 2021, 69: 101216.

[30] JANBAKHSHI P, KODRASI I, BOURLARD H. Subspace-Based Learning for Automatic Dysarthric Speech Detection[J]. IEEE Signal Processing Letters, 2021, 28: 96-100.

[31] SAHU L P, PRADHAN G. Analysis of Short-Time Magnitude Spectra for Improving Intelligibility Assessment of Dysarthric Speech[J]. Circuits, Systems, and Signal Processing, 2022, 41: 5676-5698.

[32] YUE Z, LOWEIMI E, CHRISTENSEN H, et al. Acoustic Modelling From Raw Source and Filter Components for Dysarthric Speech Recognition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30: 2968-2980.

[33] HU S J, XIE X R, JIN Z R, et al. Exploring self-supervised pretrained asr models for dysarthric and elderly speech recognition [C]// ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE, 2023: 1-5.

[34] NARENDRA N P, ALKU P. Glottal Source Information for Pathological Voice Detection[J]. IEEE Access, 2020, 8: 67745-67755.

[35] NARENDRA N P, ALKU P. Automatic intelligibility assessment of dysarthric speech using glottal parameters[J]. Speech Communication, 2020, 123: 1-9.

[36] AIRAKSINEN M, RAITIO T, STORY B, et al. Quasi closed phase glottal inverse filtering analysis with weighted linear prediction[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2013, 22(3): 596-607.

[37] DUAN S F，WANG J Q，DINGAM C，et al. Disease Degree Classification of Dysarthria Based on Spatial Features of Articulation [J]. Journal of Fudan University(Natural Science)，2021，60(3)：288-296.

[38] XIE X，RUZI R，LIU X，et al. Variational Auto-Encoder Based Variability Encoding for Dysarthric Speech Recognition[C]// Interspeech 2021. ISCA，2021：4808-4812.

[39] LEGGETTER C J，WOODLAND P C. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models[J]. Computer Speech & Language，1995，9(2)：171-185.

[40] NETO J，ALMEIDA L，HOCHBERG M，et al. Speaker-adaptation for hybrid HMM-ANN continuous speech recognition system[C]// 4th European Conference on Speech Communication and Technology(Eurospeech 1995). 1995：2171-2174.

[41] YU J，XIE X，LIU S，et al. Development of the CUHK Dysarthric Speech Recognition System for the UA Speech Corpus [C]// Interspeech 2018. ISCA，2018：2938-2942.

[42] GENG M，XIE X，LIU S，et al. Investigation of data augmentation techniques for disordered speech recognition[C]// Interspeech 2020. ISCA，2020：696-700.

[43] XIE X，LIU X，LEE T，et al Bayesian learning for deep neural network adaptation [J]. IEEE/ACM Transactions on Audio，Speech，and Language Processing，2021，29：2096-2110.

[44] ZHANG C，WOODLAND P C. DNN speaker adaptation using parameterised sigmoid and ReLU hidden activation functions [C]// IEEE International Conference on Acoustics，Speech and Signal Processing(ICASSP 2016). IEEE，2016：5300-5304.

[45] LIU S，GENG M，HU S，et al. Recent Progress in the CUHK Dysarthric Speech Recognition System[J]. IEEE/ACM Transactions on Audio，Speech，and Language Processing，2021，29：2267-2281.

[46] GAUVAIN J L，LEE C H. MAP estimation of continuous density HMM：theory and applications[C]// Speech and Natural Language：Proceedings of a Workshop Held at Harriman. New York，1992.

[47] MENGISTU K T，RUDZICZ F. Adapting acoustic lexical models to dysarthric speech[C]// IEEE International Conference on Acoustics，Speech and Signal Processing (ICASSP 2011). IEEE，2011：4924-4927.

[48] KIM M J，YOO J，KIM H. Dysarthric speech recognition using dysarthria-severity-dependent and speaker-adaptive models [C]// Interspeech. 2013：3622-3626.

[49] SEHGAL S，CUNNINGHAM S. Model adaptation and adaptive training for the recognition of dysarthric speech[C]// Proceedings of SLPAT 2015：6th Workshop on Speech and Language Processing for Assistive Technologies. 2015：65-71.

[50] TAKASHIMA R，TAKIGUCHI T，ARIKI Y. Two-step acoustic model adaptation for dysarthric speech recognition[C]// IEEE International Conference on Acoustics，Speech and Signal Processing(ICASSP 2020). IEEE，2020：6104-6108.

[51] XIONG F，BARKER J，YUE Z，et al. Source domain data selection for improved transfer learning targeting dysarthric speech recognition[C]// IEEE International Conference on Acoustics，Speech and Signal Processing(ICASSP 2020). IEEE，2020：7424-7428.

[52] SHOR J，EMANUEL D，LANG O，et al. Personalizing ASR for dysarthric and accented speech with limited data[C]// Interspeech 2019. ISCA，2019：784-788.

[53] GREEN J R，MACDONALD R L，JIANG P P，et al. Automatic Speech Recognition of Disordered Speech：Personalized Models Outperforming Human Listeners on Short Phrases[C]// Interspeech. 2021：4778-4782.

[54] DENG J，GUTIERREZ F R，HU S，et al. Bayesian Parametric and Architectural Domain Adaptation of LF-MMI Trained TDNNs for Elderly and Dysarthric Speech Recognition[C]// Interspeech. 2021：4818-4822.

[55] WANG T Z，HU S K，DENG J J，et al. Hyper-parameter Adaptation of Conformer ASR Systems for Elderly and Dysarthric Speech Recognition[J]. arXiv：2306. 15265，2023.

[56] KIM M，KIM Y，YOO J，et al. Regularized speaker adaptation of KL-HMM for dysarthric speech recognition[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering，2017，25(9)：1581-1591.

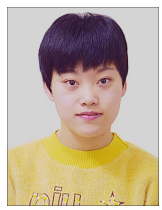
[57] QI J Z，HAMME H V. Parameter-efficient Dysarthric Speech Recognition Using Adapter Fusion and Householder Transformation[J]. arXiv：2306. 07090，2023.

[58] CHRISTENSEN H，CASANUEVA I，CUNNINGHAM S，et al. Automatic selection of speakers for improved acoustic modelling：Recognition of disordered speech with sparse data [C]// IEEE Spoken Language Technology Workshop (SLT 2014). IEEE，2014：254-259.

[59] WANG D，DENG L，YEUNG Y T，et al. Unsupervised Domain Adaptation for Dysarthric Speech Detection via Domain Adversarial Training and Mutual Information Minimization[J]. arXiv：2106. 10127，2021.

[60] WANG D，LIU S，WU X，et al. Speaker Identity Preservation in Dysarthric Speech Reconstruction by Adversarial Speaker Adaptation[C]// 2022 IEEE International Conference on Acoustics，Speech and Signal Processing(ICASSP 2022). IEEE，2022：6677-6681.

[61] TURRISI R，BADINO L. Interpretable Dysarthric Speaker Adaptation based on Optimal-Transport[C]// Interspeech 2022. ISCA，2022：26-30.



KANG Xinchen，born in 1996，postgraduate，is a member of CCF (No. P5697G). Her main research interests include information accessibility and speech recognition.

DONG Xueyan，born in 1986，Ph.D，senior lecturer. Her main research interests include information accessibility and speech recognition.

（责任编辑：喻藜）