Computer Science and Application 计算机科学与应用, 2018, 8(11), 1752-1761
Published Online November 2018 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2018.811194

Hans汉斯

Speaker Recognition System Based on Multifractal Spectrum Feature and Characters Selection Policy

Yuhuan Zhou[1], Liang Zhang[2]

[1]Institute of Command Information System, The Army Engineering University of PLA, Nanjing Jiangsu
[2]96733 Unit 74 Detachment of PLA, Huitong Hunan
Email: zhouyuhuan250@sina.com, 52761263@qq.com

Received: Nov. 5th, 2018; accepted: Nov. 15th, 2018; published: Nov. 22nd, 2018

Abstract

Speech is one kind of complicated non-linear signal, so traditional speech or speaker recognition system based on the linear theory is difficult to be further improved. In this paper, a new method based on the WTMM (wavelet transform modulus-maxima method) is proposed, which can facilitate the extraction of speech signals in the multifractal spectrum feature (MSF). The multifractal spectrum feature combined with the traditional linear features can obviously enhance performance of speaker recognition system. Experiment results show that 6-dimensional MSF combined with 13-dimensional MFCC and 16-dimensional LPC make error rate decrease to 1.2% in short speech speaker recognition. Then greedy algorithm is used to select 13 dimensional features from 101-dimensional features set. The experiment results show that the optimal feature selective method can eliminate disturbance of other redundant features, and obviously reduce the error rate, and improve the computational speed. The error rate decreases to 1.6%, and computation time decreases about 86%.

Keywords

Speaker Recognition, Multifractal Spectrum Feature, Wavelet Transform Modulus-Maxima Method, Gaussian Mixture Model, Feature Selection

基于多分形谱及特征优选的说话人识别系统

周宇欢[1]，张　亮[2]

[1]陆军工程大学指挥信息系统学院，江苏　南京
[2]中国人民解放军96733部队74分队，湖南　会同
Email: zhouyuhuan250@sina.com, 52761263@qq.com

文章引用: 周宇欢, 张亮. 基于多分形谱及特征优选的说话人识别系统[J]. 计算机科学与应用, 2018, 8(11): 1752-1761.
DOI: 10.12677/csa.2018.811194

收稿日期：2018年11月5日；录用日期：2018年11月15日；发布日期：2018年11月22日

摘　要

语音是复杂的非线性信号，这使得基于线性理论的传统说话人识别系统性能难以进一步提高。结合语音特点，基于小波极大模方法(Wavelet Transform Modulus-Maxima Method, WTMM)，提出一种语音多分形谱特征(Multifractal Spectrum Feature, MSF)提取方法，并将语音多分形谱特征与传统特征结合用于说话人识别，实验表明，在短语音说话人识别中，6维MSF与LPC结合，误识率相比单独使用LPC降低了6.4个百分点；而MSF与MFCC、LPC组合，误识率降至1.2%左右。采用贪婪策略对说话人识别的特征进行优选，从101维特征中优选出13维特征用于识别，实验结果表明优选后的特征参数能有效降低系统误识率，提高识别速度，误识率最低降至1.6%，识别时间减少约86%。

关键词

说话人识别，多分形谱特征，小波极大模方法，高斯混合模型，特征选择

1. 引言

传统的语音信号处理是基于线性理论的，它基于这样一个假设，即当分段足够小时，非线性系统可以用线性系统来近似，但这种近似必然会损失一些有用的信息，使得基于线性理论的系统性能难以进一步提高。从声学和空气动力学角度看，语音信号既非确定性的，也非完全随机的，而是一个非线性过程，因而目前研究的注意力转向非线性信号分析方法，其中的一个方向就是用分形理论。本文尝试用多分形理论分析语音信号，从而在一定程度上弥补线性理论描述语音信号的不足，基于统计理论的 GMM (Gaussian Mixture Model，高斯混合模型)模型在训练和识别时都需要充分的数据，才能达到较高的识别率，当进行短语音(2 秒左右)说话人识别时，往往难以满足这一要求，为了从有限的数据中，提取出更丰富的信息，以增加模型的识别能力，我们引入了基于分形理论的说话人特征。

