DOI:10.20033/j.1003-7241.(2024)05-0163-05.

基于PNCC特征的录音语音自动识别方法*

肖 宜，葛 罗，胡 凯，严斌俊，邵立政
（国网湖北省电力有限公司，湖北 武汉 430077）

摘 要：为了提高录音语音自动识别方法的识别精度，提出基于语音特征提取算法（Power-Normalized Cepstral Coefficients，PNCC）特征的录音语音自动识别方法。采用小波变换方法转换录音语音信号，建立语音特征序列；利用滤波器对语音信号进行滤波分析，抑制非对称噪声；采用离散余弦变换方法变换非线性信号序列，提取基于PNCC特征的录音语音特征；利用人工蜂群算法对录音语音特征参数进行矢量量化，获得录音语音最优码书；构建录音语音识别模型，将提取的特征参数与录音语音最优码书输送到模型内，实现录音语音自动识别。实验结果表明，该方法的录音语音误识率低于20%，提高了识别精度，有效性较强。

关键词：PNCC特征；录音语音；自动识别；矢量量化
中图分类号：TP391.42 文献标识码：A 文章编号：1003-7241(2024)05-0163-05

Automatic Recognition Method of Recorded Speech Based on PNCC Feature

XIAO Yi, GE Luo, HU Kai, YAN Bin-jun, SHAO Li-zheng
(State Grid Hubei Electric Power Company, Wuhan 430077 China)

Abstract: In order to improve the recognition accuracy of automatic recording speech recognition method, an automatic recording speech recognition method based on speech feature extraction algorithm (PNCC) feature is proposed. The wavelet transform method is used to transform the recorded speech signal and establish the recorded speech feature sequence. The sequence is filtered and analyzed by filter to suppress the asymmetric noise in speech. The discrete cosine transform method is used to transform the nonlinear signal sequence and extract the recorded speech features based on PNCC features. The feature parameters of recorded speech are vector quantized by artificial bee colony algorithm to obtain the optimal codebook of recorded speech. A recording speech recognition model is constructed, and the extracted feature parameters and the optimal codebook of recording speech are transmitted to the model to realize the automatic recognition of recording speech. The experimental results show that the error recognition rate of recorded speech is less than 20%, which improves the recognition accuracy and has strong effectiveness.

Keywords: PNCC features; recorded speech; automatic recognition; vector quantization

0 引言

随着人工智能技术的发展，人类与计算机的交流智能化得到越来越多的关注[1-2]。交流智能化是指机器人可以听懂人类指令，达成人机交互的目的。在科技研究人员的不懈努力下，语音识别技术已经取得了一定的成果。早在1950年，美国实验室便设计出了英文数字语音识别系统，但该技术还不够完善，识别效果存在些许欠缺。为了使录音语音自动识别技术能够发挥出最佳识别效果，需要对录音语音自动识别方法进行研究。

姜芃旭[3]等人提出一种基于卷积神经网络特征表征的语音情感识别方法。设计语音识别模型，将卷积核转换成一维卷积核，实现语音识别。该方法设计的识别模型不够完善，存在识别效果差的问题。庄志豪[4]等人提出一种基于深度自编码器子域自适应跨库语音情感识别方法。利用编码器获取语音源域中的最强表征性数据；利用训练模型对数据特征值进行训练实现语音识别。该方法获取的语音特征误差较大，存在识别率低的问题。田祥宏[5]等人提出一种局部线性嵌入的语音识别方法。该方法利用非线性无监督距离公式对不同的语音数据点的距离进行控制，实现语音识别。该方法的识别结果不够稳定，存在误识率高的问题。白璐[6]等人采用语音语谱图作为汉语单字语音识别的特征，构建基于卷积神经网络的语音识别模型，识别语音。该方法误识率较高。

