

嗓音多频带非线性分析的声带病变识别*

周 强¹ 张晓俊¹ 顾济华¹ 赵鹤鸣² 朱俊杰¹ 陶 智^{1†}

(1 苏州大学物理科学与技术学院 苏州 215006)

(2 苏州大学电子信息学院 苏州 215006)

2012 年 9 月 21 日收到

2013 年 3 月 19 日定稿

摘要 提出了一种嗓音多频带非线性分析的声带病变识别方法,以提高声带病变嗓音的识别率。首先采用 Gammatone 听觉滤波器组对嗓音信号进行滤波,求取每个频带下的最大李雅普诺夫指数;对映射到核空间的数据采用高斯最大似然度准则优化核函数,然后采用优化核主成分分析算法实现特征抽取。识别实验表明,多频带最大李雅普诺夫指数的识别率比传统的 MFCC 和最大李雅普诺夫指数分别有 6.52% 和 8.45% 的提高,且采用优化核主成分分析算法比传统核主成分分析算法有更好的抽取效果。将多频带非线性分析和优化核主成分分析算法结合,识别率提升至 97.82%。

PACS 数:

Vocal cords diseases detection by multi-band nonlinear analysis of voice

ZHOU Qiang¹ ZHANG Xiaojun¹ GU Jihua¹ ZHAO Heming² ZHU Junjie¹ TAO Zhi¹

(1 School of Physical Science and technology, Soochow university Suzhou 215006)

(2 School of Electronics and Information Engineering, Soochow University Suzhou 215006)

Received Sept. 21, 2012

Revised Mar. 19, 2013

Abstract In order to improve the recognition rate of pathological voices caused by disease of vocal cords, multi-band nonlinear analysis is proposed. Gammatone filter bank is applied to voice signal for front-end time-domain filtering, and then calculate the largest Lyapunov exponent of every band. Data is first mapped into kernel space and use Gaussian maximum likelihood rule to get the best parameter for kernel, which is used for kernel principal component analysis to extract feature. The proposed feature achieves higher recognition rate of 6.25% and 8.45% than MFCC and the largest Lyapunov exponent respectively. When the proposed kernel function is used for kernel principal component analysis, it achieves better performance than traditional function. Ultimately, we get recognition rate of 97.82% by combing them.

引言

病理嗓音的临床表现主要体现为嗓音的音质异常,从病因角度可分为功能性、器质性和神经性。声带病变属于器质性,临床常见有声带息肉、声带小结和声带水肿等疾病。如喉镜一类的电子仪器诊断方法的缺点是,肉眼很难捕捉发音瞬间,且会给病人带来不便,导致不准确的诊断结果^[1]。嗓音的声学分析是非侵入诊断,具有快速、客观、可远距离检测

等优点。因此,该方法应用在医学领域具有非常大的意义。

嗓音分析系统广泛地采用声学参数^[2-3]包括平均基频(F_0),基频微扰(Jitter),基频微扰商(PPQ),振幅微扰(Shimmer),振幅微扰商(APQ)。上述参数的有效性仅限于嗓音为周期或者准周期信号^[4]。20 世纪 80 年代,Teager 发现当气流通过伪声带和声带之间的腔体时会产生涡流^[5],这种非线性与声带的振动情况有关^[6]。在以后的研究中语音研究的专家学者也证实了在语音产生的过程中,存在涡流这种非

* 国家自然科学基金(61271359, 61071215)和苏州大学捷美生物医学工程仪器联合实验室项目资助

† 通讯作者:陶智, Email: taoz@suda.edu.cn

线性现象^[7]。因此,非线性时间序列分析技术常用来提取嗓音信号的特征。Jiang^[8]验证喉切除患者嗓音与正常嗓音的最大李雅普诺夫指数 (Largest Lyapunov Exponent, LLE) 和关联维 (Correlation Dimension, CD) 参数有显著性差异。Zhang et al^[9]将 CD 与 Jitter, Shimmer 结合起来进行单侧声带麻痹识别,实验结果表明识别率比传统参数有明显的改善。基于人耳听觉特性的 MFCC 参数广泛地应用于病理嗓音检测系统,并被证明其有效性^[10-11],而 Gammatone 滤波器组可更加精准模拟人耳的听觉特性,并被应用于语音识别^[12]、说话人识别^[13]与语种识别^[14]等系统中。将 Gammatone 滤波器组应用到声带病变患者嗓音检测系统,改进 MFCC 参数得到 Gammatone 滤波倒谱系数 (Gammatone Filtering Cepstrum Coefficient, GMCC)。

基频的影响因素除了与声带本身特征,如长度、质量和弹性等有关外,还与弹力和黏膜的黏液特性、硬度和湿度有关,也受外在环甲肌、甲构肌及声门下压的影响^[15]。基频类参数在反映声带病变程度时会受诸多其它因素的影响。与强度有关的衍生参数主要是振幅微扰,它与基频微扰均是发声系统稳定性的短期指标,是反映声带振动稳定的无意识指标。随着年龄的增长,不稳定性也会增加^[16],故振幅类参数并不能准确地反映出声带病变对嗓音的影响。声带病变使声带振动负荷增加,干扰了振动的规律性,从而对嗓音信号的非线性特性产生影响。但是由于嗓音信号产生有诸多不确定因素,并不能把整个嗓音信号的产生看成单一的非线性过程^[17]。封子阳^[18]的声带振动模型研究指出,声带各部分振动不尽相同,对各频率的影响也不一样。因此本文提出了多频带非线性分析的特征:多频带最大李雅普诺夫指数 (Multi-Band Largest Lyapunov Exponent, MBLLE)。它可细致地体现各频带下的非线性特性,并能反映出人耳的听觉感知特性。

