

基于 Mel 倒谱系数和矢量量化的昆虫声音自动鉴别

竺乐庆¹, 王鸿斌², 张 真^{2,*}

(1. 浙江工商大学计算机与信息工程学院, 杭州 310018;

2. 中国林业科学研究院森林生态与保护研究所, 国家林业局森林保护学重点实验室, 北京 100091)

摘要: 为了给生产单位害虫管理的普通技术人员提供简便易操作的昆虫种类鉴别方法, 本研究把人类语音识别领域的先进技术应用于昆虫识别, 提出了一种新颖的昆虫声音自动鉴别方法, 用声音参数化技术为昆虫声纹识别设计了一种简单易行的方案。声音信号经过预处理、分段得到一系列的声音样本, 从声音样本提取 Mel 倒谱系数(MFCC), 并用 Linde-Buzo-Gray(LBG)算法对提取的 MFCC 进行矢量量化(VQ), 所得码字作为声音样本的特征模型。特征参数之间的匹配用搜索最近邻的方法实现。本文方法在包含 70 种昆虫声音的库中进行了试验, 取得了超过 96% 的识别率和理想的时间性能。试验结果证明了该方法的有效性。

关键词: 昆虫; 声音识别; Mel 倒谱系数; LBG 算法; 矢量量化

中图分类号: Q967 文献标识码: A 文章编号: 0454-6296(2010)08-0901-07

Automatic acoustical identification of insects based on MFCC and VQ

ZHU Le-Qing¹, WANG Hong-Bin², ZHANG Zhen^{2,*} (1. College of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China; 2. Key Laboratory of Forest Protection of State Forestry Administration, Research Institute of Forest Ecology, Environment and Protection, Chinese Academy of Forestry, Beijing 100091, China)

Abstract: This study aims to provide general technicians who manage pests in production with a convenient way to recognize insects. A simple and viable scheme to identify insect voiceprints automatically is introduced using a sound parameterization technique that dominates speaker recognition technology. The acoustic signal was preprocessed and segmented into a series of sound samples. Mel-frequency cepstrum coefficient (MFCC) was extracted from the sound sample, and a feature model was trained using Linde-Buzo-Gray algorithm to generate vector quantization (VQ) codebook from above MFCC. The matching for a test sample was completed by finding the nearest neighbour in all the VQ codebooks. The method was tested in a database with acoustic samples of 70 different insect sounds. The recognition rate above 96% was obtained, and an ideal time performance was also achieved. The test results proved the efficiency of the proposed method.

Key words: Insects; sound recognition; Mel-frequency cepstrum coefficient (MFCC); Linde-Buzo-Gray (LBG) algorithm; vector quantization (VQ)

尽管研究人员在昆虫声学通讯方面已做了较多的研究工作(Drosopoulos and Claridge, 2005), 但对于发声昆虫种类的自动鉴别仍然是模式识别的边缘领域, 在这方面的文献也相对较少。昆虫的声学鉴定是基于昆虫具有发出声音的能力, 这种能力或者是出于沟通的目的有意发出, 或者是因取食、飞翔、运动而产生。昆虫发声机制归结为肌肉的有力收缩

导致发声部位的机械振动, 该声源被装载并向外发射。昆虫通过以下 5 种不同方式产生声音(Alexander, 1957):

1) 摩擦(stridulation): 身体两个不同部分的摩擦, 通常发出唧唧声(蟋蟀、纺织娘、螽斯、甲虫、蛾、蝴蝶、蚂蚁、毛虫、甲壳虫幼虫等);

2) 敲击(percussion): 通过身体某个部位如足

基金项目: 国家高技术研究发展计划(“863”计划)项目(2006AA10Z211); 中国林业科学研究院基本科研业务专项资金(CAFRIF200710)

作者简介: 竺乐庆, 女, 1972 年 7 月生, 浙江嵊州人, 博士, 讲师, 研究方向为图像处理及模式识别, E-mail: zhuleqing@zjgsu.edu.cn

* 通讯作者 Corresponding author, Tel.: 86-010-62889567; E-mail: zhangzhen@caf.ac.cn

收稿日期 Received: 2010-01-27; 接受日期 Accepted: 2010-07-04

(带翅螽斯)、腹尖(蟑螂)、或头(红毛窃蠹)敲击地面,通常发出嗒嗒声或咚咚声;

