基于非线性动力学的乐器分类方法

芮 瑞,鲍长春
(北京工业大学电子信息与控制工程学院语音与音频信号处理研究室,北京 100124)

摘 要: 本文基于非线性动力学理论,对不同乐器产生的音频时间序列进行了相空间重构,通过分析各类乐器的递归特性,提出了一个新的定量递归参数——密集度,它能够描述管乐器、弦乐器和键盘乐器在相空间中的差异,然后将密集度与传统的音色特征相结合,提出一种乐器分类方法,并将其应用于不同的分类模型.实验表明,本文所提的方法使三类乐器家族的分类准确率提高了4%~7%,单个乐器的分类准确率提高了3%左右.

关键词: 乐器分类; 非线性动力学; 相空间重构; 密集度; 递归图
中图分类号: TN912.3 **文献标识码:** A **文章编号:** 0372-2112 (2012) 07-1481-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2012.07.032

The Musical Instrument Classification Algorithm Based on Nonlinear Dynamics

RUI Rui, BAO Chang-chun
(*Speech and Audio Signal Processing Laboratory, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing* 100124, *China*)

Abstract: In this paper, the phase space reconstruction of audio time sequences producing by different instruments is discussed based on the nonlinear dynamic theory. The dense ratio, a novel quantitative recurrence parameter, is proposed to describe the difference of wind, string and keyboard instruments in the phase space by analyzing the recursive property of every instrument. In addition, a method of musical instrument family classification is developed using several classification models by combining the dense ratio with traditional timbre features. The experiments indicate that the accuracy of the proposed method is improved by 4% – 7% and 3% in the instrument family classification and individual instrument classification, respectively.

Key words: musical instrument classification; nonlinear dynamic theory; phase space reconstruction; dense ratio; recurrence plot

1 引言

音乐是人们文化娱乐活动中不可缺少的一部分,然而,我们只能够通过实时试听的方式才能了解音乐信息,从中获取自己感兴趣的内容.伴随着科技的发展,人们能够接触到的信息正以指数级增长着,因此,在海量的数据中通过实时试听来搜索某一类音乐不是一件容易的事,这就迫切地需要寻找一种自动的音乐分类方法来有效地管理数据.基于内容的音乐分类技术就是为了解决上述问题而提出的[1,2].

音乐大多需要依赖于乐器才能表达出来.很多学者对乐器分类(这里指西洋乐器)做了研究,A Eronen[3]使用美尔倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC)和时频特征,结合 k 近邻(k-Nearest Neighbors, kNN)对29种乐器进行分类,准确率达到35%.J J Deng等人[4]在MFCC和时频特征的基础上加入了MPEG-7音色特征,结合支持向量机(Support Vector Machine, SVM)对20种乐器进行分类,取得86.9%的准确率.B Kostek[5]使用多层神经网络训练小波特征和MPEG-7特征,研究12种乐器的分类,平均准确率可以达到70%.但在已有的研究成果中,不同家族的乐器之间常常发生错误分类的情况,比如:钢琴错分为双簧管;单簧管和双簧管错分为大提琴;吉他错分为小号和钢琴[4,6]等等,这与现实中人耳的听觉特性并不相符.这些现象说明,时频特征、倒谱特征和MPEG-7特征并没有很好的刻画各种乐器家族之间的差异.

随着非线性理论研究的深入,非线性分析方法在音频信号处理中得到了广泛的应用.目前的研究成果已

收稿日期:2011-10-11;修回日期:2012-04-16
基金项目:国家自然科学基金(No.60872027,No.61072089);北京市教育委员会科技计划重点项目(No.KZ201110005005);北京市属高等学校人才强教计划

经表明音频信号的时间序列具有典型的非线性特征[7,8].因此,本文首次将非线性动力学理论引入乐器分类中,对不同乐器产生的时间序列进行相空间重构,揭示了每一类乐器家族的特有属性.通过分析各类乐器的递归特性,提出了一个新的定量递归参数——密集度,它能够描述三类乐器家族在相空间的差异.最后,将密集度与传统的音色特征相结合,提出一种乐器分类方法,并将其应用在不同的分类模型.实验证明,本文提出的方法不仅能有效地减少三类乐器家族之间的错误分类现象,还对同类家族的乐器分类准确率有一定的提升.

2 相空间重构

非线性系统产生的轨迹经过一定时期的演化,最终会做一种有规律的运动.由于文献[7,8]验证了音频信号具有非线性特征,因此本文引入了相空间重构理论,它能够将一维的音频时间序列张开到高维空间,来描述音频信号的运动轨迹.F Takens 的延迟嵌入定理[9]指出,只要选择合适的嵌入维数 m 和延迟时间 τ,就可以将一维音频时间序列 $\boldsymbol{x}=(x_1,x_2,x_3,\cdots,x_K)^{\mathrm{T}}$ 嵌入到 m 维空间 $\boldsymbol{Y}$ 中,即:

$$\boldsymbol{Y}=(\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_N)=\begin{pmatrix}\boldsymbol{x}(1) & \boldsymbol{x}(2) & \cdots & \boldsymbol{x}(N)\\ \boldsymbol{x}(1+\tau) & \boldsymbol{x}(2+\tau) & \cdots & \boldsymbol{x}(N+\tau)\\ \vdots & \vdots & & \vdots\\ \boldsymbol{x}(1+(m-1)\tau) & \boldsymbol{x}(2+(m-1)\tau) & \cdots & \boldsymbol{x}(N+(m-1)\tau)\end{pmatrix} \tag{1}$$

其中,T 表示转置;K 为音频时间序列 $\boldsymbol{x}$ 的长度;$N=K-(m-1)\tau$ 是相空间中相矢量的个数.