不同维数间具有相关性,会导致特征冗余,进行特征抽取很有必要。核主成分分析 (Kernel Principal Component Analysis, KPCA) 是近年来发展起来的一种特征抽取技术,并被成功应用于模式分析和模式识别中^[19]。而核函数的形式及参数对 KPCA 的性能有很重要的影响,因此核函数的形式和参数的选取显得尤为重要。交叉验证是常用的核参数调节方法^[20-21],但是存在计算量大和数值的选取需要经验的缺点。Keerthi^[22]采用最小化半径-间隔界的算法进行参数选取,但存在着算法复杂且效率低的缺点。Olivier Chapelle^[23]提出采用梯度下降法解决核

函数多个参数的选取问题,但当偏导数不存在或难以求解时,此方法将失效。针对上文提到的核参数选取存在的缺点,本文提出了基于最大高斯似然度的核参数确定算法,并用到组合核函数中得到优化核 (Majorization Kernel, MK),实验结果表明采用 MK 的 KPCA 的有效性。

本文提出多频带非线性分析算法提取嗓音的 MBLLE 参数,细致地刻画声带病变对嗓音信号各频带非线性影响。提出优化核主成分分析算法抽取有显著性差异频带的非线性特性,进而提高了声带病变嗓音的识别率。

1 听觉感知特性的多频带非线性分析

1.1 Gammatone 滤波器组

Gammatone 函数最早由 Johannesma 提出^[24],该滤波器的时域脉冲响应可以表示为:

$$h_i(t) = Gt^{s-1} \exp(-2\pi B_i t) \cos(2\pi f_i t + \phi_i) U(t), \quad (1)$$

$$(1 \leq i \leq I)$$

其中 G 表示输出增益, s 是滤波器的阶数,决定了边缘处的斜度, B_i 表示滤波器的带宽, f_i 是滤波器的中心频率, ϕ_i 是相位, $U(t)$ 为阶跃函数, I 为滤波器的个数。当 $s = 3, 4, 5$ 时, Gammatone 滤波器可精确地模拟人耳的听觉特性。

一般情况下选择听觉临界频带的中心频率作为 Gammatone 滤波器的中心频率,然后根据语音信号的采样频率来确定滤波器的个数。根据式 (1) 的冲激响应函数,即可得到滤波器的频率响应特性。Gammatone 函数是一个最大幅度出现在中心频率位置的带通滤波器,不同中心频率的 Gammatone 滤波器具有不同的带宽。带通滤波器中心频率两侧都具有较陡的边沿,表明 Gammatone 滤波器具有尖锐的频率选择特性,集中表现中心频率附近的语音特征。但同时两边过渡圆滑,又减少了频谱的泄露。这些幅频响应特性与基底膜的滤波特性是一致的。

采用 Gammatone 滤波器组对 MFCC 参数改进得到 GMCC。嗓音信号经过 Gammatone 滤波器组后,对每个频带信号分帧、加窗,然后进行 DFT 变换并求和得到每个频带的能量 $EN(i)$,最后进行对数处理和 DCT 变换:

$$C_{GMCC}(p) = \sum_i^N \lg EN(i) \cos \left[\left(i - \frac{1}{2} \right) \frac{p\pi}{N} \right], \quad (2)$$

其中 p 为阶数, $p = 1, 2, \dots, P$, P 一般取 12~15。

1.2 多频带非线性特征提取

在非线性动力系统中,李雅普诺夫指数描述了

相空间相邻轨道的平均指数发散率和平均指数收敛率^[25], 其中最大的指数称为 LLE, 可以定量地表示非线性动力系统在其相空间中的稳定状态, 如表 1 所示。

表 1 LLE 反应相空间吸引子的特征

LLE	稳定状态
$\lambda < 0$	稳定的不动点
$\lambda = 0$	稳定的极限环
$\lambda > 0$	混沌

LLE 的定义如下:

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{|\delta \mathbf{X}(t)|}{|\delta \mathbf{X}_0|}$$

(3)

其中 $|\delta \mathbf{X}_0|$ 表示相空间中两个轨道之间的初始距离, $|\delta \mathbf{X}(t)|$ 表示 t 时刻后的距离。关于 LLE 的计算方法有很多, 本文采用的是 Wolf 算法^[26]。

(1) 对离散时间序列 $x_1, x_2, x_3, \dots, x_N$ 采用虚假邻点法确定重构维数 m , 平均互信息法确定延迟时间间隔 τ , 重构相空间 $\mathbf{X}(t) = (x_t, x_{t-\tau}, \dots, x_{t-(m-1)\tau})$, 相点数为 $N_m = N - (m - 1)\tau$ 。