3) 振动(vibration and tremulation): 通过身体某些部位的振动发声,如翅膀在空气中的振动,通常发出嗡嗡声或隆隆声(蚊子、苍蝇、黄蜂、蜜蜂等);

4) 鼓室结构(tymbal mechanism): 快速收缩和舒张鼓室肌(振动类似鼓面的薄膜);通常发出一系列的嘀嗒声(蝉、叶蝉、角蝉、沫蝉);

5) 喷气(air expulsion): 通过身体收缩喷出气体或液体,通常发出哨叫声或嘶嘶声(螳螂、短角蝗)。

昆虫发出声音与其特定的行为模式相联系。根据不同行为主要将其分为以下4大类(Alexander, 1957):

1) 合唱(chorus): 昆虫群体同时发声,主要目的是召唤雄虫和雌虫聚集到一起(蝉);

2) 召唤声(courtship calling): 这是昆虫结对的第一步,用于将较大范围内雌虫吸引到邻近区域(如蟋蟀和蝉发出的配对声),一些雌虫也会发出声音以帮助雄虫确定其所在位置(斜面蝗)或对雄虫作出应答(纺织娘);

3) 求偶声(mating calling): 雄虫在小范围内发出的声音,目的是在交配前吸引有反应的雌虫。当雌虫接近到1 m 范围内时,雄虫即由召唤声转变为求偶声;

4) 争斗声(competing and alarm sound): 这种声音当昆虫在飞行中被捕获或扰乱、或雄虫通知其他雄虫侵入了它的活动范围时发出的宣告干扰的鸣叫(通常称为警告、胁迫或争斗声)。这种声音也用于向同类对危险发出报警,这种抗议声有些缺乏节奏感,因为鼓室的振动频率(蝉)或摩擦器官(蟋蟀、纺织娘)因种类而异,这种声音具有种类独特性。

如果昆虫产生的生物声学信号遵循一致的模式并具有种类上的独特性,那么这些信号就可以用于鉴定和检测的目的。Riede (1998)表明昆虫发出的声音能提供可靠的分类学线索,因而可用于生物多样性度量。韩萍(2003)用频谱分析和BP神经网络对仓储害虫的声音识别进行了研究,在包含米象 *Sitophilus oryzae* Linne.、玉米象 *S. zeamais* Motsch. 和赤拟谷盗 *Tribolium castaneum* Herbst 3种仓虫的数据库中取得了81%的识别率。Chesmore 和 Nellenbach (2001)用声音信号的时域信号分析法和人工神经网络对直翅目昆虫(蝗虫及蟋蟀)的自动鉴别进行了研究,并在包含25类英国直翅目昆虫的库中进行了测试,取得较高的识别率;Chesmore (2004)又用同样方法精确识别出在较高干扰的自然环境下的4类英国直翅目昆虫鸣叫声;Pinhas 等

(2008)则对红棕象甲 *Rhychophorus ferrugineus* (Oliver)自动声音识别进行了研究,最后分别用矢量化(vector quantization, VQ)技术及高斯混合模型(gaussian mixture modeling, GMM)技术实现识别;Ganchev 等(2007)用主谐波、振动持续时间、以及23个线性倒谱系数(linear frequency cepstral coefficients, LFCCs)经归一化构成声音片段的特征向量,分别用似然神经网络(probabilistic neural network, PNN),混合高斯模型(GMM),隐马尔可夫模型(hidden markov model, HMM)分类器进行识别,在包含了313种昆虫(蝉、蟋蟀和纺织娘)的北美鸣虫库(singing insects of the North America collection, SINA)中进行了试验验证,取得了理想的识别精度。

昆虫声音识别的相关应用领域包括:特定区域的自动环境检测和生物多样性分类与编目;濒危物种的生存能力分析;生存环境健康程度和恶化评估,因为某些物种是居住环境质量和环保的标志;农业害虫的监测和预警;用非专家方式识别和分类相当范围的昆虫种类。