2.1 延迟时间 τ 的求取

选择 τ 的基本思想是使相空间中 $\boldsymbol{y}_n$ 的各相邻矢量之间在某种程度上相对独立但又不完全无关.如果 τ 取值过小,则会使得 $\boldsymbol{y}_n$ 的各相邻矢量之间的值过分靠近;而 τ 取值过大,则又会使得 $\boldsymbol{y}_n$ 的各相邻矢量之间完全无关,不能正确反映相空间轨迹的演化规律.τ 的求取方法有很多种,比如自相关函数法和平均互信息法[10].由于自相关函数法实现简单、方便,计算复杂度较低,因此本文选择自相关函数法来求取最佳的延迟时间 τ,如式(2):

$$R(\tau)=\frac{1}{K-\tau}\sum_{n=1}^{K-\tau}x(n)x(n+\tau) \tag{2}$$

对实测的音频时间序列,可以选取 $R(\tau)$的第一个过零点为最佳的延迟时间.

2.2 嵌入维数 m 的求取

相比 τ 而言,嵌入维数 m 的选择更加重要,若选择太小,将导致那些原本不相邻的两个相矢量投影到低维空间上变成相邻的两个相矢量,我们称这样的相矢量为虚假近邻点.若 m 选择太大,又容易破坏相矢量间的真实结构.确定 m 的方法有虚假近邻点(False Nearest Neighbors,FNN)法和奇异值分解(Singular Value Decomposition,SVD)法[10]等.本文选择虚假近邻点法来求取嵌入维数 m,其主要思想是,随着 m 的增加,原来交叉的轨迹不再重叠,虚假近邻点不断减少,直到虚假近邻点的数目不会随着 m 的增加而减少,相空间的几何结构被完全打开,此时就可以确定合适的嵌入维数.

在 m 维相空间中,对任意的相矢量 $\boldsymbol{y}_n$ 都存在一个最近邻相矢量 $\boldsymbol{y}_n^{NN}$,记它们的距离为 $D_m(n)=\|\boldsymbol{y}_n-\boldsymbol{y}_n^{NN}\|$.当相空间维数从 m 增加到 $m+1$ 时,这两个相矢量间的距离变为:

$$D_{m+1}^2(n)=D_m^2(n)+\|\boldsymbol{y}_{n+m\tau}-\boldsymbol{y}_{n+m\tau}^{NN}\|^2 \tag{3}$$

进一步定义 m 每增加 1 引起的此相矢量与其最近邻相矢量间距离的相对变化值 $\Delta D_m(n)$为:

$$\Delta D_m(n)=\left[\frac{D_{m+1}^2(n)-D_m^2(n)}{D_m^2(n)}\right]^{1/2}=\frac{\|\boldsymbol{y}_{n+m\tau}-\boldsymbol{y}_{n+m\tau}^{NN}\|}{D_m(n)} \tag{4}$$

此处 τ 为前面求得的最佳延迟时间,当 $\Delta D_m(n)>f_D$ 时,记 $\boldsymbol{y}_n^{NN}$ 为 $\boldsymbol{y}_n$ 的虚假近邻点.根据经验,阈值 f_D 通常在[10,50]之间选取.对 m 维空间中的每一个相矢量判别其近邻点是否是虚假的,从而统计虚假近邻点占全部相矢量的比例.对实测时间序列,当虚假近邻点的比例小于 5% 时,可以认为相轨迹已经完全展开,此时的 m 即为最佳嵌入维数.

对不同种类的乐器信号而言,它们均源于不同的非线性动力系统,即使是同一个乐器,由于演奏者的不同,重构的相空间结构同样会有所不同.所以,每一帧乐器信号的嵌入维数 m 和延迟时间 τ 都是不同的.然而,在模式识别的研究中,大量的训练数据使得对每帧信号单独求取 m 和 τ 是不现实的,因此本文使用统计的方法,对 3 小时的采样数据,计算每一帧信号的 m 和 τ 取值的概率,并统计最大概率出现的位置,预先设定唯一的嵌入维数和延迟时间.图 1 给出了 6 种常见西洋

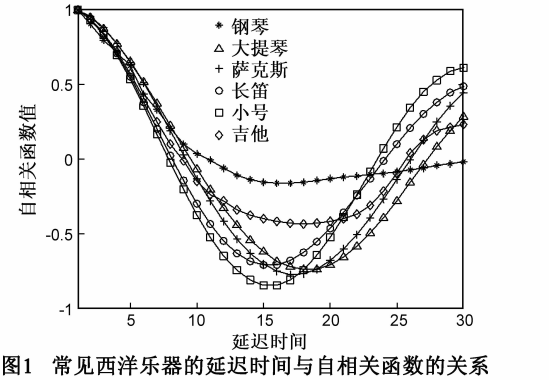


图1 常见西洋乐器的延迟时间与自相关函数的关系

乐器的自相关函数与延迟时间之间的关系.可以看出自相关函数的第一个过零点大多数出现在 $\tau = 7 \sim 10$ 的范围内,因此这里选择出现频率最多的 $\tau = 9$ 为最佳延迟时间.下面讨论嵌入维数的最佳值,图 2 给出了 6 种常见西洋乐器的虚假近邻点比例与嵌入维数之间的关系.可以看出,大多数乐器信号在嵌入维数大于 6 以后虚假近邻点的比例在 5% 以下,此时相空间重构可以反映出原始系统的动力学规律.最后,本文选定参数为:$m = 6, \tau = 9$.