(2) 以初始相点 $\mathbf{x}_0$ 为基点, 在点集中选取一个与 $\mathbf{x}_0$ 最近的点 $\mathbf{x}_1$ 作为端点, 构成初始向量, $\mathbf{x}_0, \mathbf{x}_1$ 间欧氏距离可记为 $L(t_0)$ 。

(3) 时间步长或演化时间 k , 初始向量沿轨线向前演化得到一新向量, 其相应点与端点间欧氏距离可记为 $L(t_1)$, 在相应时段内系统线度指数增长率记为:

$$\lambda = \frac{1}{k} \ln \frac{L(t_1)}{L(t_0)},$$

(4)

(4) 如此继续, 直至所有相点, 然后取各指数增长率的平均值为最大 Lyapunov 指数估计值:

$$\lambda = \frac{1}{N_m} \sum_{i=1}^{N_m} \frac{1}{k} \ln \frac{L(t_i)}{L(t_{i-1})}.$$

(5)

嗓音信号通过 Gammatone 滤波器组后, 分别对每个频带信号求取最大李雅普诺夫指数。图 1 给出了声带水肿 (左) 和正常嗓音 (右) MBLLE 参数的提取过程的实例。例举两嗓音在第 1 和 19 频带有显著性差异, 而第 7 和 24 频带无显著差异, 第 14 频带有一定差异性。因此, 特征抽取很有必要, 可提取有显著性差异频带的信息。

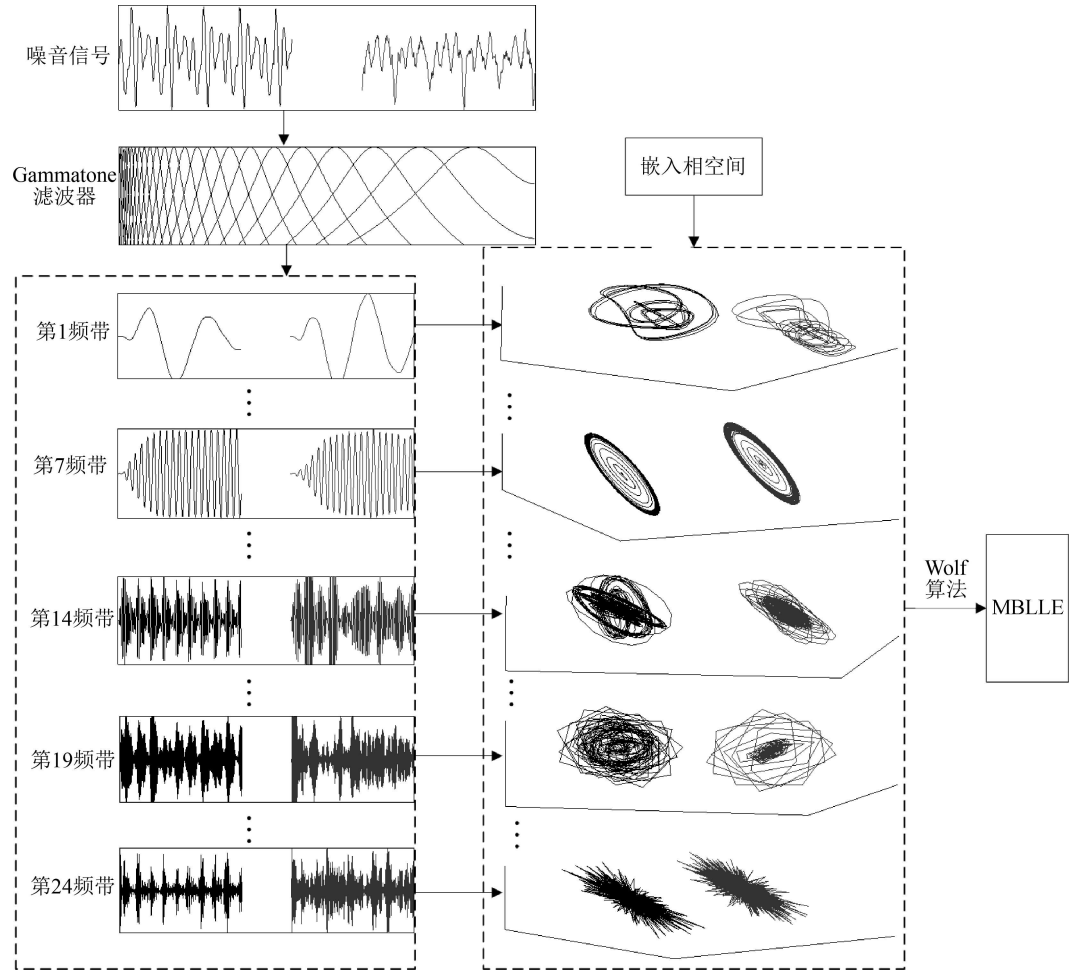


图 1 参数 MBLLE 的提取过程

2 KPCA 特征优化

2.1 KPCA 介绍

传统 PCA 算法是基于原始特征的一种线性变换,当原始数据存在非线性属性时,使用 PCA 后的显著成分却不能反映这种非线性属性。

KPCA 的基本思想是通过一个非线性变换把输入数据映射到核空间,再利用 PCA 进行特征提取^[27]。语音信号中含有相当部分的非线性成分,因此 KPCA 方法将优于 PCA。