上世纪九十年代中后期，语音分形特征已应用于端点检测，语音分割，语音合成，语音编码，语音识别等方面，取得了不错的效果。在说话人识别方面，Jungpa Seo 等人[1]将元音的相关维作为特征参数，能有效区分具有相似声音的说话人；Petry 等人[2]认为传统方法得到的倒谱参数没有反映出系统的非线性动力进化，因此将非线性动力信息——分形维数与倒谱参数结合起来，利用 Bhattacharyya 距离进行说话人识别，改善了系统性能；近年来，在分形特征的提取方式，分形与传统线性特征组合等方面进行了更深入的研究，取得一定的进展[3] [4]。目前的研究主要集中在计算语音的分形维数，提取语音的单分形维数、熵、Lyapunov 参数等作为非线性特征，分形维数的提取方法主要是借鉴图像处理方法，采用盒覆盖的方式，这样提取的分形特征不能全面表征语音的细节信息，提取方式也不符合语音一维信号的特点。

本文将多分形理论引入说话人识别系统，结合语音的时变特性，基于小波极大模方法(Wavelet Transform Modulus- Maxima Method, WTMM) [5]，提出一种语音局部多分形谱的计算方法，并应用于短

语音(2 秒)说话人识别。实验表明，将多分形谱特征与语音线性特征结合，可以有效提高短语音的说话人识别率。另外，由于引入了多分形谱特征，使得特征的总维数增高，影响了系统的实时识别，因此本文还利用贪婪算法，从大量的特征参数中优选出识别性能较好的说话人特征参数，实验表明，优选出的特征用于识别，可以提高了识别速度，而且去除了不良特征的干扰，与相同维数的单类特征相比，识别率有较大提高。

2. 多分形理论

简单的分形维数对所研究的对象只能作平均性的描述，无法反映分形结构全面精细的信息，为此人们提出了多分形的概念[6] [7] [8]。多分形通过一个谱函数来描述不同分形结构混合而成的复杂系统，它能够提供系统更精细的分形信息。多分形谱主要有两种语言描述：$\alpha - f(\alpha)$ 语言和 $q - D_q$ 语言，它们之间可通过多分形热力学公式相互转换，其中 $\alpha - f(\alpha)$ 语言有比较清晰的物理意义，而 $q - D_q$ 语言更适合快速计算。

多分形理论用 $f(\alpha)$ 描述不同单分形结构其数量随尺度的变化：$N(\varepsilon) \sim \varepsilon^{-f(\alpha)}$。图 1 中 b 为钟状曲线，两头的 $f(\alpha)$ 均降到 0，它表示的是以(0.6, 0, 0.4)比率构造的 Cantor 集；c 为向右的钩状曲线，左侧 $f(\alpha)$ 降到 0，右侧降到 0.63，它表示的是以(0.2, 0.6, 0.2)比率构造的 Cantor 集。而 b′和 c′分别是按(0.7, 0, 0.3)和(0.4, 0.2, 0.4)比率构造的 Cantor 集。

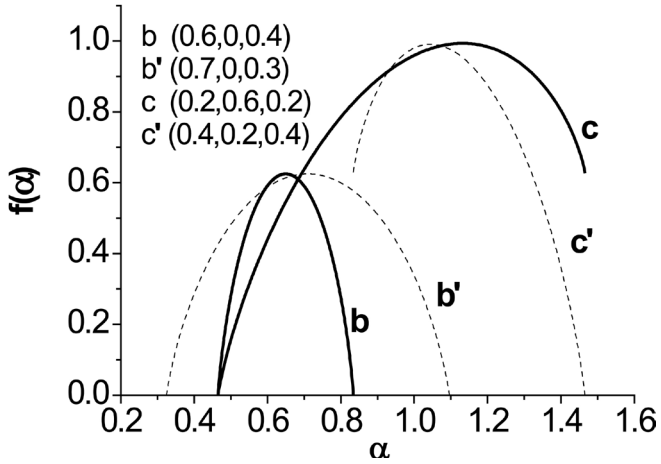

Figure 1. Multifractal spectrum of different Cantor sets
图 1. 不同 Cantor 集的多分形谱图

多分形谱图包含丰富的信息，由图 1 中不同 Cantor 集对应的多分形谱结构，可以分析得到一些有用的信息：

1) 多分形谱中 α 表示的是不同的分形结构。在尺度变化过程中，α 越小表示结构的概率测度越大。

2) 多分形谱中的 $f(\alpha)$ 表示不同分形结构其数量与尺度的对数关系。$f(\alpha)$ 越小说明对应分形结构数量较少。

3) 多分形谱的最大宽度 $\Delta\alpha = \alpha_{\max} - \alpha_{\min}$ 是多分形的一个重要参量，它描述了不同分形结构的差异程度，一般地讲，$\Delta\alpha$ 越大表明该信号包含的分形结构的差异度大。