本研究用人类语音处理中行之有效的方法,来解决昆虫声学检测中遇到的挑战性课题。我们认为在自动语音/说话人识别和昆虫声学识别之间的交叉应用是有根据的,因为两者有相同的目的,只是在发出声音的机制上有所不同。此处把语音处理任务中涉及的信号处理方法进行了调整以适应于昆虫识别的具体应用。参考语音识别的特征提取方法,本文的参数化技术为昆虫声学识别应用作了专门设计。本文方法在仓储昆虫取食和行动的声音、土壤无脊椎动物的取食和移动声音、树木上昆虫取食与移动声音、作物上昆虫取食与移动声音以及昆虫翅膀和腹部振动的声音混合的数据库中进行了测试,取得了理想的识别效果。

1 材料与方法

1.1 声音样本来源和预处理

本文试验所用材料采用美国农业部 Richard Mankin 学科组录制的各类昆虫声音库(Mankin, 2009)以及本课题组录制的一些小蠹声音文件。这些录制的昆虫声音文件长度介于3~60 s之间,过长的声音文件直接进行特征提取一方面会增加计算过程的处理负荷,另一方面背景噪声作为声音的一部分进行分析会影响识别精度,取这些样本中包含

昆虫发声时的活跃状态 1.2 s 的片段, 便足以提取出声音的特征参数。因此, 我们把声音文件进行预处理, 并把每一个声音文件进行分割, 删除无用成分后划分成若干个样本, 由这些划分得到样本集合组成新的数据库, 本文试验在预处理和分割后的样本库上进行。

假定输入的昆虫声音信号为经过采样和量化的数字信号, 预处理主要包括振幅归一化、样本分割等步骤。

1.2 声音信号归一化

按信号幅度绝对值的最大值做归一化处理, 得到幅值范围统一的信号, 即:

$$\tilde{x}(i) = x(i) / \max_{0 \leq i < n-1} x(i) \quad (1)$$

其中 $x(i)$ 为原信号, $\tilde{x}(i)$ 为归一化后的信号, n 为信号长度。

1.3 声音信号分割

分割信号为一定长度范围内的帧, 提取最能体现其特点的部分, 以降低信号长度, 减少运算量。分割基于声学行为的探测器, 估计 K 个连续采样点的经 $1 - az^{-1}$ ($a=0.9375$) 预加重的短时能量, 即:

$$E(k) = \sum_{i=1}^K (x(kL + i) - ax(kL + i - 1))^2 \quad (2)$$

$k = 0, \dots, M - 1$

其中 x 是输入信号, k 是帧号, L 是一个预定义的步距, 该值也决定了两个连续帧的交叠程度,

$$F = \lfloor (N - K + L) / L \rfloor \quad (3)$$

是包含 N 个采样点的音频信号的总帧数, $\lfloor \cdot \rfloor$ 表示向下取整。 K 为每一帧包含的采样点数。边界的精度取决于步距 L 的大小, 本研究取 $L=80$ (在 25 000 Hz 的采样频率下, 相当于 3.2 ms 的时间分辨率), 这是为了取得计算代价和时间分辨率之间比较好的折衷。每一帧的窗宽取 $K=240$, 即对应帧持续时间为 9.6 ms。我们用短时能量序列中的最大值 E_{max} 的 10% 作为阈值 th , 小于 th 的信号看作是静音, 相邻两段静音之间的数据为一个样本, 如果该段样本持续时间过短, 便当作噪音滤除, 样本最大长度取 1.2 s, 如果超过 1.2 s, 则只截取以窗内最大值为中心的 1.2 s 信号作样本。