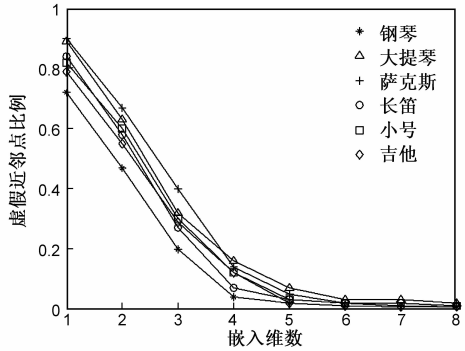


图2　常见西洋乐器的嵌入维数与虚假近邻点比例的关系

2.3　乐器信号的相空间重构

确定了最优的嵌入维数和延迟时间之后,便可以对乐器信号进行相空间重构.本文首先将重构的相空间投影到三维空间中,使得系统的动力学结构能够可视化,进而采取适当的分析手段来处理乐器信号.

三种常见乐器的时域波形和相空间轨迹如图 3 所示,其中图 3(a)(b)(c)表示圆号、小提琴和钢琴的时域波形,图 3(d)(e)(f)表示对应的相空间轨迹.由时域波形图可见,圆号、小提琴和钢琴具有明显的周期或类周期型信号的特征,但是很难从中分辨出这三种乐器.然而从三维相空间中却可以看出,圆号的相空间轨迹具有很强的周期递归性质,表现出明显的周期环状运行轨迹,该信号的周期对应的样点数 T 很容易计算出来(比如自相关法[11]),这 T 个点就构成了一条完整的轨迹.在相空间中,一共有 $n = \lfloor N/T \rfloor$ 条这样的轨迹,并且每一条轨迹分布都十分紧密.小提琴的相空间轨迹依然能够看到周期递归的性质,但是每条轨迹分布比较松散.钢琴的相空间中,每条轨迹分布杂乱无章,但仍然遵循一定的规律,轨迹分布在相对稳定的范围内.

由于这三类乐器的发音机制不同,导致了它们在相空间分布的差异.管乐器是通过演奏者口腔内的气流在管腔内共振而发音;弦乐器是通过琴弓与琴弦的摩擦或者演奏者弹拨琴弦,使张紧的弦线振动而发音;键盘击弦乐器(这里特指钢琴,简称键盘乐器)是通过演奏者按压键盘,使得榔头击弦而发音.

经验证,随着乐器演奏的音高、旋律和风格的变化,它的相空间轨迹分布也有所改变,但有一点是不变的,即:管乐器(包括铜管乐器和木管乐器)、弦乐器和键盘乐器的相空间轨迹分别具有与图 3(d)(e)(f)相似的轨迹分布.图 4 分别给出了双簧管演奏的四个音符:$d\#4, e4, f\#4$ 和 $a4$ 的相空间轨迹.尽管演奏的音高不同,但是这四个音符的相空间轨迹的密集程度十分相似.以上结论为区分不同家族的乐器提供了可能.

3　递归特性分析

3.1　递归图

递归现象是确定性动力学系统中的一个最基本的特征.为了揭示高维相空间轨迹的运行方式,Eckmann 等人[12]提出了一种从二维图形上观察非线性时间序列的动力学特征的分析方法:递归图(Recurrence Plot,RP)法.RP 的数学表达式为:

$$r_{i,j} = \Theta(\varepsilon - \| \boldsymbol{y}_i - \boldsymbol{y}_j \|), i, j = 1, \cdots, N \qquad (5)$$

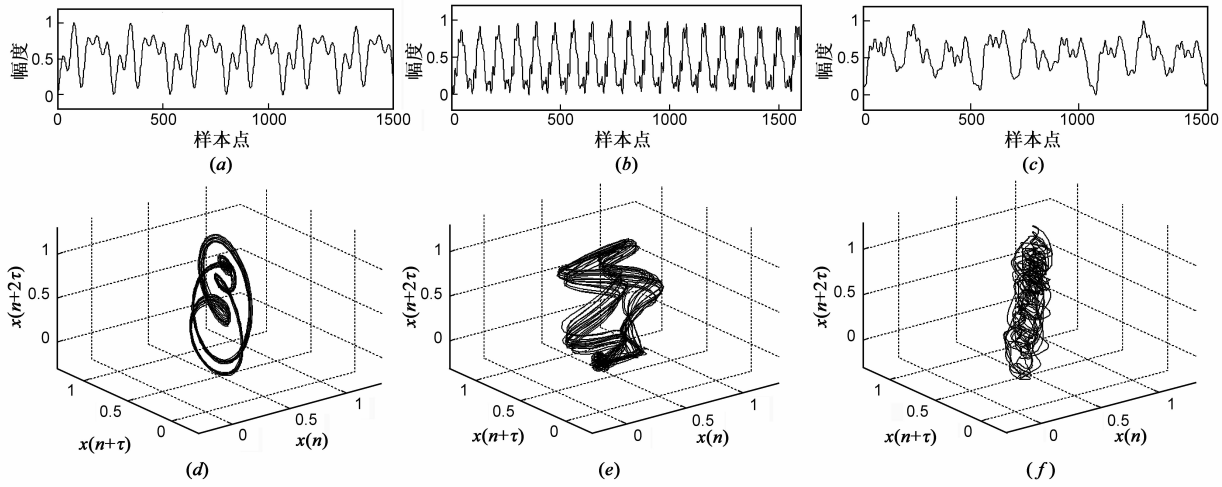
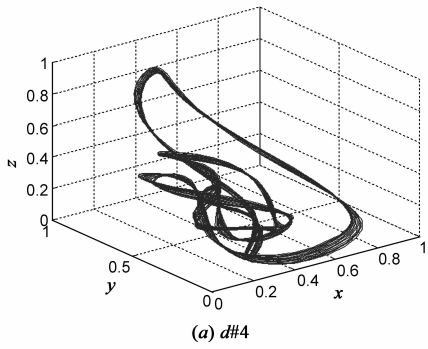
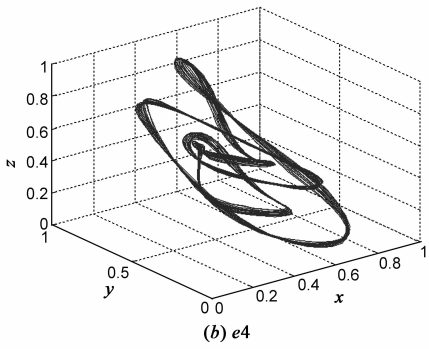