4) $f_{\max}(\alpha)$ 代表了最主要的分形结构，可以表示信号的平均分形维数，反映了分形体的总体特性。其值越接近 0，表示分形体形态越离散；其值越接近 1，表示分形体形态越接近于线；其值越接近 2，表示分形体形态越接近于面。

3. 语音多分形谱特征的计算

目前分形的计算方式很多，比较常用的是盒维数。但是语音信号是一维信号，将其表示为二维开曲线再利用盒覆盖方法求分形维数，不仅不方便，而且不符合语音一维信号的性质。在进行湍流等复杂系统的多分形谱分析中，最为著名的是由法国学者 A. Arneodo 等人提出的基于小波分析的“小波极大模理论(WTMM)”。

WTMM 可以很方便的提取出一维信号的多分形谱结构。在固定尺度 a 时，对信号作小波变换，得到小波系数 $W_f(x,a)$，在 x 的邻域内，若有 $|W_f(x_0,a)|>|W_f(x,a)|$，则称 x_0 为一个“局部极大值”点，极大值点的连线就是“极大模线族 $l(a)$”。定义配分函数：

$$Z(q,a)=\sum_{x\in l(a)}\left(|W_f(x,a)|^q\right)=a^{\tau(q)} \tag{1}$$

即 $W_f(x,a)$ 在极大模线 $l(a)$ 上求和，得到 q 与 $\tau(q)$ 的关系，$\tau(q)$ 可以从 $Z(q,a)$ 和 a 的双对数曲线中求出。利用多分形热力学公式求取多分形谱：

$$\begin{cases} D(h)=\min\limits_q\left[qh-\tau(q)\right] \\ h=\dfrac{\partial\tau}{\partial q} \end{cases} \tag{2}$$

其中 $h,D(h)$ 的地位相当于测度论中 $\alpha,f(\alpha)$ 的地位。

以上是 WTMM 求取多分形谱的方法，本文结合语音信号特点，构建了一种语音信号多分形特征(MSF)的提取方法。为了使 MSF 与传统的特征在特征域结合，MSF 提取方式也基于帧。经过初步的实验，在进行 MSF 的计算时，语音帧长度的选取非常重要，如果语音帧长度太短，提取的 MSF 就不太准确，而如果语音帧太长，导致 MSF 的特征量变少，由于 GMM 模型训练和识别需要大量的数据，又会导致 GMM 模型训练不准确，综合以上因素，本文对 MSF 取两倍于传统特征的帧长，但是两者的帧移相等，也就是 MSF 的帧计算时会有一半重合，而传统特征帧无重合，这样得到 MSF 和传统特征的长度相同，便于在特征域进行融合。

为了得到特定时间段内的多分形谱特征，先对语音进行分帧，对每一帧分别提取 MSF，从而得到局部的 MSF，其计算流程如图 2 所示。

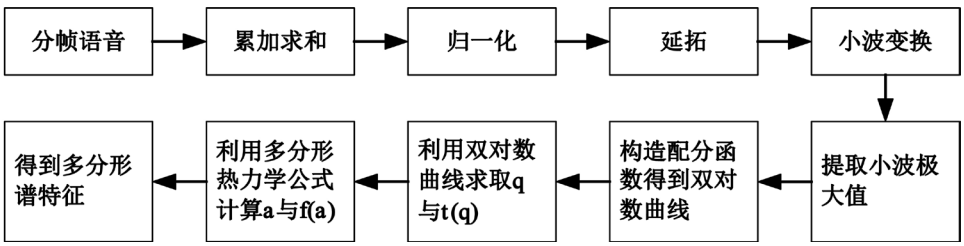

Figure 2. Extraction process of the MSF
图 2. 多分形谱特征提取流程

1) 累加求和。设某段分帧语音信号为：$x(i)$，$i=1,2,\cdots,N$，i 为序号，N 为帧长，利用公式 $s(n)=\sum_{i=1}^{n}x(i)$ 对信号累加求和，其中 n 为序号，与 i 一一对应。

2) 归一化。利用公式 $s(n)=s(n)/s(N)$，对 $s(n)$ 归一化，其中 $s(N)$ 是语音信号 $x(i)$ 累加和。归一化的目的的去除语音波形的幅度对计算的影响。