为了解决上述方法中存在的问题，提高语音识别效果，提出基于PNCC特征的录音语音自动识别方法。为了避免识别期间录音语音被干扰，提高识别精度，创新性

*基金项目：国网湖北省电力有限公司科技项目研究成果（521505210004）
收稿日期：2022-07-18

地利用PNCC方法提取录音语音的特征。采用小波变换方法对录音语音信号进行时频域转换处理，建立录音语音特征序列；利用滤波器对序列进行滤波分析，抑制语音中的非对称噪声；归一化处理录音语音的多通道时频、功率；采用离散余弦变换方法变换录音语音非线性信号序列，提取基于PNCC特征的录音语音特征；创新性地利用人工蜂群算法，将矢量量化后的录音语音特征参数进行迭代，自动获取失真率最低的最优码书，解决了不同设定条件下录音语音误识率高的问题。将最优码书作为搜索索引输入到识别模型中进行训练和识别，实现录音语音自动识别。

1 基于PNCC特征的录音语音特征提取

通常情况下录音语音在识别时，若识别期间录音语音所处的状态为正常状态，那么就不会对录音语音的识别效率造成影响，若识别期间录音语音处于干扰状态，语音的识别精度就会大幅度降低。所以为了避免这种问题的发生，需要利用PNCC方法对录音语音的特征进行提取。

由于录音语音是一种非平稳的信号，所以具体的特征提取流程为：

（1）需要对输入的录音语音信号进行预处理操作，其中包括信号滤波、端点检测等；

（2）录音语音信号预处理后，时频域转换分析预处理后的录音语音数字信号序列，主要表现在：

首先采用小波变换方法对录音语音信号实行时频域转换，从中取得各个信号的小波系数，通过对小波系数的计算取得小波系数频谱，将其从低到高进行排列后拼接，依据拼接结果建立出录音语音特征序列；

（3）计算录音语音特征序列，该计算结果就是功率谱计算结果，用方程表达式定义如下：

$$P(\omega)=\lim_{T\to\infty}\frac{F_T(\omega)^2}{2\pi T} \tag{1}$$

式中，$P(\omega)$描述的是功率谱，ω描述的是录音语音初始信号小波系数频谱，$F_T(\omega)$描述的是录音语音初始信号时频转换后所产生的值，T描述的是录音语音的总帧数；

（4）利用滤波器对语音信号进行滤波分析，这样滤波器所发生的时域冲击响应可以用方程表示如下：

$$g(t)=at^{(n-1)}+e^{-2\pi bt}+\cos(2\pi f_0 t) \tag{2}$$

式中，$g(t)$描述的是在t时刻下的滤波器，b描述的是滤波器带宽，a描述的是系数，n描述的是滤波器阶数，e描述的是冲波响应，f_0t描述的是滤波器函数；

（5）为了取得录音语音的信号频谱序列，需要对录音语音所产生的背景噪声进行补偿，那么这时就要对录音语音的长时帧功率进行计算，而出现的非对称噪声要进行抑制，那么利用方程表达式定义为：

$$\tilde{Q}(s,c)=1/(2s+1)+\sum_{s=1}^{s}P[s,c] \tag{3}$$

式中，$\tilde{Q}(s,c)$描述的是录音语音长时帧功率，$P(s,c)$描述的是某一帧的功率谱，s描述的是录音语音的帧。

根据计算的长时帧功率，将其以非对称滤波器形式进行表示，定义为：

$$\tilde{Q}_{out}(s,c)=\begin{cases}\lambda_a+\tilde{Q}_{out}[s-1,c]-(1-\lambda_a)+\tilde{Q}_{in}(s,c)\\ \lambda_b+\tilde{Q}_{out}[s-1,c]-(1-\lambda_b)+\tilde{Q}_{in}(s,c)\end{cases} \tag{4}$$

式中，$\tilde{Q}_{out}(s,c)$描述的是语音滤波输出，$\tilde{Q}_{in}(s,c)$描述的是语音滤波输入，λ_a、λ_b均描述的是对滤波系数的调整；

（6）对录音语音的多通道时频、功率进行归一化处理，处理完成后再对功率进行调整，具体过程如下所示：

$$\begin{cases}\tilde{M}(s,c)=1/(c_2-c_1+1)+\sum_{c=1}^{c_2}\tilde{N}(s,c)+\tilde{Q}(s,c)\\ T(s,c)=P(s,c)+\tilde{M}(s,c)\end{cases} \tag{5}$$