设 $\mathbf{x}_i \in \mathbf{R}^N$ ($i = 1, 2, \dots, n$) 为样本数据向量,把输入空间 $\mathbf{R}^N$ 通过非线性变换 ϕ 投影到空间 $\mathbf{F}$, 即 $\mathbf{x}_i \rightarrow \phi(\mathbf{x}_i)$ 。然后求其协方差矩阵 $\mathbf{C}$ 的特征值 λ 和特征向量 μ :

$$\mathbf{C}\mu = \lambda\mu, \quad (6)$$

上式的解应在由 $\phi(\mathbf{x}_i)$ 组成的子空间内,故存在系数 $\alpha_1, \alpha_2, \dots, \alpha_n$, 使得 $\mu = \sum_{i=1}^n \alpha_i \phi(\mathbf{x}_i)$, 从而转换为:

$$n\lambda\alpha = \mathbf{K}\alpha, \quad (7)$$

其中:

$$\mathbf{K} = (k_{ij})_{n \times n} = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle_{n \times n}.$$

对式 (7) 求解得到特征值, 选取正的并按从大到小排列 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, 与对应的特征向量 $\alpha^1, \alpha^2, \dots, \alpha^p$ 。通过设定一定的阈值 θ ($0 < \theta \leq 1$), 使得 $\sum_{i=1}^q \lambda_i / \sum_{i=1}^p \lambda_i \geq \theta$, 便可以从挑选 q ($q < p$) 主成分。

样本在特征空间 $\mathbf{F}$ 的特征向量 μ^m 上的投影为:

$$\mathbf{y} = \langle \phi(\mathbf{x}) \cdot \mu^m \rangle = \frac{1}{\sqrt{n\lambda_m}} \sum_{i=1}^n \alpha_i^m k(\mathbf{x}_i, \mathbf{x}), \quad (8)$$

$$(m = 1, 2, \dots, q)$$

而当 $\sum_{i=1}^n \phi(\mathbf{x}_i) \neq \mathbf{0}$ 时, $\mathbf{K}$ 应由 $\widehat{\mathbf{K}} = \mathbf{K} - \mathbf{I}\mathbf{K} - \mathbf{K}\mathbf{I} + \mathbf{I}\mathbf{K}\mathbf{I}$ 代替, 其中 $\mathbf{I}$ 为所有元素都为 $1/n$ 的 $n \times n$ 矩阵。

2.2 优化核函数

常用的核函数有多项式、径向基 (Radial Basis Function, RBF) 函数和 Sigmoid 函数。当选择了一种核函数后, 也就选定了一种学习模型, 而一个学习模型性能的好坏是由学习能力和推广能力 (对未知样本的预测能力) 两方面决定的, 其对应于插值理论就是核函数的插值能力和外推能力。为了建立一个既有较好插值能力又有较好外推能力的学习模型, 可采用组合核函数形式, 由局部核函数与全局核函数共同组成^[28], 本文采用如下形式:

$$k(\mathbf{x}, \mathbf{y}) = \theta \langle \mathbf{x} \cdot \mathbf{y} \rangle^d - (1 - \theta) \|\mathbf{x} - \mathbf{y}\|^s + 1, \quad (9)$$

$$(d \in \mathbf{N}, 0 < s \leq 2)$$

其中 θ 为比例系数, d 和 s 均为幂指数。而核参数的选择也尤为重要, 如果选择不当, 可导致异类间数据在映射的核空间严重混叠。

对于 KPCA 而言, 总要假设映射数据 $\{\phi(\mathbf{x}_i)\}_{i=1}^n$ 服从高斯分布, 因为只有这样不相关才等同于统计独立。本文提出一种基于高斯最大似然度的核参数选取算法, 并用到组合核函数中得到 MK:

Step1: 对于给定的核参数 θ , 求解训练样本集对应的数据集 $\{\mathbf{y}_i\}_{i=1}^n$;

Step2: 求取数据集的均值 $\bar{\mathbf{y}}$ 和协方差矩阵 Σ , 从而得到目标高斯分布函数:

$$b(\mathbf{y}) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \bar{\mathbf{y}})^T \Sigma^{-1} (\mathbf{y} - \bar{\mathbf{y}}) \right\}.$$

Step3: 求取对数似然函数

$$J_\theta = \sum_{i=1}^n \log p(\mathbf{y}_i | \bar{\mathbf{y}}, \Sigma)$$

来表征高斯性程度;

Step4: 按照 $\hat{\theta} = \arg \max J_\theta$ 确定核参数;

Step5: 确定 MK:

$$k(\mathbf{x}, \mathbf{y}) = \hat{\theta} \langle \mathbf{x} \cdot \mathbf{y} \rangle^2 - (1 - \hat{\theta}) \|\mathbf{x} - \mathbf{y}\|^{1/2} + 1.$$