3) 延拓。延拓的目的是为了准确提取小波系数，特别是信号边缘部分的小波系数，公式如下：

$$\begin{cases} S(n)=s(1), 1<n\le N \\ S(n)=s(n-N), N<n\le 2N \\ S(n)=s(N), 2N<n\le 3N \end{cases} \tag{3}$$

4) 小波变换。对 $S(n)$ 进行小波变换，本文采用墨西哥帽小波。

5) 提取小波极大值。遍历得到的小波系数，提取局部极大值。

6) 双对数曲线。根据公式(1)构造配分函数，其中求和只在局部极大值点处。由配分函数得到关于 $Z(q,a)$ 与 a 的双对数曲线，每一个 q 对应一条曲线。

7) 求取 q 与 $\tau(q)$。计算每条双对数曲线的斜率，用来估计 $\tau(q)$ 的值，从而得到 q 与 $\tau(q)$ 的关系。

8) 求取 α 与 $f(\alpha)$。利用多分形热力学公式(2)求取该语音段的 MSF。

9) 简化 MSF。由于计算出来的 MSF 维数较高，达到 30 维，因此根据本文第二部分对多分形谱图的分析，对 MSF 进行简化，只提取多分形谱中 $f_{\max}(\alpha)$，$\alpha_{\min}$ 和 $\alpha_{\max}$ 三点信息来表征语音多分形信息，此三点信息可以反映语音的主要分形结构，分形结构的分布状态以及分布状态的不对称程度等信息。

4. MSF 与传统特征结合实验

MSF 与传统特征结合进行说话人识别实验，语音数据库选用的是 TIMIT，数据库选用其中 50 人，每个人有 10 句语音，长度均为 2 秒。训练和识别分别选 5 句语音。所有的语音均为 16,000 Hz 采样，16 位精度，提取特征参数时，MSF 的帧长为 32 ms，也就是 512 个采样点，帧移 16 ms (256 个采样点)；而 MFCC (Mel Frequency Cepstrum Coefficient，梅尔频率倒谱系数)等传统特征的帧长和帧移均为 16 ms (256 个采样点)。

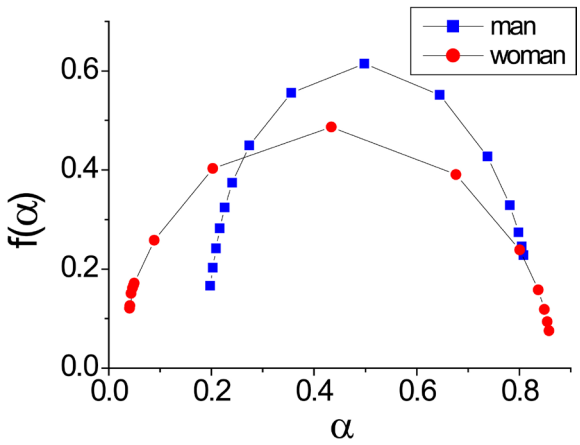

Figure 3. The multifractal spectrum feature of Male and Female
图 3. 男声和女声的多分形谱特征

图 3 显示了男声和女声的多分形谱图，由图可看出男声的 $f_{\max}(\alpha)$ 较大而 $\Delta\alpha$ 较小，说明男声语音相对连续，结构变化较小；女声的 $f_{\max}(\alpha)$ 较小而 $\Delta\alpha$ 较大，说明语音的结构相对离散，结构变化较大。这从另一侧面反映出男声基音频率低，语音平缓；女声基音频率高，语音起伏大的特点。

本文说话人识别的流程如图 4 所示。说话人模型采用 GMM，训练时迭代次数为 10，模型混合度的选取是遍历 1 到 64，选取误识率最低的混合度。实验发现一般情况下，当混合度大小在特征长度一半左右时，可以得到最低的误识率。

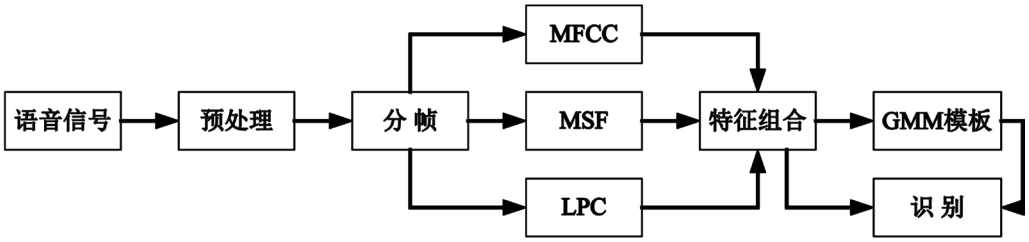

Figure 4. Flow chart of speaker recognition
图 4. 说话人识别流程图

Table 1. False acceptance rate of different parameters combinations
表 1. 不同特征参数及其组合的误识率