式中，c描述的是滤波器中的通道数量，c_1和c_2分别描述的是归一化处理前后的通道数量，N描述的是滤波器的平滑通道数量，$\tilde{N}(s,c)$描述的是背景噪声系数，$T(s,c)$描述的是调整后的功率谱值。

对录音语音功率进行归一化的流程为：优先对录音语音样本的平均功率进行估算，利用目前功率减去估算结果，即$\mu(s)$。取得归一化处理后的录音语音功率，用方程定义如下：

$$\begin{cases}\mu(s)=\lambda_\mu T(s,c)+(1-\lambda_\mu)/c+\sum_{c=0}^{C-1}T(s,c)\\ U(s,c)=k\times[T(s,c)/\mu(s)]\end{cases} \tag{6}$$

式中，$U(s,c)$描述的是归一化处理后的功率，k描述的是调整参数，λ_μ描述的样本平均功率滤波调整系数；

（7）基于式(6)对其进行幂函数非线性处理，采用离散余弦变换(Discrete Cosine Transform，DCT)变换方法对录音语音非线性处理信号序列进行变换，这样就能够取得W-PNCC特征的录音语音参数，实现基于PNCC特征的录音语音特征提取。该提取流程如下所示。

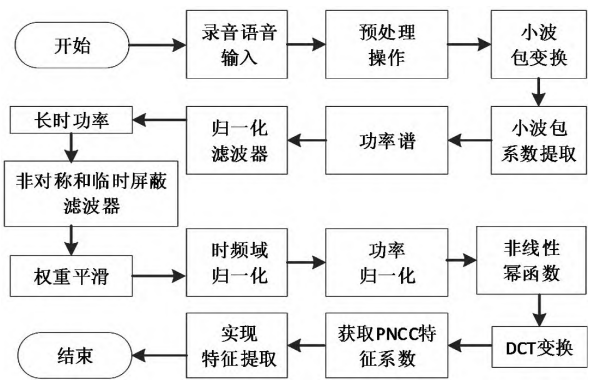

图1 基于PNCC的录音语音特征提取流程

2 录音语音自动识别方法

基于上述提取的录音语音信号特征，利用人工蜂群算法（Artificial bee colony algorithm，ABC）[7]，对录音语音特征参数进行迭代优化，从中取得最佳码书后，将其输入到自动识别模型中进行训练和识别，实现录音语音自动识别。

2.1 人工蜂群

在ABC算法中，负责迭代优化搜索的蜂群与食物源有着密切关联。设置在人工蜂群搜索空间中$x_i(i=1,2,\cdots,SN)$是它的初始解，其中SN描述的是食物源数量。那么进行初始化后，蜂群就会跟随蜂头进行反复搜索，直至达到最大循环次数MCN后停止。通过ABC算法进行矢量量化在最优个体解的空间范围内具有较高的搜索能力，能够避免搜索过早陷入局部最优，收敛速度较快，具有较好的全局搜索性，所以需要利用ABC算法对录音语音特征进行矢量量化，从中获取最佳码书。

2.2 基于ABC矢量量化搜索最佳码书

ABC矢量量化算法可以利用蚁群中的峰头对需要聚类的录音语音特征数据进行搜索，用码书中与特征数据输入矢量最匹配的码字索引代替输入矢量进行传输与存储，而解码时仅需要简单地查表操作。最匹配的码字索引，即最佳码书，最匹配实质为平均失真值最小。ABC矢量量化算法设置搜索最优码书作为适应度目标函数，在食物源中自动获取最优码书，将失真降到最低。

假设第z维的码字大小为f，那么食物源所处的位置就是$z\times f$维。在通过ABC算法矢量量化，实现自动化搜索最优码书之前，需要转换录音语音作为初始化训练矢量，设置最小平均失真值作为搜索的目标函数，计算食物源的选择概率作为目标函数迭代的最大适应度值。

（1）录音语音的初始化

将提取的录音语音特征值转换成训练矢量，该训练矢量由K进行描述，在K中随机选择一个由$z\times f$个矢量构成的码字。采用最近邻法则[8]，对各个矢量中的码书进行规划，并归类相同的码书计算该码书的聚类中心，将其用作原始食物源所处的位置。重复上述流程，就能够得到SN个食物源；