3 实验及分析

3.1 实验数据

本实验采用 MEEI (Massachusetts Eye and Ear Infirmary) 数据库^[29] 中的样本, 该数据库包含 1384 例病理嗓音 /a/, 给出了患者的年龄、性别、是否抽烟等信息, 并给出了诊断结果。本文选取嗓音具体统计数据见表 2。正常嗓音的采样频率为 50 kHz, 病理嗓音为 25 kHz, 对正常嗓音进行滤波和降采样统一采样频率为 25 kHz。

从表 2 中选取 53 例正常嗓音与 53 例声带病变嗓音 (声带小结 9 例, 声带息肉 12 例, 声带水肿 16 例, 声带麻痹 16 例), 对 MBLLE 参数采用 SPSS17.0 进行配对样本 t 检验, 其中置信区间设为 95%。图 2 给出了各频带下的 Sig(双侧) 值, Sig 值越小表明差异越显著。可以看出 MBLLE 参数在部分频带的差异性较显著, 这对识别有重要的贡献, 而其它频带则并没显著性差异。可以看出声带病变对嗓音各频带非线性的影响存在着差异性。而且说明特征的抽取很有必要。

表 2 嗓音的统计情况

嗓音类别	男		女	
	人数	年龄均值, 标准差 (年)	人数	年龄均值, 标准差 (年)
正常	21	38.8, 8.5	32	34.2, 7.9
声带小结	2	46.5, 4.8	7	40.3, 4.7
声带息肉	9	37.1, 9.5	3	38.7, 8.5
声带水肿	8	45.2, 10.1	24	41.2, 8.4
声带麻痹	16	45.5, 10.6	16	43.8, 8.2
病理合计	35	43.3, 10.5	50	41.7, 7.8

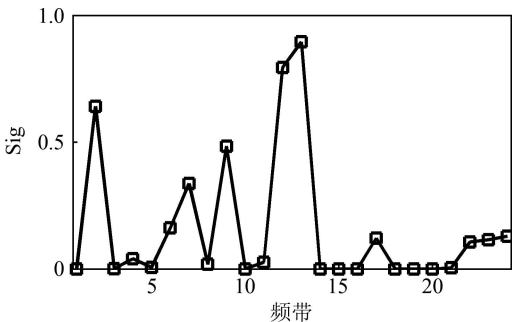


图 2 各频带差异性分析

3.2 实验环境及设置

Gammatone 设置为 10-12500Hz 范围内的 24 个滤波器组。在提取参数 MFCC 和 GMCC 参数时采用 Hamming 窗，窗长设为 20 ms，帧移为 10 ms。阶数 P 取为 12，加上差分系数后特征维数为 24，对嗓音样本每一帧进行提取，然后求所有帧的平均值作为此嗓音的特征。在提取参数 MBLLE 过程中采用的是截取后时长为 160ms 的嗓音信号。为了更好地进行比较，本文采用了三种分类器进行实验：支持向量机 (Support Vector Machine, SVM)、朴素贝叶斯 (NativeBayes) 和最佳优先树 (Best First Tree, BFTree)。

本文在采用 KPCA 特征抽取时采用了不同的核函数进行对比实验，如表 3 所示。

3.3 实验评价指标

受试者工作特征曲线 (Receiver Operating Characteristic curves, ROC) 被用来形象地描述识别的效果。ROC 曲线^[30]的横坐标是假阳性率 (正常嗓音

被识别为声带病变患者嗓音占正常嗓音总数的百分比)，纵坐标为真阳性率 (被正确识别的声带病变患者嗓音占总声带病变患者嗓音的百分比)。而且 ROC 曲线下的面积 (the Area Under the ROC Curve, AUC) 也被用来表征识别的效果的优劣，其值越接近 1 表明识别结果越好。Kappa 也被用来对识别效果进行评价^[31]，它用来表征识别效果和随机识别的差别，越接近 1 表明识别结果越好，其定义为：

$$\text{Kappa} = \frac{P_0 - P_c}{1 - P_c}, \tag{10}$$

其中 P_0 为观测一致性， P_c 为期望一致性。

表 3 KPCA 中不同核函数的参数设置

名称	核函数
RBF	$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\ \mathbf{x} - \mathbf{y}\ ^2}{16}\right)$
Sigmoid	$k(\mathbf{x}, \mathbf{y}) = \tanh(\langle \mathbf{x} \cdot \mathbf{y} \rangle + 1)$
MK	$k(\mathbf{x}, \mathbf{y}) = \hat{\theta} \langle \mathbf{x} \cdot \mathbf{y} \rangle^2 - (1 - \hat{\theta}) \ \mathbf{x} - \mathbf{y}\ ^{1/2} + 1$

3.4 实验结果及分析

采用 10 次交叉验证的方法进行实验，实验样本集为表 2 中 85 例声带病变嗓音和 53 例正常嗓音。所有样本被均匀地分成 10 个子样本集，取其中 9 个子集进行训练，留下一个子集进行测试，如此重复 10 次，每个子样本测试 1 次。表 4 给出了三种识别方法和不同特征参数的识别结果，包括 Accuracy, AUC, Kappa 和三种识别方法的平均识别率。