特征参数(维数)	混合度	误识率
MSF (6)	3	71.2%
MFCC (13)	7	9.2%
LPC (16)	7	9.6%
MFCC + LPC (29)	13	2.8%
MFCC + MFCC 一阶差分 (26)	13	13.2%
MSF + MFCC (19)	10	8.0%
MSF + LPC (22)	12	3.2%
MSF + MFCC + LPC (35)	15	1.2%

表 1 是不同特征参数及各种组合下的误识率，由表 1 看出，单独使用 MSF 进行识别，效果并不好，误识率达到 71.2%，原因是 MSF 不是针对说话人特征设计的，可能包含了大量与说话人特征无关的信息，比如语义信息，从而导致单独使用 MSF 时对说话人个体的区分性能不好；但是 MSF 与现有语音特征组合时，能大幅提高识别率，这说明 MSF 与线性特征有很好的互补性，MSF 中包含了传统特征丢失的信息，特别是 MSF 与 LPC (Linear Prediction Coefficient，线性预测系数)参数组合时，误识率下降了 6.4 个百分点。而将 MSF 与 MFCC、LPC 组合用于识别，误识率为 1.2%。另外由表 1 也可以看出，并不是增加特征维数就一定可以降低误识率，可以发现，MFCC 与 MFCC 一阶差分相结合在短语音说话人识别中，误识率相比单独用 MFCC 反而有所增加。

5. 特征优选

增加特征的维数，在一定程度上降低了误识率，但是也大大增加了系统的计算负担，MSF 特征有 30 维，全部用于识别，会导致识别时间过长，无法实现实时应用，虽然利用对多分形谱结构的分析，将 MSF 的特征减少到了 6 维，但如果加上其它传统特征，维数仍然很高。

目前，常用的语音特征有基音，共振峰，LPC 及 LPCC (Linear Prediction Cepstrum Coefficient，线性预测倒谱系数)，MFCC，PLP (Perceptual Linear Predictive，感知线性预测)等参数。许多文献[9] [10] [11] [12]通过综合运用加权、微分、组合等方法，进行二次特征提取，在一定程度上提高了系统识别性能，但是随之而来的是过高的系统消耗。另外，有文献通过因子分析、主成分分析[13]、线性判别分析[14]、聚类分析、相关分析等方法选择最优的特征参数组合。但是以上方法所选择的特征维数仍然很高，使系统在实用化方面受到影响。

特征选择[15]在模式识别中方法很多，从特征集合的评价策略上主要可分滤波器(Filter)方法和嵌入式(Wrapper)方法，这两者的区别在于特征的评价准则是否以分类器性能为准则，如果以分类器性能为准则就是 Wrapper 方法，否则就属于 Filter 方法。本文基于 Wrapper 方法，采用启发式搜索策略中的序列前向选择方法(Sequential Forward Selection SFS)，利用贪婪算法，挑选出 13 维左右特征，在不显著增加误识率的前提下，极大提高了系统的识别效率。

5.1. 特征优选算法

假设特征集合中共有 W 维特征，本文优选特征基本策略是：每一轮计算都从特征集合中找到一维最低误识率对应的特征，并在下一轮与特征集合中的其它特征依次组合，计算误识率，每一轮得到的优选特征组合起来成为最终的特征 S。其计算流程如图 5 所示。设有 N 个说话人，每个说话人 GMM 模型参数用 λ_k 表示，其中 $k=1,2,\cdots,N$。特征集合用 T 表示，$T=\{t_1,t_2,\cdots,t_j,\cdots,t_W\}$，其中 t_j 是一维特征矢量，$j=1,2,\cdots,W$。$t_{j,k}$ 表示第 k 个说话人对应的第 j 维特征参数，挑选的特征集合用 $S=\{s_1,s_2,\cdots\}$ 表示，初始时 S 为空。

1) 初始化，令 $i=0$，S 为空，预定达到的误识率为 R，预定挑选特征的数目为 K；

2) 依次取 $j=1,2,\cdots,W$，即从 T 中的取 t_j 与 S 组合，作为训练和识别的特征 τ_j，不同说话人的特征用 $\tau_{j,k}$ 表示；