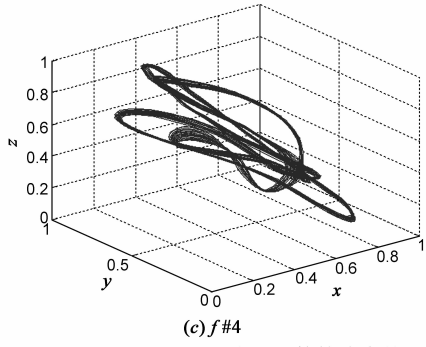
图3　圆号、小提琴和钢琴的时域波形和相空间重构图



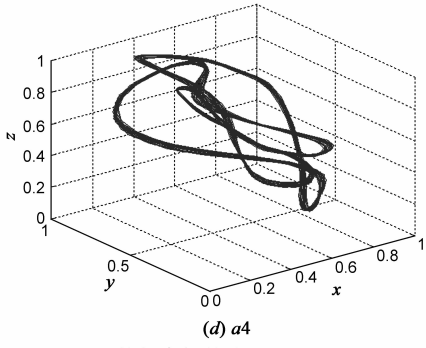
(a) d#4



(b) e4

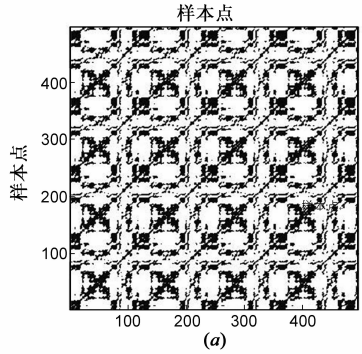


(c) f#4

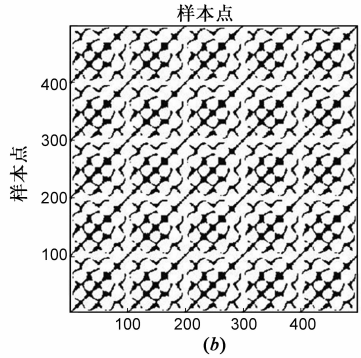


(d) a4

图4 双簧管演奏的d#4, e4, f#4和a4的相空间轨迹



(a)



(b)

图5 吉他(a)和长号(b)的递归图, 参数m=6,τ=9,$\varepsilon=\sigma$

其中,ε 表示预先设定的阈值,$\|\cdot\|$ 表示范数,$\Theta(\cdot)$表示 Heaviside 函数,定义为:

$$\Theta(z)=\begin{cases}1, z\geq 0\\ 0, z<0\end{cases} \tag{6}$$

从 i 时刻到 j 时刻的递归状态可以通过一个二维方阵用黑点或白点来表示.如果相空间中的相矢量 $\boldsymbol{y}_i$ 和 $\boldsymbol{y}_j$ 的距离在设定的阈值ε 之内,那么 $r_{i,j}$的值为 1,递归图中的(i,j)位置上表示成一个黑点;反之,(i,j)位置上则表示成一个白点.因此,递归图将一个 m 维相空间轨迹的分布情况映射到了一个二维图上.图 5 给出了吉他和长号的递归图,从图中能够清晰地看到这两种信号具有明显的周期递归特性.平行于 45°主对角线的每一条线段间隔为信号周期对应的点数.

3.2 定量递归分析

为了量化递归图中表现出来的系统递归现象,Zbilut 和 Webber[13]提出了定量递归分析(Recurrence Quantification Analysis,RQA)的方法,主要包括五种常用的量化参数:递归度(Recurrence Ratio,RR)、确定性(Determinism,DET)、最长对角线(Maximum Diagonal Line,MDL)、熵(Entropy,ENT)和递归趋势(Recurrence Trend,RT).随后,文献[14,15]又补充了层状度(Laminarity,LAM)、1 型和 2 型平均递归时间等参数来获得更丰富的信息.上述参数定量地描述了系统的确定性和稳定性,对于区分确定性信号和随机信号具有很好的效果[15].然而,本文研究的乐器信号都是确定性系统中的周期或类周期型信号(图 5),现有的量化参数无法识别三类乐器家族,如表 1 所示,递归参数依然选择 $m=6,\tau=9,\varepsilon=\sigma$,可以看出最长对角线(MDL)和递归趋势(RT)在三类乐器家族中有特定的取值范围,但是这个范围存在很大的重叠,容易造成错误分类的现象;其他量化参数在三类乐器家族的取值范围则是完全混淆的.