表 4 三种识别算法下各参数的识别结果

	SVM			NativeBayes			BFTree			平均识别率 (%)
	Accuracy (%)	AUC	Kappa	Accuracy (%)	AUC	Kappa	Accuracy (%)	AUC	Kappa	
MFCC	90.58	0.89	0.77	89.13	0.94	0.75	85.51	0.84	0.67	88.41
GMCC	92.03	0.92	0.82	90.58	0.96	0.80	86.96	0.83	0.70	89.86
LLE	81.88	0.87	0.64	88.41	0.97	0.76	89.13	0.92	0.76	86.47
MBLLE	94.93	0.94	0.90	94.20	0.97	0.87	95.65	0.92	0.91	94.93

从表 4 可以看出:

(1) GMCC 特征在各种分类器下都比 MFCC 有更好的表现, 平均识别率提高了 1.45%。这说明采用 Gammatone 滤波器组提取的特征参数比采用三角滤波器组提取的特征有更好的区分性。这主要是因为 Gammatone 滤波器组能更好的模拟人耳的听觉特性。

(2) MBLLE 特征的识别率比 LLE 平均有 8.45% 的提高。特征参数可以更细腻地反映发声的非线性特性。MBLLE 的平均识别率比 MFCC 和 GMCC 分别提高了 6.52% 和 5.07%, 证明了非线性分析技术在声带病变患者嗓音识别中的有效性。

(3) MBLLE 比其它三个参数的平均识别率都有提高, 但比 LLE 的平均识别率提高程度最大。其一是因为引入了分频带思想 (MFCC 和 GMCC 本来就有分频带处理), 从每个频带上表述嗓音信号的非线性特性; 其二是因为频带划分可以精准地模拟人耳的特性。

图 3 给出了各参数在三种识别算法的 ROC 曲线。也可以看出本文提出的 MBLLE 参数比传统参数在识别中的优越性。

图 4 给出了特征参数 MFCC, GMCC 和 MBLLE

在不同的核函数下 1~12 维的识别率 (SVM 分类器)。可以看出采用 MK 的 KPCA 算法在各维数的识别率最高, 说明 MK 的学习能力比较强。虽然 Sigmoid 核函数在特征 MBLLE 中取得了较好的效果, 特别是当维数为 1~3 时比 MK 的识别率要高, 但是 Sigmoid 在其他特征抽取时则表现得不尽人意, 说明核 Sigmoid 外推能力比较差, 当应用到分类器中时会影响分类器的泛化能力。而 MK 在这三种参数中表现得比较稳定, 有较好的外推能力。

表 5 给出了不同核函数下各特征参数的最佳识别率。可以看出 MBLLE 参数在进行抽取后识别率还是最高的, 且当 MBLLE 和 MK 相结合时识别率被提高达到 97.82%。贡献来源于两方面: 一是多频带非线性分析技术可以更好地表征声带病变嗓音的特性; 二是本文提出的 MK 在特征抽取方面有更好

表 5 各参数在不同核函数下的最佳识别率

Kernel function	Accuracy (%)		
	MFCC	GMCC	MBLLE
PCA	92.75	95.65	96.37
Sigmoid	83.33	84.06	97.10
Rbf	92.75	95.65	97.10
MK	93.48	97.10	97.82