如图 1 所示, 长度为 6 s 的原音频数据经过归一化, 将被分割为 3 个样本来处理。

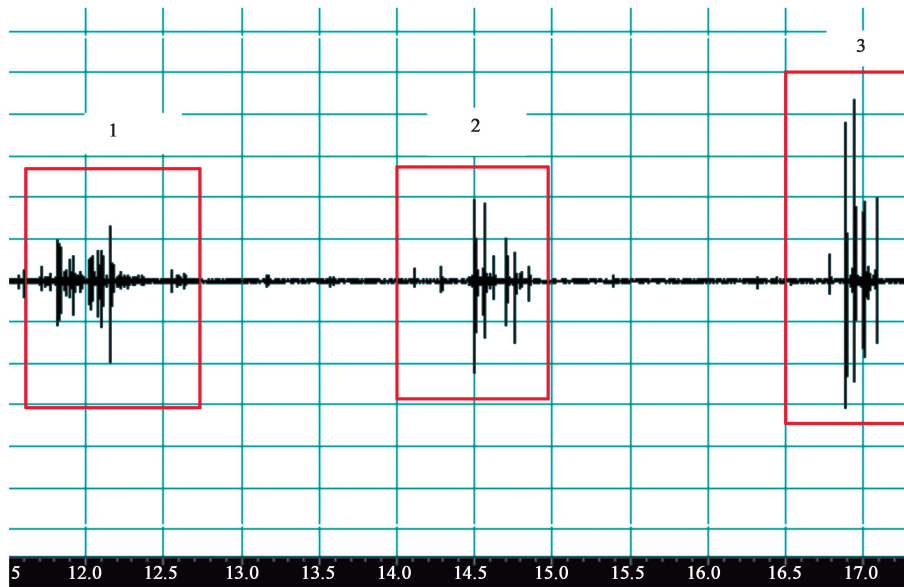


图 1 声音采样数据分割示意图

Fig. 1 The segmentation of a sound sample

2 自动昆虫识别

图 2 给出了本系统所采用的昆虫声纹识别过程的流程图。

2.1 特征提取

声音特征参数可以是能量、基音频率、共振峰值等语音参数, 目前在声音识别中较为常用的特征参数为线性预测倒谱系统 LPCC 和 Mel 倒谱系数 (MFCC)。二者都是将声音信号变换到倒谱域上,

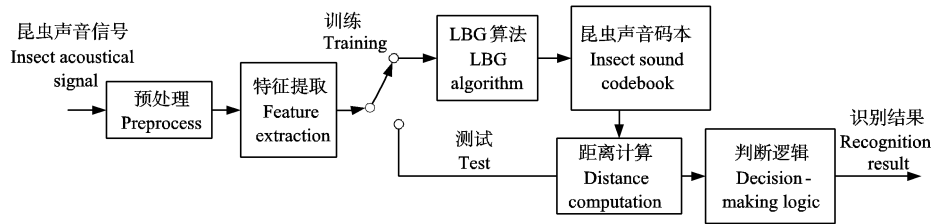


图2 基于 VQ 的昆虫声纹识别流程图

Fig.2 Flowchart of the insect acoustic recognition based on VQ

前者是从发声模型角度出发,利用线性预测编码(LPC)技术求倒谱系数;后者则构造人的听觉模型,人的听觉系统是一个特殊的非线性系统,它响应不同频率信号的灵敏度是不同的,基本上是一个对数的关系。MFCC 充分利用人耳这种特殊的感知特性,以声音通过该模型(滤波器组)的输出为声学特征,直接通过离散傅里叶变换(DFT)进行变换。MFCC 与 LPCC 相比具有其优势:MFCC 采用 Mel 频标,突出有利于识别的低频信息,而屏蔽了包含于高频部分的噪声干扰,基于线性频标的 LPCC 无此特点;MFCC 无任何前提假设,而 LPCC 假定所处理的信号为自回归信号,对于动态较强的信号,或噪声存在时,这一假设并不成立。MFCC 和线性频率的转换关系(如图3)是:

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (4)$$

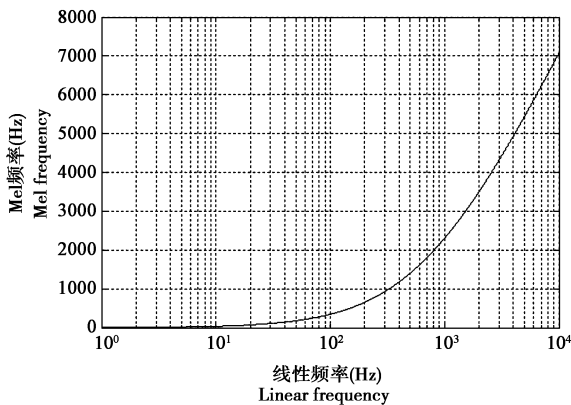


图3 Mel 频标与线性频率的关系

Fig.3 The relation between Mel pitch and frequency

鉴于上述因素,本文采用 MFCC 提取特征参数, MFCC 的计算过程如图4所示。

1) 预加重: 功率谱随频率的增加而减小,其大部分能量集中在低频范围内。这就造成消息信号高频端的信噪比可能降到不能允许的程度。预加重的目的是提升高频部分,使信号的频谱变得平坦,以便于进行频谱分析或声道参数分析。预加重在采样