表 1 三类乐器家族的递归定量参数值

	RR	DET	MDL	ENT	RT	LAM	T2
弦乐器	29% ~ 32%	0.99 ~ 1	> 260	2 ~ 3.5	8 ~ 23	0.98 ~ 1	0.5 ~ 0.9
管乐器			> 280		7 ~ 24		
键盘乐器			> 220		11 ~ 25		

3.3 密集度

为了描述各类乐器家族的差异,本文基于相空间重构和递归特性分析,提出一种新的定量递归参数——密集度(Dense Ratio,DR).它表示相空间中,每条轨迹分布的密集程度,其定义式如下:

$$DR=\frac{1}{n(n-1)T}\sum_{i=1}^{T}\sum_{\substack{j,k=0\\ j\neq k}}^{n-1} r_{i+jT,i+kT} \tag{7}$$

其中,T 为信号的周期对应的样点数,$n=\lfloor N/T\rfloor$.如果相空间中的两个相矢量 $\boldsymbol{y}_{i+jT}$和 $\boldsymbol{y}_{i+kT}$的距离很小,其中$\boldsymbol{y}_{i+jT}$表示第 j 条轨迹的第 i 个相矢量,则 $r_{i+jT,i+kT}$的值为 1,递归图中的$(i+jT,i+kT)$位置上表示为一个黑点;反之,则表示为一个白点.$(i+jT,i+kT)$即为图 5 中平行于 45°主对角线的每一条线段.因此,密集度统计了递归图中的$(i+jT,i+kT)$位置为黑点的比例,从而反映

了相空间中任意两条轨迹的第 i 个相矢量之间的距离.

密集度能否准确地区分三类乐器家族,主要取决于 ε 的取值.如果 ε 太大,则所有乐器信号的 $r_{i+jT,i+kT}$ 在递归图上均表示为黑点,如图 5 所示;如果 ε 太小,则所有乐器信号的 $r_{i+jT,i+kT}$ 在递归图上均表示为白点.因此,必须要选择一个适中的 ε,使得三类乐器家族的密集度具有可分性.设 σ 为乐器序列的方差,并且从 $\varepsilon \in \{0.1\sigma,0.2\sigma,0.3\sigma,0.4\sigma,0.5\sigma,0.6\sigma\}$ 的范围内确定阈值,分别计算三类乐器家族的分类准确率,如图 6 所示,可以看出,ε 取值太小或太大,都会导致三类乐器家族的分类准确率明显下降.因此,本文选择 $\varepsilon = 0.3\sigma$ 作为最佳的阈值.

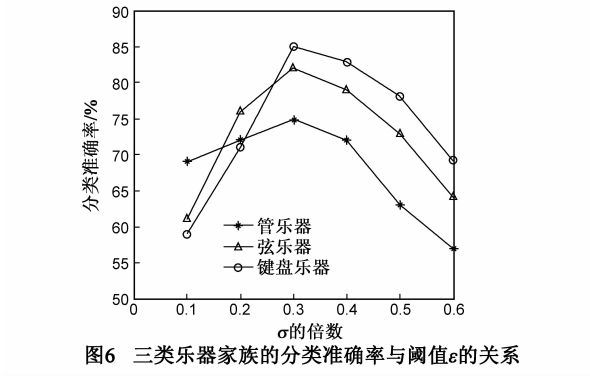


图6　三类乐器家族的分类准确率与阈值ε的关系

图 7 给出了单簧管、大提琴和钢琴的递归图,递归参数选择 $m=6,\tau=9,\varepsilon=0.3\sigma$.图 7($a$)中平行于 45°主对角线的每一条线段间隔即为信号周期对应的点数 T,构成这些线段的黑点即为 $r_{i+jT,i+kT}$.从图 7 能够明显地看出,由于管乐器的相空间轨迹分布紧密,每条轨迹之间的距离都在阈值 ε 之内,所以平行于主对角线的位置上都是黑点.弦乐器的相空间轨迹分布相对比较松散,平行于主对角线位置上的大部分黑点是孤立的,没有连成线段,说明每条轨迹之间的距离都在阈值 ε 附近上下波动.键盘乐器的相空间轨迹杂乱无章,平行于主对角线位置上基本没有黑点,说明每条轨迹之间的距离都在阈值之外.

由递归图的定义式可知,$r_{i,j}$ 是关于主对角线对称的,这里只需考虑递归图中 $i<j$ 这一部分的分布规律,从而减少了一半的计算量.此时,将式(7)变为:

$$DR=\frac{2}{n(n-1)T}\sum_{i=1}^{T}\sum_{\substack{j,k=0\\ j<k}}^{n-1} r_{i+jT,i+kT} \tag{8}$$

三类乐器家族的密集度分布曲线如图 8 所示,由于管乐器的相空间轨迹分布紧密,所以密集度最高,取值介于 0.8 ~ 1 之间;弦乐器的密集度取值介于 0.3 ~ 0.7 之间;键盘乐器的密集度最小,取值在 0.3 以下.