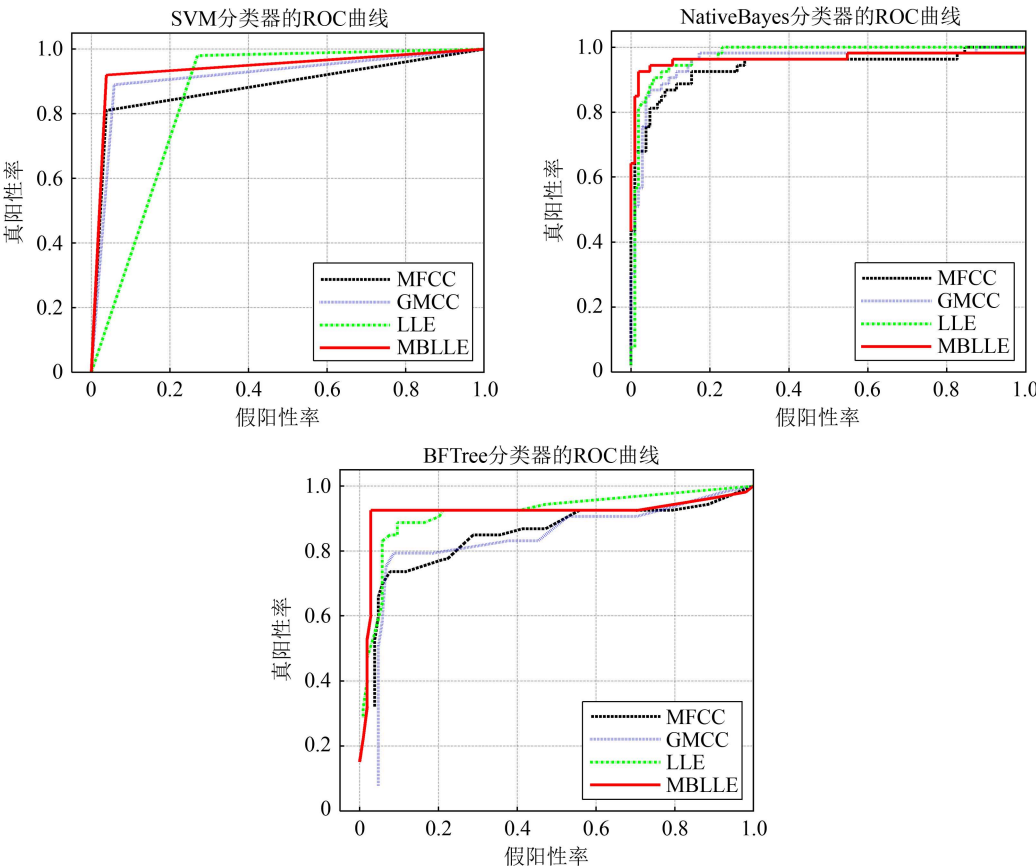


图 3 不同分类器下各参数的 ROC 曲线

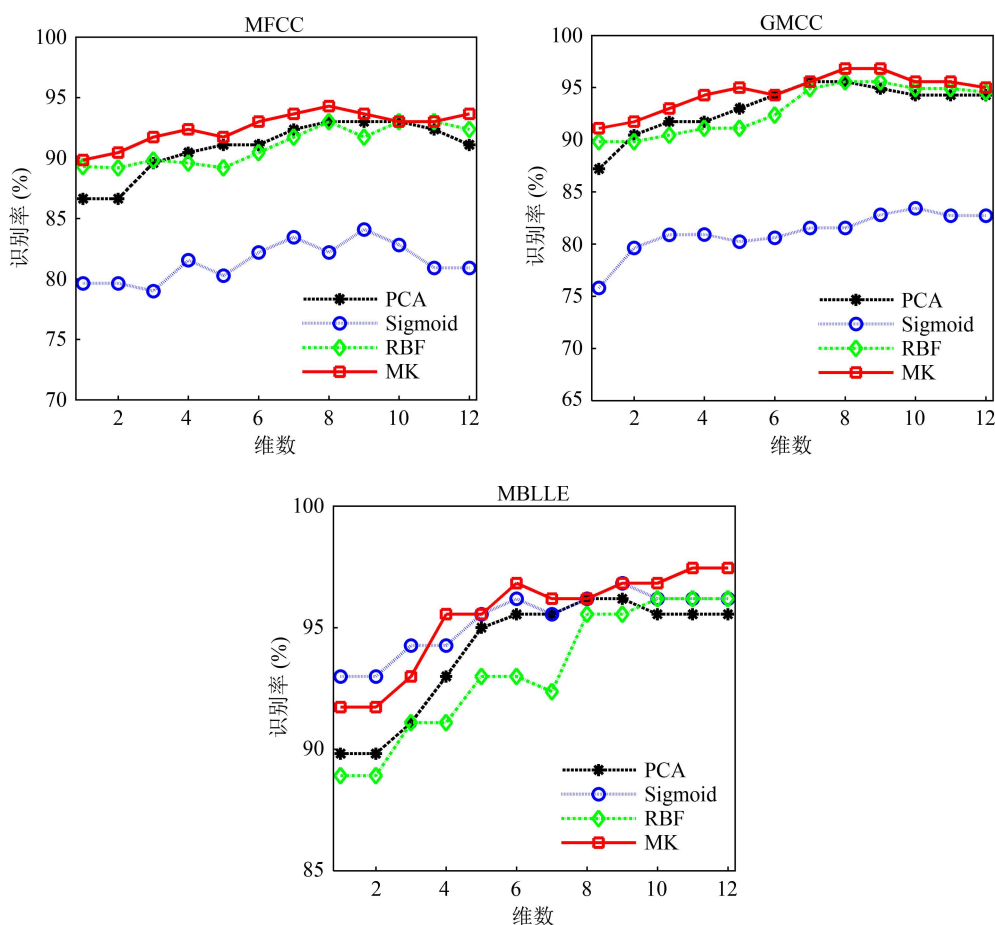


图 4 采用不同核函数在 SVM 分类器下的识别结果

的性能,更精准地提取出有显著性差异频带的非线性信息。

4 结论

针对声带病变嗓音识别中特征参数存在的局限性,根据人耳听觉感知特性与发声非线性特性,将 Gammatone 听觉滤波器组和非线性时间序列分析结合提出了 MBLLE 特征参数。声带病变患者嗓音检测结果表明,采用本文提出的 MBLLE 的识别率相对于 MFCC 和 LLE 有了显著的提高。本文提出的 MK 用于 KPCA 进行特征抽取,识别实验表明 MK 比传统核函数有更高的识别率提升。将 MBLLE 和 MK 结合最佳识别率可提升至 97.82%。

本文提出的 MBLLE 参数对其他与生理病变相关的生物信号,如表面肌电信号的识别具有一定的参考价值,核参数选取算法可应用于其它识别系统的特征抽取中。本文只是利用了声带病变引起的非线性特性,仅局限于声带病变类嗓音信号的检测,下一步将考虑精确识别出疾病的具体种类。核参数的选取只是针对单参数的,如何进行多核参数的选取也是下一步的研究重点。