（2）录音语音自动识别的适应度目标

最佳码书即码书中最小平均失真值，因而在ABC算法中，录音语音自动识别的适应度目标函数用方程表达式定义如下：

$$f(j)=\sum_{j=1}^{M}\left(x-y_j\right),j=0,1,\cdots,SN \tag{7}$$

式中，x描述的是矢量量化后的录音语音特征序列最后一个录音语音特征的失真值，$f(j)$描述的是录音语音自动识别的适应度目标函数，j描述的是食物源搜索自动划分的第j个码字空间，y_i描述的是第j个码字空间中最小失真值，M描述的是码字空间数。

通过式(7)可知，$f(j)$的值越小，就说明码书的性能越强。而在ABC算法中，$f(j)$的值越大越好。所以ABC算法下的适应度fit_i表示为：$fit_i=1/1+f(i)$；

（3）食物源的选择概率

为了能够自动获取最优食物源，利用下述方程计算食物源的选择概率，定义如下：

$$p=0.2+0.8\cdot\left(fit_i/\max fit\right) \tag{8}$$

式中，p描述的是选择概率，$max\,fit$描述的是最大适应度值，fit_i描述的是录音语音识别最优解。

在语音初始化训练矢量、自适应目标函数、食物源选择概率的基础上对录音语音特征参数进行矢量量化，自动获取最优码字的流程如下所示：

（1）首先对SN、MCN等多种参数进行设置；

（2）对录音语音进行种群初始化，从中获取到录音语音最优解，即对fit_i进行计算；

（3）依据最近邻法则对录音语音的训练矢量进行确立，并对其进行码书划分，依据划分结果，对码书进行更新，达到最大适应度值后取得最优食物源；

（4）在反复循环更新中，需要判定在操作中是否丢失最优解。若丢失，自动从其他参与划分的码字中取出距离该区域质心最远的码字，将该码字放入没有码字的空区域，然后根据步骤(3)重新生成一个最优解来代替丢失的最优解，保证最优搜索的自动化循环正常运行；

（5）对最佳食物源位置及其相对应的码字进行记录；

（6）满足最小平均失真值目标条件后，就可以对最优码字进行输出，若不满足条件，需要重复返回步骤(3)。

2.3 实现录音语音自动识别

将上节矢量量化后获得的最优码字作为搜索索引，建立一个语音自动识别模型，如图2所示。

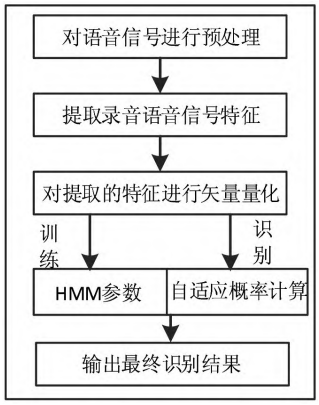

图2 录音语音识别模型

将上述提取的基于PNCC的录音语音特征参数转化成一组随机向量序列，标记为：$K_1,K_2,\cdots,K_T$，即$U(s,c)$

特征在矢量量化后转换后的序列；再把特征参数经矢量量化后的获取的最优码字转换成一组符号序列，即$R=r_1,r_2,\cdots,r_T$。输入$R=r_1,r_2,\cdots,r_T$到录音语音识别模型中，根据HMM参数进行自动识别训练，训练的最大迭代次数即自适应概率最大值，完成训练，输出最终识别结果，以此实现录音语音自动识别。

3 实验与分析

为了验证基于PNCC特征的录音语音自动识别方法的有效性，需要对该方法进行实验对比测试。采用本文方法（方法1）、文献[4]方法（方法2）和文献[5]方法（方法3）进行实验测试。

为了验证本文采用的人工蜂群算法对录音语音特征参数具有较好的优化性能，利用方法1、方法2、方法3分别通过Ackley函数对语音数据聚类特征进行优化测试。测试维数为60维，算法的种群规模设定为100，最大迭代次数为3 000。三种算法优化后的迭代收敛曲线如图3所示。