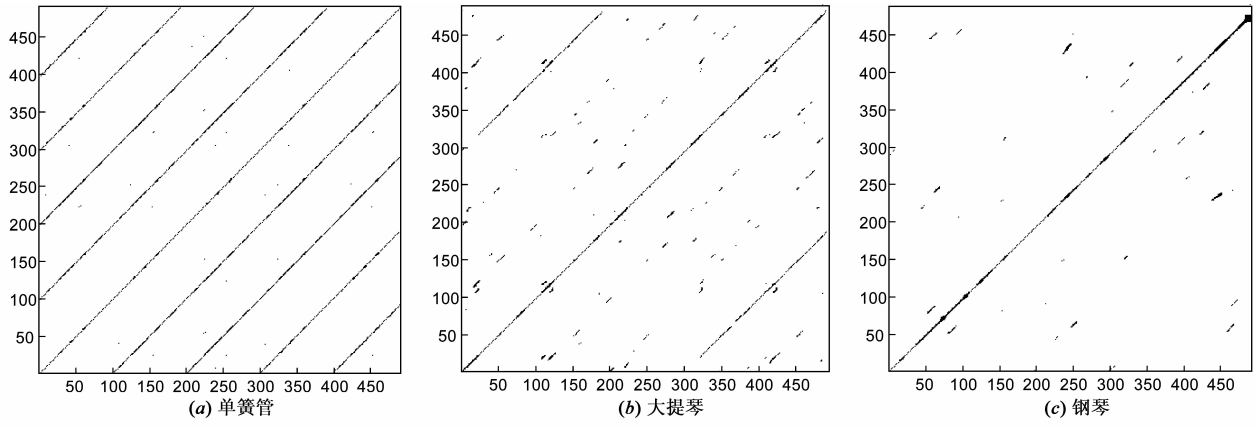


图7　单簧管、大提琴和钢琴分别演奏音符$a4$的递归图

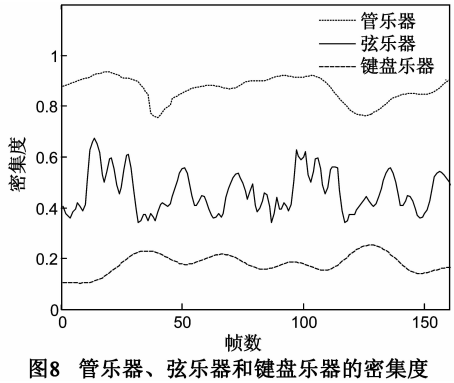


图8　管乐器、弦乐器和键盘乐器的密集度

4　乐器分类方法

为了和传统方法做比较,本文还提取了时频特征,倒谱特征和 MPEG-7 特征,包括:过零率,均方根,子带能量、带宽,谱质心,频谱流量,美尔倒谱系数和一阶差分系数,谐波谱质心,谐波谱离差,谐波谱延伸,谐波谱方差,对数起始时间和时域质心统称为音色特征(每个特征的定义请参考文献[4,16]).然后,取密集度和上述 41 维音色特征的均值和方差,构成一个 84 维的特征集合.最后,使用主成分分析[4](Principal Component Analysis,PCA)法对特征集合进行降维处理,这里对 PCA 原理

不做详细描述.通过实验发现,在使用 PCA 进行降维的过程中,当维数降至 55 时,单个乐器的分类准确率从 90.9%缓慢降至 90.4%;当维数降低至 50 时,分类准确率显著下降至 86.7%;若维数继续减少,则准确率随之大幅下降.因此,本文选择 55 维特征来折中上述矛盾.

分类过程包含训练和测试两部分,图 9 给出了乐器分类的原理框图,实线和虚线分别表示训练和测试的执行步骤.训练阶段先将乐器信号分别经过加矩形窗和哈明窗被分割为 30ms 的音频帧,相邻帧之间重叠一半.随后计算密集度特征和音色特征,共同构成特征矢量,再将训练样本进行降维处理后,输入到分类器进行训练,对每一类样本建立各自的模型并储存起来.测试阶段先将待分类的乐器序列进行特征提取,再将降维后的特征矢量与建立好的每一类模型进行匹配,得到最终的分类结果.

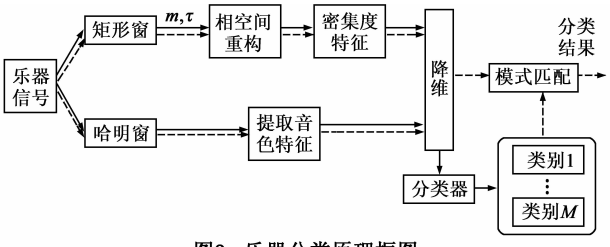


图9 乐器分类原理框图

5 实验结果

实验数据选用采样率 44.1kHz,WAV 格式的数字信号,取自 CD 光盘的西洋乐器独奏序列,共有三类乐器家族:弦乐器、管乐器和键盘乐器.其中弦乐器包括大提琴、小提琴和吉他;管乐器包括单簧管、双簧管、萨克斯、长笛、小号、圆号和长号;键盘乐器指钢琴.数据长度约 3 小时,采用留一法[4](leave-one-out)测试,即假设样本总容量为 N,轮流将 $N-1$ 个样本作为训练数据,留下 1 个样本作为测试数据.第 i 类样本的分类准确率 R_i 如下所示:

$$R_i = \frac{n_i}{N_i} \times 100\% \tag{9}$$

其中,n_i 表示判断为类别 i 并且实际是类别 i 的个数,N_i 表示类别 i 的总个数.

5.1 测试序列长度的选择

分类准确率的高低与测试序列的长度密切相关.选取的测试序列越长,获得的信息越多,准确率就越高.但是,无限的增加测试序列的长度就失去了研究的意义.本实验分别选择从 1s 到 10s 长度的测试数据,得到三类乐器家族的平均分类准确率,如图 10 所示.随着测试序列长度的增加,分类准确率也呈现上升趋势.从 1s 到 4s,准确率大幅增长,高达 7% 左右;选择 4s 到 5s 时,准确率基本保持平稳;在 6s 之后,准确率缓慢增长,并且接近 100%.测试数据的长度和分类准确率是对立的,需要折中考虑.图 10 中,4s 是一个分界点,在它之前准确率大幅增长,在它之后准确率平稳增长.因此后续实验都使用长度为 4s 的测试数据.