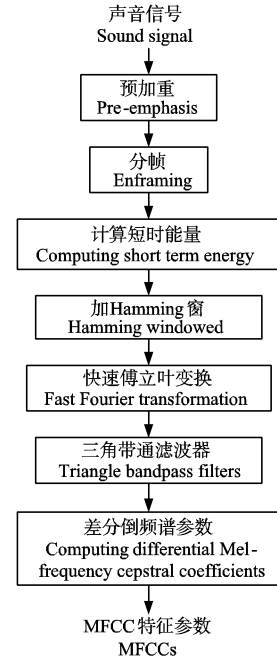


图4 MFCC 流程图

Fig.4 Flowchart of the MFCC extraction

/量化之后进行,用具有 6 dB/倍频程的提升高频特性的预加重数字滤波器实现,其 z 传递函数为:

$$H(z) = 1 - az^{-1} \quad (5)$$

上式中, a 值接近于 1, 本文取 $a = 0.9375$ 。

2) 加窗: 声纹信号的准平稳特性,使得只在短时段上才可视为一个平稳过程。可以用平稳过程的分析方法进行分析,因此需将声音信号划分为一个一个的短时段,每一短时段称为一帧,为从声音信号中切取含有 N 个样本的声音信号波形,需要用时间窗函数乘以原来的语音信号。而矩形窗的采用使得每一帧信号在起始点和终点处发生突变,不再连续,从而导致 Gibbs 现象的出现。为了减小起始点和终点处的不连续性,可以改变窗函数的时域特性,使其在起始点和终点处逐渐变为零,如 Hamming 窗(即升余弦窗):

$$\omega_H = 0.54 - 0.46 \cos\left(\frac{2\pi n}{n-1}\right) \quad n = 0, 1, \dots, N-1 \quad (6)$$

其中 N 称作帧长, 为时间窗的窗宽。为尽可能不丢失语音信号动态变化的信息, 常采用滑动窗, 即帧与帧之间有一定的重叠。

3) 取帧长 $N = 256$ 点, 对每一帧作 FFT 变换, 则第 m 帧声纹的频谱为:

$$S(k, m) = \sum_{n=0}^{255} s(n, m) \exp(-j \frac{2\pi nk}{256}) \quad (7)$$

上式中, $\{s(n, m) | n=0, 1, \dots, 255\}$ 为第 m 帧声纹的 256 个采样点。对频谱取模平方就会得到离散功率谱。

4) 用 M 个 Mel 带通滤波器 (如图 5) 进行滤波, 得到一组系数 $m_1, m_2, \dots$ 。滤波器在频域上为简单的三角形, 在 Mel 频率轴上是均匀分布的。在本文中, 滤波器的个数 M 为 20。由于每一个频带中的分量作用在人耳中是叠加的, 所以将每个滤波器频带内的能量进行叠加。

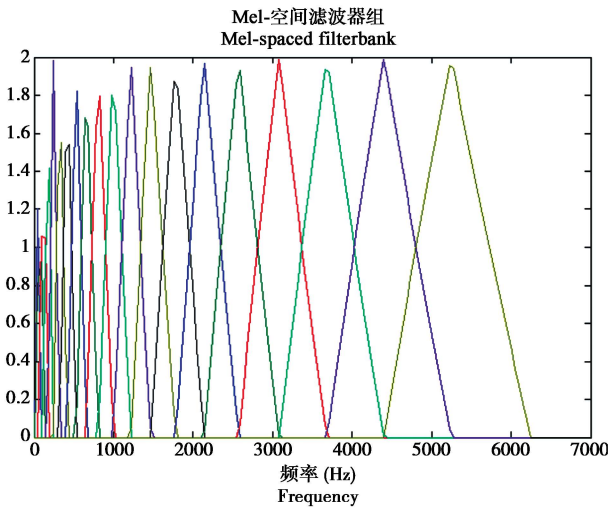


图 5 Mel 滤波器组

Fig. 5 Bank of Mel filters

5) 将每个滤波器的输出取对数, 得到相应频带的对数功率谱, 为避免复数运算, 利用反离散余弦变换得到 L 个 MFCC, 其中 M 是三角滤波器的个数。在实际应用中, 并不需要取全部维数的 MFCC, 试验表明最前若干维的 MFCC 对识别的区分性较大, 通常取 12 维的 MFCC 即可 (甄斌等, 2001)。MFCC 为:

$$C_n = \sum_{k=1}^M \ln x'(k) \cos[\pi(k - 0.5)n/M] \quad (8)$$

$n = 1, 2, \dots, L$

其中, $x'(k)$ 为第 k 个滤波器的输入功率谱。

标准的 MFCC 只反映了语音参数的静态特性, 一阶差分 MFCC (Δ MFCC) 是一种动态参数, 能反

映了语音参数的动态特性, 有较好的鲁棒性。差分参数的计算公式为:

$$d(n) = \frac{1}{\sqrt{\sum_{i=-k}^k i^2}} \sum_{i=-k}^k i \times c(n+i) \quad (9)$$

其中 c 是 MFCC 参数, d 为一阶差分 MFCC, k 为常数, 通常取 2。

2.2 矢量量化

矢量量化 (VQ) 是 20 世纪 80 年代发展起来的信源压缩编码技术, 它具有很好的分类特性, 在语音编码、语音识别、语音合成和图像数据压缩等领域有着广泛的应用。基于 VQ 法的声纹识别, 特别是小样本库声纹识别无需考虑复杂的统计模型和复杂的时间归整问题, 其运算过程也较为简单, 因此, VQ 法在声纹识别领域有着广泛的应用 (Burton *et al.*, 1985; He *et al.*, 1999)。

本研究采用 LBG 算法 (Linde-Buzo-Gray algorithm) 形成代表声纹特征的 VQ 码本, LBG 算法是由 Linde Y, Buzo A 和 Gray RM 三人提出的 (Linde *et al.*, 1988)。此算法按最近邻准则用初始码本中的各个码字对训练序列进行 Voronoi 划分, 从而形成 J 个子集, 每一子集为一类, J 为码本容量; 计算各类的形心和平均失真, 迭代计算下去, 不断对码本进行修正直到性能满足要求或不能再有明显改进为止。VQ 码本形成的 LBG 算法 (杨行峻和迟惠生, 1995) 具体过程为:

(1) 输入形成码本所需的全部特征矢量 X , X 的集合记作 S ;

(2) 设置最大迭代次数 N , 畸变初值 $D^0 = \infty$, 畸变改进阈值 δ , 迭代初值 $n = 1$;

(3) 设置 J 个码字的初值 $Y_1^0, Y_2^0, \dots, Y_J^0$;

(4) 根据最近邻准则将 S 分成 J 个子集 $S_1^n, S_2^n, \dots, S_J^n$, 当 $X \in S_k^n$ 时, 有:

$$d(X, Y_k^{n-1}) \leq d(X, Y_i^{n-1}) \quad \forall i, i \neq k \quad (10)$$

其中 d 为欧氏距离。

(5) 计算总畸变 D^n

$$D^n = \sum_{k=1}^J \sum_{X \in S_k^{n-1}} d(X, Y_k^{n-1}) \quad (11)$$

(6) 计算畸变改进阈值 δ^n

$$\delta^n = \frac{\Delta D^n}{D^n} = \frac{|D^{n-1} - D^n|}{D^n} \quad (12)$$

(7) 计算新码字 $Y_1^n, Y_2^n, \dots, Y_J^n$, 即求各类新的聚类中心;

(8) 判断 $\delta^n < \delta$ 是否成立, 如果成立, 转入步

骤(10), 反之, 转入步骤(9);

(9)判断 n 是否小于最大迭代次数 N , 是, 则令 $n = n + 1$, 转入步骤(4), 否, 则转入步骤(10);