3) 利用 $\tau_{j,k}$ 训练得到每个说话人的 GMM 参数 $\lambda_{j,k}$，并做识别实验，得到 τ_j 对应的误识率 r_j；

4) j 依次取 1 至 W，得到一组误识率 r_j。令 $i=i+1$，取最小的误识率 $\min\{r_j\}$ 对应的特征 t_J 为此轮挑选的特征 s_i，即 $s_i=t_J$，将 s_i 加入 S 中，即 $S=S+s_i$，并在 T 中去掉 τ_J，即 $T=T-t_J$。

5) 如果 $\min\{r_j\}>R$ 而且 $i<K$，那么回到第(2)步，重复(2)到(4)步，如果 $\min\{r_j\}\le R$ 或者 $i\ge K$，则输出 S。

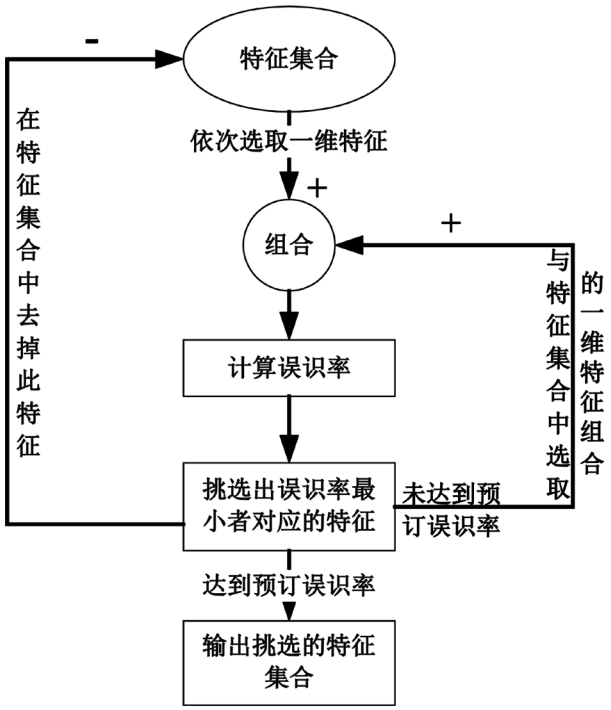

Figure 5. Flow chart of feature selection method
图 5. 特征挑选算法流程图

5.2. 实验结果

实验仍选用 TIMIT 数据库，共 500 句语音，训练 GMM 模板的语音 250 句，其中挑选特征时选用 125 句，测试优选特征识别率时选用另外的 125 句，这样使得挑选特征的数据集与测试特征的数据集不同，以保证结果的可信性。

初始特征集合包括：30 维 MSF，13 维 MFCC，13 维 MFCC 一阶差分，13 维 MFCC 二阶差分，16 维 LPC 和 16 维 LPCC，共 101 维，依次用：MSF1...MSF30；MFCC1…MFCC13，MFCC_D1… MFCC_D13；MFCC_ DD1…MFCC_DD13；LPC1…LPC16；LPCC1… LPCC16 表示。目标误识率预定为 0%，最大挑选特征集合维数为 13。之所以设定为 13，是为了便于与其它特征的误识率进行比较，另外实验发现，当挑选的特征维数超过 13 维时，误识率的降低变得不明显。

Table 2. False acceptance rate of the selection features from 101 dimensional features set
表 2. 从 101 维特征中依次挑选出的特征及误识率

特征	LPC6	LPCC9	MSF17	LPC3	MFCC12	MFCC11	LPC15
误识率	84%	57.6%	43.2%	30.4%	18.4%	11.2%	7.2%
特征	LPC9	MSF11	MFCC_DD10	LPCC1	LPC12	MFCC_D8	
误识率	5.6%	4.8%	4.0%	3.2%	2.4%	1.6%	

由表 2 可以看出，挑选的特征中，LPC 特征有 5 维，MFCC，LPCC，MSF 分别有 2 维，其余特征各 1 维。当选用前 7 维特征时，误识率已经低于单独使用 13 维 MFCC 或 16 维 LPC 的误识率，达到 7.2%；当选用前 12 维特征时，误识率低于 29 维 MFCC 和 LPC 特征组合的误识率，达到 2.4%；选用前 13 维特征时，误识率达到 1.6%，略高于 35 维 MSF+MFCC+LPC 组合的误识率。

从 101 维特征中挑选 13 维特征仍然十分耗时，观察挑选特征的分布，只从 MSF，MFCC，LPC 三种特征，共 59 维特征中优选特征，目标误识率预定为 0%，最大挑选特征集合维数为 13。