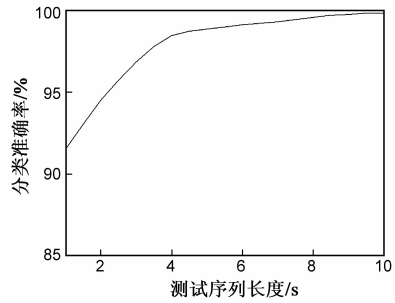


图10 分类准确率与测试序列长度的关系

5.2 乐器家族分类

使用单一分类模型得到的结果不能很好地说明本文方法的性能,因此这里分别选择朴素贝叶斯[17](Naïve Bayes,NB),线性逻辑回归[18](Linear Logistic Regression,LLR),高斯混合模型[6](Gaussian Mixture Model,GMM),支持向量机[19](Support Vector Machine,SVM)和隐马尔科夫模型[20](Hidden Markov Model,HMM)5 种分类器.本文使用 M Hall 等人提供的软件包 Weka[21],它是一个用于机器学习和数据挖掘的开源软件,其中包含了上述 5 种分类模型,所有参数均使用软件包中默认的参数.随后分别使用音色特征和递归特征对三类乐器家族进行训练,再将两者结合,最终给出测试结果.表 2 给出了不同分类器下三种乐器家族的平均分类准确率,可以看出尽管不同的分类器性能有所差异,但是将音色特征和密集度结合在一起之后,所有分类器得到的准确率明显地比单独使用音色特征提升了 4% ~ 7%,使用 HMM 分类时准确率最高,达到 97.8%.使用密集度得到的错分类现象,主要发生在一些音符的起始帧,起始阶段的不稳定影响了相空间轨迹的分布,使得密集度不能很好地区分各类乐器,进而出现错判的情况.

表 2 不同分类器下三类乐器家族的平均分类准确率(%)

特征集合	分类器				
	NB	LLR	GMM	SVM	HMM
音色特征	82.7	91.4	92.6	94.2	94.2
密集度	72.8	76.2	78.3	80.9	81.5
音色特征 + 密集度	89.5	94.6	95.9	97.4	97.8

5.3 单个乐器分类

本实验对 11 种乐器的独奏序列进行分类,将音色特征和密集度结合起来,使用 HMM 得到分类结果见表 3.此表给出了一个混淆矩阵(confusion matrix),它是评估分类器性能的基本工具,行代表实际的乐器类型(Real Instrument Type,RIT),列代表判别的乐器类型

(Discriminative Instrument Type, DIT). 其中,行中的数据代表实际乐器被判为不同乐器的比例,每一行数据之和为 100%. 如表 3 第三数据行所示,当实际的乐器类型是吉他时,判为大提琴的比例为 1.0%,判为小提琴的比例为 3.5%,判为吉他的比例为 93.7%,判为圆号的比例为 0.9%,判为钢琴的比例为 0.9%. 主对角线的数据表示每一种乐器的分类准确率,括号内数据表示单独使用音色特征[4,16]时的判别结果. 可以看出,不同乐器家族之间的错分现象普遍存在,错分比例最高达到 6.5%. 而本文提出的方法取得了很好的效果,括号外数据表示在音色特征中加入密集度的错误分类比例,可以看出,不同乐器家族之间的错分比例明显下降,尤其是误判为钢琴的情况,错分比例可以控制在 0.6%以下;其他乐器家族的错分比例在 1%左右. 单个乐器的平均分类准确率达到 90.4%,相比单独使用音色特征提高了 3%左右. 另外,加入密集度也有利于区分同一家族的乐器. 比如,管乐器中的萨克斯,在对单簧管和双簧管的错分中减少了 1% ~ 2%.

表 3　11 种乐器的分类结果(单位:%)

RIT \ DIT	大提琴	小提琴	吉他	单簧管	双簧管	长笛	萨克斯	小号	圆号	长号	钢琴
大提琴	**97.6**(93.8)	1.7(1.7)	0.6(0.8)	0	0	0	0	0	0	0.1(3.7)	0
小提琴	0	**95.0**(91.4)	4.4(4.4)	0	0	0.4(1.8)	0	0	0	0	0.2(2.4)
吉他	1.0(1.0)	3.5(3.5)	**93.7**(87.1)	0	0	0	0	0	0.9(3.6)	0	0.9(4.8)
单簧管	1.1(5.4)	0	0	**80.5**(74.9)	0	4.7(4.7)	9.4(10.7)	0	0	4.3(4.3)	0
双簧管	0.8(3.8)	0	0	4.9(4.9)	**87.5**(79.4)	2.1(2.1)	1.0(1.0)	0	0	3.7(4.6)	0(4.2)
长笛	0	0.5(6.1)	0	5.2(5.2)	0	**88.7**(83.1)	3.2(3.2)	0	0	2.4(2.4)	0
萨克斯	0	0	0	6.6(8.0)	0.9(2.7)	4.3(4.3)	**85.2**(82.0)	0	3.0(3.0)	0	0
小号	0	0	0	0	0	0	2.2(2.2)	**91.2**(91.2)	2.6(2.6)	3.9(3.9)	0
圆号	0	0	0	3.8(3.8)	0	0.9(0.9)	0	4.3(4.3)	**84.7**(84.7)	6.3(6.3)	0
长号	0.3(6.5)	0	0	0	0	2.3(2.3)	0	4.7(4.7)	2.1(2.1)	**90.1**(80.9)	0(3.0)
钢琴	0	0	0	0	0(3.4)	0	0	0	0	0	**100**(96.6)

6　结束语

本文基于非线性动力学理论,提出了一个新的递归定量参数——密集度,它能够准确地描述三类乐器家族在相空间的区别. 使用本文采集的数据库得到的实验结果表明,将本文提出的特征与传统的音色特征相结合,不仅使三类乐器家族的分类准确率提高了 4% ~ 7%,单个乐器的分类准确率提高了 3%,还对同类家族的乐器分类准确率有一定的提升. 这体现了非线性动力学理论在乐器信号分析中的有效性,并为乐器特征提取和分类的研究提供了新方法. 最后,如何能更好地区分同一家族的乐器,是未来研究的一个重点.