参 考 文 献

- 1 Gavidia-Ceballos L, Hansen J, Kaiser J. Vocal fold pathology assessment using an autocorrelation analysis of the teager energy operator. *ICSLP*, 1996: 757—760
- 2 Prosek R, Montgomery A, Walden B, Hawkins D. An evaluation of residue features as correlates of voice disorders. *J. Commun. Disord*, 1987; **20**(2): 105—117
- 3 Parsa V, Jamieson D. Identification of pathological voices using glottal noise measures. *J. Speech, Lang., Hear. Res.*, 2000; **43**(2): 469—485
- 4 Jiang J J, Zhang Y, McGilligan C. Chaos in voice, from modeling to measurement. *J. Voice*, 2006; **20**(1): 2—17
- 5 Teager H M, Teager S M. Evidence for nonlinear production mechanisms in vocal tract. In: *Speech Production and Speech Modeling*. Boston: Kluwer Academic Publishers, 1990: 241—261
- 6 Titze I R. The myoelastic aerodynamic theory of phonation. Iowa, IA: *National Center for Voice and Speech*, 2006
- 7 Titze I R, Baken R J, Herzel H. Evidence of chaos in vocal fold vibration, in *Frontiers in Basic Science*, I. R Titze, Ed., San Diego, CA: Singular Publishing Group, 1993: 143—188
- 8 J J Jiang, Y Zhang, Ford C N. Nonlinear dynamics of phonations in excised larynx experiments. *J. Acoust. Soc. Am.*, 2003: 2198—2205

- 9 Zhang Y, Jiang J J, Biazzo L, Jorgensen M. Perturbation and non-linear dynamic analysis of voices from patients with unilateral laryngeal paralysis, *The Journal of Voice*, 2005; **19**(4): 519—528
- 10 Godino-Llorente J I, Gomez-Vilda P. Automatic detection of voice impairments by means of short-term cepstral parameters and neural network-based detectors. *IEEE Trans. Biomed. Eng.*, 2004; **51**(2): 380—384
- 11 Godino-Llorente J, Gómez-Vilda P, Blanco-Velasco M. Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Trans. Biomed. Eng.*, 2006; **53**(10): 1943—1953
- 12 Shao Yang, JIN Zhaozhuang, WANG Deliang *et al.* An auditory-based feature for robust speech recognition. ICASSP, Taipei, Taiwan, 2009: 4625—4628
- 13 Shao Yang, WANG Deliang. Robust speaker identification using auditory features and computational auditory scene analysis. ICASSP, Las Vegas, Nevada, USA. 2008: 1589—1592
- 14 Zhang Weiqiang, He Liang, Hou Tao *et al.* Fractional Fourier transform based auditory feature for language identification. IEEE Asia Pacific Conference on Circuits and Systems. Macao, China, 2008: 209—212
- 15 Manfredi C. Adaptive noise energy estimation in pathological speech signals. *IEEE Transactions on Biomedical Engineering*, 2000; **47**(11): 1538—1543
- 16 Decoster W, Debruyne F. Changes in spectral measures and voice-onset time with age: A cross-sectional and a longitudinal study. *Folia Phoniatr Logop*, 1997; **49**(6): 269—280
- 17 Little M A, McSharry P E, Roberts S J, Costello D A, Moroz I M. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomed. Eng. Online*, 2007; **6**(23)
- 18 封子阳. 一种新的声带振动模型振动产生声场的研究, 陕西: 陕西师范大学, 2012
- 19 Shawe-Taylor J, Cristianini N. Kernel methods for pattern analysis. Cambridge U.P, 2004
- 20 Tax D, Duin D. Support vector data description. *Pattern Recognition Letters*, 1999: 1191—1199
- 21 Tax D, Duin D. Support vector data description. *Machine learning*, 2004; **54**(1): 45—66
- 22 Keerthi S S. Efficient tuning of SVM hyper parameters using radius/margin bound and iterative algorithm. *IEEE Transactions on Neural Networks*, 2002; **13**(5): 1225—1229
- 23 Olivier Chapelle, Vladimir Vapnik, Oliver Bousquel *et al.* Choosing multiple parameters for support vector machines. *Machine Learning*, 2002; **46**(1): 131—159
- 24 Aertsen A M H J, Johannesma P I M. Spectro-temporal receptive fields of auditory neurons in the grassfrog. *Biological Cybernetics*, 1980; **38**(4): 223—234
- 25 Giovanni A, Ouaknine M, Triglia J M. Determination of largest Lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis. *J. Voice*, 1999; **13**(3): 341—454
- 26 Wolff A, Swift J, Swinney H, Vastano J. Determining Lyapunov exponents from a time series. *Physica D*, 1985; **16**(3): 285—317
- 27 Scholköpfung B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 1998; **10**(6): 1299—1319
- 28 Smits G F, Jordan E M. Improved SVM regression using mixtures of kernels. IEEE Proceeding of IJCNN 02 on Neural Networks, 2002: 785—790
- 29 Massachusetts Eye and Ear Infirmary. Voice disorders database. Version 1.03. Kay Elemetrics Corp., Lincoln Park, NJ, 1994
- 30 Fawcett T. ROC graphs: Notes and practical considerations for re searches, HP Laboratories, Palo Alto, CA, 2004
- 31 Siegel S, Castellan N J. Non-parametric statistics for the behavioral sciences. 2nd edition. McGraw-Hill, New York, N.Y, 1988