Table 3. False acceptance rate of the selection features from 59 dimensional features set
表 3. 从 59 维特征中依次挑选出的特征及误识率

特征误识率	LPC6 84%	LPC3 59.2%	LPC10 40%	MSF13 24.8%	LPC15 19.2%	LPC8 13.6%	LPC4 9.6%
特征	MSF27	MFCC7	LPC13	MFCC10	MSF2	MFCC12	
误识率	5.6%	4.0%	3.2%	3.2%	2.4%	2.4%	

由表 3 可以看出，从 59 维特征集合中挑选的特征，LPC 特征有 7 维，MFCC，MSF 分别有 3 维。与表 2 比较，虽然挑选出的特征有所不同，但对应维数的识别率相差不大，最高识别率略低于表 2。

优选特征的计算复杂度与特征集合的总特征维数有关，维数越高，计算耗时越长，本文最后设计一种特征优选方法，将特征集合一分为二，分别优选出一组特征，然后再结合起来进行识别。本文实验是将 59 维特征集合分成 30 维 MSF 集合和 29 维线性特征集合，并运用贪婪策略分别优选特征 3 维和 10 维，结果如表 4，表 5 所示，然后结合成 13 维特征进行识别，最终误识率为 3.2%。

Table 4. False acceptance rate of the selection features from 30 dimensional MSF
表 4. 从 30 维 MSF 中依次挑选出的特征及误识率

特征	MSF19	MSF11	MSF13
误识率	88.8%	76%	74.4%

Table 5. False acceptance rate of the selection features from 29 dimensional MFCC+LPC
表 5. 从 29 维 MFCC+LPC 中依次挑选出的特征及误识率

特征	LPC6	LPC3	LPC10	MFCC3	MFCC9
误识率	84%	59.2%	40%	28%	20%
特征	MFCC4	MFCC12	LPC13	MFCC6	LPC9
误识率	16%	11.2%	10.4%	6.4%	6.4%

5.3. 计算时间分析

在训练阶段，GMM 包括：去均值，计算协方差矩阵，计算正交转换矩阵，正交化变换训练特征，初始化高斯分布，EM 算法重估高斯混合模型的参数等，计算复杂度为：

$$O\left(L*d^2\right)+O\left(d^3\right)+O\left(d*M\right)+O\left(I*L*d*M\right)+O\left(I*L*M^2\right)$$

其中，d 为特征参数的维数，M 为混合阶数，I 是 EM 过程的迭代数，L 是与语音长度相关的帧数。

在识别阶段，GMM 需要计算每一个人的模型得分，计算复杂度为：

$$O\left(N*L*d^2\right)+O\left(N*L*d*M\right)$$

其中 N 为总的说话人数量。

分析可知，特征维数 d 对于计算时间的影响非常大。因此，与十几维甚至几十维的特征参数相比，本文选用的 13 维特征参数可以将训练和识别的速度大大提高，同时也减少每个说话人 GMM 模型的存储空间。实验显示在 MATLAB 平台上，用 13 维特征训练和识别一个说话人的平均时间大约为 10 秒，而用 35 维 MSF + MFCC + LPC 特征训练和识别一个说话人的平均时间大约为 72 秒。因此，优选的特征在不明显增加误识率的基础上，能大大提高计算速度。

6. 结论

语音信号具有多分形特征，可以将之与传统特征组合，进一步提高识别率。本文利用 WTMM，结合语音信号的特点，提出一种计算语音多分形特征的方法，并用于说话人识别实验，由实验结果可以看出：将 MSF 与传统的线性特征组合，可以不同程度的提高识别率，实验发现，MSF 与 LPC 互补性强，两者组合相比单独使用 LPC 误识率降低了 6.4 个百分点，而将 MSF 与 MFCC、LPC 组合用于识别，误识率降低了 1.6 个百分点，达到 1.2%。

特征的维数对计算时间影响很大，利用 MSF 与线性特征组合，虽然可以达到很高的识别率，但计算时间过长，不能满足实时要求。本文运用贪婪策略，从大量的特征中，挑选出若干维较优的特征参数组合，使得在减少模型存储空间和计算时间的同时，识别率能够满足实用要求。实验发现，对于一个中等大小的说话人集合，优选的 13 维特征参数，其误识率为 1.6%，识别时间相比 35 维特征减少了约 86%。另外本文进一步研究了特征集合的设计和优选策略，实验了不同特征集合下挑选的特征，并采用二分特征集合，以减少单个特征集合的维数，从而有效降低了特征挑选的计算复杂度。