参考文献

[1] Mark R Every. Discriminating Between Pitched Sources in Music Audio [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2008, 16(2): 267 – 277.

[2] 张一彬,周杰,边肇祺,张大鹏. 一种新的基于分类的音频流分割方法[J]. 电子学报, 2006, 34(4): 612 – 617.
ZHANG Yi-bin, ZHOU Jie, BIAN Zhao-qi, ZHANG Da-peng. A novel classification-based audio segmentation algorithm [J]. Acta Electronica Sinica, 2006, 34(4): 612 – 617. (in Chinese)

[3] A Eronen. Comparison of features for musical instrument recognition [A]. Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics [C]. New York: IEEE Press, 2001. 19 – 22.

[4] Jeremiah D Deng, C Simmermacher, S Cranefield. A study on feature analysis for musical instrument classification [J]. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, 2008, 38(2): 429 – 438.

[5] B Kostek. Musical instrument classification and duet analysis employing music information retrieval techniques [J]. Proceedings of the IEEE, 2004, 92(4): 712 – 729.

[6] J Garcia A Barbedo, G Tzanetakis. Musical instrument classification using individual partials [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(1): 111 – 122.

[7] Y T Sha, C C Bao, et al. High frequency reconstruction of audio signal based on chaotic prediction theory [A]. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Dallas: IEEE Press, 2010. 381 – 384.

[8] J Serra, C A Santos, et al. Nonlinear audio recurrence analysis

with application to genre classification [A]. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Prague: IEEE Press, 2011. 169 – 172.

[9] F Takens. Detecting strange attractors in turbulence [J]. Lecture Notes in Mathematics, Springer, Berlin, 1981, 898: 366 – 381.

[10] H Kantz, T Schreiber. Nonlinear Time Series Analysis [M]. Cambridge, UK: Cambridge University Press, 2nd edition, 2004. 64 – 151.

[11] T Kitahara, M Goto, K Komatani, T Ogata. Musical instrument recognizer "instrogram" and its application to music retrieval based on instrumentation similarity [A]. Proceedings of IEEE International Symposium on Multimedia [C]. Washington DC: IEEE Press, 2006. 265 – 274.

[12] J P Eckmann, S O Kamphorst, D Ruelle. Recurrence plots of dynamical systems [J]. Physics Letters, 1987, 4: 973 – 977.

[13] J P Zbilut, C L Webber Jr. Embeddings and delays as derived from quantification of recurrence plots [J]. Physics Letters, 1992, 171: 199 – 203.

[14] N Marwan, N Wessel. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data [J]. Physical Review, 2002, 66: 1 – 8.

[15] J B Gao, H Q Cai. On the structures and quantification of recurrence plots[J]. Physics Letters A, 2000, 270: 75 – 87.

[16] ISO/MPEG N4224. Text of ISO/IEC Final Draft International Standard 15938-4 Information Technology Multimedia Content Description Interface—Part 4 Audio [S]. MPEG Audio Group, Sydney, 2001.

[17] Zhou-yu Fu, Guo-jun Lu, et al. Learning naïve bays classifiers for music classification and retrieval [A]. Proceedings of IEEE International Conference on Pattern Recognition [C]. Istanbul: IEEE Press, 2010. 4589 – 4593.

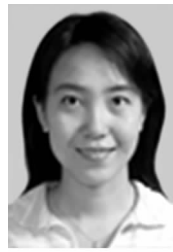
[18] Khe Chai Sim, Kong-Aik Lee. Adaptive score fusion using weighted logistic linear regression for spoken language recognition [A]. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Dallas: IEEE Press, 2010. 5018 – 5021.

[19] 齐峰岩,鲍长春. 一种基于支持向量机的含噪语音的清/浊/静音分类的新方法[J]. 电子学报, 2006, 34(4): 605 – 611.
QI Feng-yan, BAO Chang-chun. A method for voiced / unvoiced / silence classification of speech with noise using SVM [J]. Acta Electronica Sinica, 2006, 34(4): 605 – 611. (in Chinese)

[20] 卢坚,陈毅松,孙正兴. 基于隐马尔可夫模型的音频自动分类[J]. 软件学报, 2002, 13(8): 1593 – 1597.
LU Jian, CHEN Yi-song, SUN Zheng-xing. Automatic audio classification by usig hidden Markov model [J]. Journal of Software, 2002, 13(8): 1593 – 1597. (in Chinese)

[21] M Hall, E Frank, et al. The Weka data mining software: An update [J]. ACM SIGKDD Explorations Newsletter, 2009, 11(1): 10 – 18.

作者简介



芮 瑞 女, 1983 年出生, 北京人. 北京工业大学博士研究生, 主要研究方向为基于内容的音频分类检索.
E-mail: rr_2006@emails.bjut.edu.cn

鲍长春 男, 1965 年出生, 内蒙古赤峰人. 博士, 北京工业大学教授、博士生导师, 国际语音通信学会(ISCA)会员, 中国电子学会理事, 信号处理学会委员, 《通信学报》编委会副主任委员、《信号处理》和《数据采集与处理》编委. 主要研究方向为语音与音频编码, 语音与音频增强, 语音与音频频带扩展, 基于内容的音频分类检索.
E-mail: baochch@bjut.edu.cn