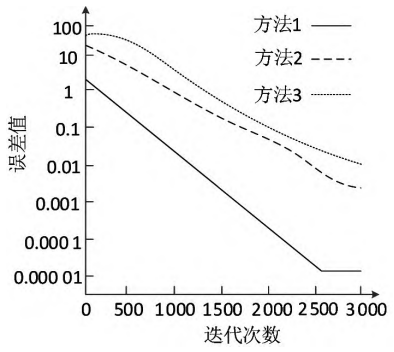

图3 迭代收敛曲线

由图3可知，方法1可收敛到最优解，且收敛速度明显快于其他算法，并且更接近于最优值，说明人工蜂群算法具有更好的收敛性能。

本次实验在中国科学院自动化研究所录制的CASIA汉语情感语料库选择1 050句语音作为训练数据，语料库中剩余的150句语音作为测试数据。利用语音测评软件OpenKWS16进行实验验证，语音特征由软件OpenSMILE进行提取，并将特征归一化到[0,1]，每次识别重复10次取均值以减少误差。根据设立的条件对录音语音自动识别性能进行实验测试。

（1）在实验期间，语音数据获取的输入矢量帧数设置为$p=5$，状态数为$L=3$，语音数据加权指数为$M=4$。

测试方法1、方法2和方法3对聚类类别数c的语音误识效果。误识效果越低，说明该方法的语音自动识别效果越佳，测试结果如图4所示。

分析图4中的数据发现，当$c=2$时，方法1的初始误识率就要低于方法2和方法3，随着聚类类别数c的不断增加，方法1的上升速度较慢，整体误识率最低。这表明聚类类别数c对方法1的识别性能带来的影响小，致使方法1的语音自动识别效果最佳；

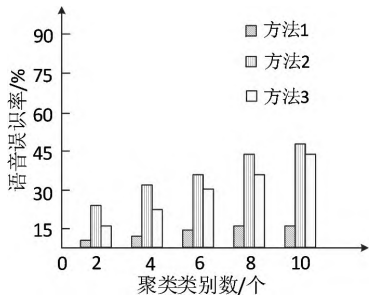

图4 不同聚类类别数下的语音识别误识率测试

（2）在实验测试中，设定$c=2$，输入矢量帧数$p=6$，$M=4$，此次实验主要验证语音数据状态数L对录音语音自动识别性能带来的影响，利用方法1、方法2和方法3分别对不同状态数L下的录音语音误识率进行有效测试，测试结果表1所示。

表1 不同状态数下的语音误识率测试

状态数取值	不同方法的语音误识率/%		
	方法1	方法2	方法3
5	5.0	11.2	12.3
10	6.2	12.2	15.9
15	6.6	12.5	16.9
20	6.9	12.7	19.5
25	7.2	13.2	20.6
30	7.9	13.6	22.8
35	8.2	13.9	23.9
40	8.9	14.9	25.9
45	9.6	15.3	26.7
50	9.9	16.2	28.9
55	10.0	16.9	29.5

根据表1中的数据发现，当状态数L的取值不断变化时，三种方法的误识率均有所提升。但在整体测试中，方法3的初始误识率与最终误识率之间的跨度较大，可以看出方法3的误识率上升速度较快，因而判定L对方法3带来的影响较大，导致方法3的误识率最高、识别效果最差。反观方法1的整体误识率较低，这说明即使L在不断变化，方法1所产生的误识率仍要低于其余两种方法，表明方法1的识别效果强；

（3）设定基础实验参数$c=4$、$L=3$、$M=4$，对输入矢量帧数p进行改变，采用三种方法对不同输入矢量帧数下的语音误识效果进行测试，具体测试结果如图5所示。