(10)迭代终止, 输出 $Y_1^n, Y_2^n, \dots, Y_I^n$ 作为最终码字。

本文中初始码本的选择采用了分裂法, 20 组滤波器输出的 MFCC 作为特征矢量集合 S 。

2.3 特征匹配

用于测试的声纹数据同样要经过特征提取过程得到原始特征矢量即 MFCC, 然后计算特征矢量与各码本的距离, 该距离为各帧声音的原始特征矢量与码本中码字最小距离的总和, 即:

$$d_n = \sum_{i=1}^I \min_{1 \leq j \leq J} \sqrt{\sum_{k=1}^L (x_{ik} - y_{jk}^n)^2} \quad 1 \leq n \leq N \quad (13)$$

其中, I 为测试样本提取得到的 MFCC 矢量数, J 为训练得到的 VQ 码本中矢量个数, L 为 MFCC 及 VQ 矢量的长度, N 为库中模板个数, d_n 为测试样本特征矢量与第 n 个 VQ 码本的距离, 在 $n = 1, \dots, N$ 中, 使 d_n 当取最小值时的 n 对应的昆虫即为识别结果。

3 识别试验结果

以上算法在以 Matlab 为平台的 PC 机上进行了测试, 试验所用昆虫声纹数据库中共录制了 70 种昆虫发出的不同声音, 经过预处理和分割共得到 488 个声音样本, 其中每种声音的第一个样本用于训练得到 VQ 码本, 其余样本用于测试。表 1 显示了 VQ 分群数 k 分别为 8, 16, 32, 64 时所得的识别结果。从表 1 可知, 随着 VQ 分群数的增加, VQ 训练时间及匹配时间成倍增加, 而识别率并没有显著的提高, 因此, 综合考虑识别时间性能和识别正确率, 取 $k = 16$ 较为理想。

表 2 为本文试验所用数据库中各类别识别性能比较, 库中录制声音分为仓储昆虫活动和取食声音 8 种、土壤无脊椎动物活动和取食声音 16 种、土壤昆虫防御性鸣叫声 2 种、树林昆虫活动和取食声音 14 种、作物昆虫活动和取食声音 1 种、翅膀或腹腔振动声音 22 种, 小蠹声音 7 种。

表 1 本文昆虫声纹自动鉴别算法性能
Table 1 Performance of the proposed automatic acoustic insect identification algorithm

VQ 分群数 Number of VQ groups k	8	16	32	64
预处理时间 Preprocessing time (ms)	364	364	364	364
MFCC 特征提取时间 MFCC feature extraction time (ms)	370	370	370	370
VQ 训练时间 VQ training time (ms)	19	47	75	134
匹配时间 Matching time (ms)	103	200	350	1 114
识别精度 Recognition accuracy (%)	95.41	96.17	96.26	95.91

表 2 各类昆虫声纹识别率比较
Table 2 Comparison among recognition accuracy of different kinds of insect sounds

昆虫声音种类 Insect sound types	$k = 8$	$k = 16$	$k = 32$	$k = 64$
仓储昆虫活动和取食声音 Stored product insect movement and feeding sounds	0.980	1.000	0.99	1.000
土壤无脊椎动物活动和取食声音 Movement and feeding sounds of soil invertebrates	1.000	1.000	1.000	1.000
土壤昆虫防御性鸣叫声 Defensive stridulation of soil insects	0.600	0.636	0.867	0.909
树林昆虫活动和取食声音 Movement and feeding sounds of insects in wood	1.000	1.000	1.000	1.000
作物昆虫活动和取食声音 Movement and feeding sounds of insects in plants	1.000	1.000	1.000	1.000
翅膀或腹腔振动声音 Wing and abdominal vibration sounds	0.889	0.903	0.898	0.876
小蠹声音 Scolytidae sounds	0.872	0.971	0.986	1.000

k = VQ 分群数 Number of VQ groups.