MSF 的计算复杂度比较高，小波变换和利用配分函数求取 MSF 还没有比较成熟的快速算法，这大大影响了特征提取的计算时间，因此下一步研究重点是简化算法复杂度，开发快速算法。

融合策略有很多种，本文只是将 MSF 与传统特征在特征域上简单的组合，下一步考虑赋予各个特征不同的权值，以及引入 PCA、LDA 等特征优选方法进一步提高特征组合的效果，或者可以考虑在模型域、决策域上进行融合。

贪婪策略容易局部收敛，下一步考虑引入遗传算法等全局性算法优选特征，使得特征优选的计算时间更少和准确性更高。

基金项目

江苏省自然科学基金青年基金面上资助项目(BK20140075)；中国博士后科学基金第八批特别资助项目(2015T81081)；第 54 批中国博士后面上项目一等资助(2013M542425)；江苏省自然科学基金青年基金面上资助项目 (BK20140073)。

参考文献

[1] Seo, J.P., Kim, M.S., Baek, I.C., *et al.* (2004) Similar Speaker Recognition Using Nonlinear Analysis. *Chaos*, *Solitons and Fractals*, **21**, 159-164.

[2] Petry, A. and Barone, D.A.C. (2002) Speaker Identification Using Nonlinear Dynamical Feature. *Chaos*, *Solitons & Fractals*, **13**, 221-231. https://doi.org/10.1016/S0960-0779(00)00260-5

[3] Fan, Y.L., Yi, L. and Tong, Q.Y. (2008) Speaker Gender Identification Based on Combining Linear and Nonlinear Features. 7*th World Congress on Intelligent Control and Automation*, Chongqing, 25-27 June 2008, 6739-6744.

[4] Hou, L.M. and Wang, S.Z. (2004) Generalized Dimensions Applied to Speaker Identification. *Biometric Technology for Human Identification*, Orlando, FL, 12-13 April 2004, 555-560. https://doi.org/10.1117/12.542828

[5] Arneodo, A., Audit, B., Bacry, E., *et al.* (1997) Thermodynamics of Fractal Signals Based on Wavelet Analysis: Application to Fully Developed Turbulence Data and DNA Sequences. *Physica A*: *Statistical Mechanics and its Applications*, **254**, 24-45. https://doi.org/10.1016/S0378-4371(98)00002-8

[6] Kestener, P. and Arneodo, A. (2008) A Multifractal Formalism for Vector-Valued Random Fields Based on Wavelet Analysis: Application to Turbulent Velocity and Vorticity 3D Numerical Data. *Stochastic Environmental Research and Risk Assessment*, **22**, 421-435. https://doi.org/10.1007/s00477-007-0121-6

[7] 李彤, 商朋见. 多重分形在掌纹识别中的研究[J]. 物理学报, 2007(8): 4393-4400.

[8] 叶吉祥, 王聪慧. 多重分形在语音情感识别中的研究[J]. 计算机工程与应用, 2012, 48(13): 186-189.

[9] 刘婷婷. 基于因子分析的与文本无关的说话人辨认方法研究[D]: [硕士学位论文]. 合肥: 中国科学技术大学, 2014.

[10] 张庆芳, 赵鹤鸣, 龚呈卉. 基于因子分析和特征映射的耳语说话人识别[J]. 数据采集与处理, 2016, 31(2): 362-369.

[11] 张翔. 基于因子分析的鲁棒性说话人识别技术研究[D]: [博士学位论文]. 北京: 中国科学院研究生院, 2011.

[12] 徐利敏, 唐振民, 何可可, 等. 基于加权特征补偿变换的说话人识别仿真研究[J]. 系统仿真学报, 2008, 20(3): 616-619.

[13] 俞一彪, 芮贤义, 许允喜. 说话人语音特征子空间分离及识别应用[J]. 电路与系统学报, 2008(1): 7-11.

[14] Kim, M.-S., Yu, H.-J., Kwak, K.-C., *et al.* (2006) Robust Text-Independent Speaker Identification Using Hybrid PCA & LDA. *Mexican International Conference on Artificial Intelligence*, Mexico, 13-17 November 2006, 1067-1074.

[15] 毛勇, 周晓波, 夏铮, 等. 特征选择算法研究综述[J]. 模式识别与人工智能, 2007(2): 211-218.