输入矢量帧数p的取值不断增加后，三种方法所产生的语音误识率大不相同。从图5中可以看出，方法3的误识率要高于方法1与方法2，同时p值越大，方法3的误识

率越多。经对比发现，方法1的误识率较低，这主要是因为方法1对录音语音特征进行矢量量化，以此提升了录音语音识别效率，降低了该方法的录音语音误识率。

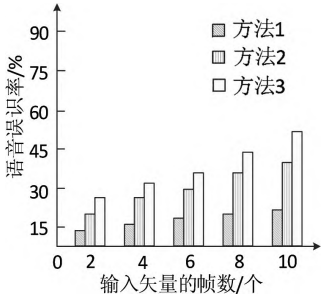

图5 不同输入矢量帧数下的语音误识效果测试

4 结束语

现如今录音语音识别技术已经被人类广泛应用，针对录音语音自动识别方法识别精度不高的问题，提出基于PNCC特征的录音语音自动识别方法。基于PNCC对录音语音的特征进行提取，并利用人工蜂群算法获取到录音语音最佳码书，将其输入到构建的识别模型内，实现录音语音自动识别。实验结果证明，该方法提高了识别精度和识别效率，为录音语音自动识别方法提供了重要信息，在今后录音语音自动识别方法中有着长远的发展前景。在未来的研究工作中，考虑将语音识别解码生成的Lattice用于数据选择，进一步提高语音识别的性能。

参考文献：

[1] 周红锴.基于单片机控制的孤立词语音自动识别系统设计[J].现代电子技术，2020，43(18)：4.

[2] 杨志杰，张梅，李冠龙，等.基于长短时记忆元的语音智能识别系统设计[J].电子设计工程，2020，28(1)：5.

[3] 姜芃旭，傅洪亮，陶华伟，等.一种基于卷积神经网络特征表征的语音情感识别方法[J].电子器件，2020，42(4)：998-1001.

[4] 庄志豪，傅洪亮，陶华伟，等.基于深度自编码器子域自适应的跨库语音情感识别[J].计算机应用研究，2021，38(11)：3279-3282，3348.

[5] 田祥宏.一种结合局部线性嵌入与支持向量机的语音识别方法[J].电视技术，2020，43(2)：61-65.

[6] 白璐，王连明.基于卷积神经网络的大容量汉语孤立字语音识别方法[J].东北师大学报：自然科学版，2020，52(2)：6.

[7] 程龙，张方华.用于混合储能系统平抑功率波动的小波变换方法[J].电力自动化设备，2021，41(3)：100-104，128.

[8] 张佳宁，严冬梅，王勇.基于word2vec的语音识别后文本纠错[J].计算机工程与设计，2020，41(11)：6.

作者简介：肖宜(1985-)，男，硕士，高级工程师，研究方向：智能电网调度。

通信作者：葛罗(1989-)，男，硕士，工程师，研究方向：智能电网调度。

(上接第69页)

tion[J].IEEE Transactions on Pattern Analysis&Machine Intelligence，2017，39(11)：2314-2320.

[6] Redondo C，Baptista M and López-Sastre R J.Learning to exploit the prior network knowledge for weakly supervised semantic segmentation[J].IEEE Trans.Image Process.2019，28(7)：3649-3661.

[7] 郭艳婕，杨明，侯宇超.改进的SLIC算法在彩色图像分割中的应用[J].重庆理工大学学报(自然科学)，2020，34(2)：158-164.

[8] 张蕊，李锦涛.基于深度学习的场景分割算法研究综述[J].计算机研究与发展，2020(26)：1-15.

[9] 王季峥，尹丽菊，咸日常，等.基于改进SLIC算法的电力设备故障区域分割方法[J].计算机应用与软件，2021，38(1)：222-226，237.

[10] 刘连忠，李孟杰，宁井铭.基于改进SLIC的光照干扰下茶树冠层图像分割[J].江苏农业学报，2020，36(4)：1022-1027.

[11] 廖苗，李阳，赵于前，等.一种新的图像超像素分割方法[J].电子与信息学报，2020，42(2)：364-370.

[12] 宋熙煜.基于超像素的图像分割技术研究[D].郑州：解放军信息工程大学，2015.

[13] 韩纪普，段先华，常振.基于SLIC和区域生长的目标分割算法[J].计算机工程与应用，2021，57(1)：213-218.

[14] 韩剑辉，吕郅强.融合FPGA技术的改进SLIC超像素分割算法[J].哈尔滨理工大学学报，2020，25(1)：59-65.

[15] 詹琦梁，陈胜勇，胡海根，等.一种结合多种图像分割算法的实例分割方案[J].小型微型计算机系统，2020，41(4)：837-842.

作者简介：王静(1980-)，女，本科，高级实验师，研究方向：图像识别、计算机软件等。