试验表明, 同类昆虫的雌虫与雄虫发出的声音所提取的特征比较接近(尤其是翅膀或腹腔振动声音), 有时甚至难以区分并影响了最终的识别精度; 另一方面也说明, 用本文方法提取的特征能代表昆虫声纹的本质属性, 同类昆虫的雌虫和雄虫虽然在生理结构上存在差异, 但总体的发声机制还是相近的, 因此声纹提取的特征信息也比较接近。

4 结论与讨论

本文借鉴了人类语音识别方法中所采用的模式识别技术, 通过信号参数化方法来实现昆虫声纹的自动鉴定和识别。该方法表现较高的识别精度与较理想的时间性能。该方法可以直接扩展到其他能发出固有声音的生物的种类鉴别, 在非侵犯性声学环境检测方面存在应用前景。

本研究对仓储昆虫活动和取食声音, 土壤无脊椎动物活动和取食声音, 土壤昆虫防御性鸣叫声, 树林昆虫活动和取食声音, 作物昆虫活动和取食声音、翅膀或腹腔振动声音以及小蠹声音等 70 种仓虫进行了识别试验, 总的识别率超过 96%, 在包含 70 种昆虫的声纹库中作一次匹配用时仅需 0.2 s 左右, 试验表明该方法切实可行, 有较大的实用推广价值。

同类昆虫的雌虫与雄虫发出的声音在本研究中难以区分, 说明本文提出的方法尚不能达到同类昆虫的性别识别精度, 尤其是对翅膀或腹腔振动声音。那么, 自然界中昆虫翅膀或腹腔振动声音是否具有吸引配偶的意义? 或者自然界中昆虫吸引配偶的声音特征是否需要更好的分析方法? 这些问题都是昆虫的性生态学感兴趣的。

此外, 本文试验录制的声音文件都只包括单个昆虫发出的声音, 将来我们将从多种混合声音包括背景声音中对某一种或几种昆虫的“存在”和“状态”的检测与判断进行探索与研究。

参 考 文 献 (References)

- Alexander R, 1957. Sound production and associated behavior in insects. *The Ohio Journal of Science*, 57(2): 101–113.
- Burton DK, Shore JE, Burk JT, 1985. Isolated word speech recognition using multi-VQ code books. *IEEE Trans. ASSP*, 33(4): 837–849.
- Chesmore D, 2004. Automated bioacoustic identification of species. *Annals of the Brazilian Academy of Sciences*, 76(2): 435–440.
- Chesmore ED, Nellenbach C, 2001. Acoustic methods for the automated detection and identification of insects. *Acta Horticulturae*, 562: 223–231.
- Drosopoulos S, Claridge M, 2005. Insect Sounds and Communication: Physiology, Behaviour, Ecology, and Evolution. Contemporary Topics in Entomology. CRC Press, Boca Raton, FL. 552 pp.
- Ganchev T, Potamitis I, Fakotakis N, 2007. Acoustic monitoring of singing insects. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing, IV. Honolulu. 721–724
- Han P, 2003. Voice-pattern recognition of stored-product insects. *Computer Engineering*, 29(22): 151–154. [韩萍, 2003. 仓储物害虫声音的模式识别. 计算机工程, 29(22): 151–154]
- He JL, Liu L, Palm G, 1999. A discriminative training algorithm for VQ-based speaker identification. *IEEE Transactions on Speech and Audio Processing*, 7(3): 353–356.
- Linde Y, Buzo A, Gray RM, 1988. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28(1): 84–95.
- Mankin R, 2009. Sound Library. <http://www.ars.usda.gov/pandp/docs.htm?docid=10919#albopictus>.
- Pinhas J, Soroker V, Hetzroni A, Mizrach A, Teicher M, Goldberger J, 2008. Automatic acoustic detection of the red palm weevil. *Computers and Electronics in Agriculture*, 63(2): 131–139.
- Riede K, 1998. Acoustic monitoring of Orthoptera and its potential for conservation. *Journal of Insect Conservation*, 2: 217–223.
- Yang XJ, Chi HS, 1995. Voice Digital Signal Processing. Publishing House of Electronics Industry, Beijing. 454 pp. [杨行峻, 迟惠生, 1995. 语音信号数字处理. 北京: 电子工业出版社. 454 页]
- Zhen B, Wu XH, Liu ZM, Chi HS, 2001. On the importance of components of the MFCC in speech and speaker recognition. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 37(3): 371–378. [甄斌, 吴玺宏, 刘志敏, 迟惠生, 2001. 语音识别和说话人识别中各倒谱分量的相对重要性. 北京大学学报(自然科学版), 37(3): 371–378]

(责任编辑: 袁德成)