

DOI: 10.3969/j.issn.1003-0972.2015.01.031

# 基于分频段 ABAP 谱减法的鸟叫声分类研究

张妍琰<sup>1\*</sup>, 刘建芳<sup>2</sup>, 郭力争<sup>1</sup>

(1. 河南城建学院 计算机科学与工程学院, 河南 平顶山 467036;

2. 平顶山学院 软件学院, 河南 平顶山 467000)

**摘要:** 针对带噪鸟叫声, 用自适应 Bartlett 平均周期图 (ABAP) 完成噪声估计, 进行分频段谱减去噪后, 提取经过二维离散余弦变换的 Mel 频率倒谱系数动态声音特征 (TDMFCC)。最后, 利用支持向量机 (SVM) 分别结合 MFCC、TDMFCC 以及经过分频段 ABAP 谱减法降噪后的 MFCC 和 TDMFCC 对 30 种鸟叫声进行不同背景环境和信噪比情况下的对比实验。结果表明, 分频段 ABAP 谱减法降噪后提取 TDMFCC 结合 SVM 的方法可以取得较好的分类效果, 适用于噪声环境下的鸟叫声分类。

**关键词:** 鸟叫声分类; 自适应 Bartlett 平均周期图; 分频段谱减法

中图分类号: TP391.4 文献标志码: A 文章编号: 1003-0972(2015)01-0127-06

## Study of the Bird Sound Classification Based on ABAP Multi-band Spectral Subtraction

Zhang Yanyan<sup>1\*</sup>, Liu Jianfang<sup>2</sup>, Guo Lizheng<sup>1</sup>

(1. Institute of Computer Science and Engineering, Henan University of Urban Construction, Pingdingshan 467036, China;

2. School of Software, Pingdingshan University, Pingdingshan 467000, China)

**Abstract:** To the bird sounds mixed with noises, the estimation of noise is done by using the adaptive Bartlett averaging periodogram (ABAP), and noise reduction is made by using multi-band spectral subtraction, then the dynamic feature of Mel-frequency spectrum coefficients by two-dimensional discrete cosine transform (TDMFCC) was extracted. The comparison experiments of 30 bird sounds classification in different environments under different SNRs were constructed based on the combination of SVM classifier and different features, namely MFCC, TDMFCC and MFCC, TDMFCC through ABAP multi-band spectral subtraction respectively. The results showed that the method of TDMFCC through ABAP multi-band spectral subtraction combined SVM can achieve better classification effect, which is very suitable for the bird sound classification in noisy environments.

**Key words:** bird sound classification; adaptive bartlett averaging periodogram; multi-band spectral subtraction

## 0 引言

鸟类是自然生物的重要成员之一, 在物种演化的漫长过程中, 大多鸟类都形成了具有自身特异性的鸣叫声。通过对鸟叫声的研究, 有利于提高对鸟类资源、种类、数量、生存区域统计调查的准确性、真实性和可靠性, 同时可将其作为分析自然声音的组成结构和动物行为的典型样例。

自然环境中的鸟叫声通常伴有杂音, 为了去除噪声, 得到较为纯净的声音信号, 必须借助于各种声音增强算法。谱减法<sup>[1]</sup>是声音增强算法中重要的非参数方法, 谱减法以其原理简单、运算量小、性能稳定而受到最为广泛的关注和研

究。传统谱减法<sup>[2-3]</sup>采用的周期图估计并非一致估计, 方差很大, 并造成“音乐噪声”, 影响谱减后声音分类效果, 尤其是鸟类声音的分类, 因为鸟叫声中夹杂的自然环境噪声组成复杂、种类繁多, 且在频域上分布较广。为了抑制“音乐噪声”, 降低方差, Hendriks 等人通过搜索功率谱的最长平稳段, 进行周期图平滑, 进而减小周期图估计的方差<sup>[4]</sup>。但是因果系统并不能得到理论上的最长平稳段, 该搜索过程还需要较大的数据存储量和计算量, 并且这种周期图平滑利用的只是时域的平稳性, 却没有考虑功率谱频域上的结构特征, 无法应对复杂的自然声音。Ephraim 等人提出的引导决策 (Decision-Directed, DD) 算法能够较好地估计高信

收稿日期: 2014-04-02; 修订日期: 2014-08-24; \* . 通信联系人, E-mail: zzyknyl@163.com

基金项目: 河南省科技计划重点科技攻关项目 (122102210412)

作者简介: 张妍琰 (1981-), 女, 河南平顶山人, 讲师, 硕士, 主要从事计算机应用技术研究。

噪比情况下的功率谱<sup>[5]</sup>. 但是当信噪比较低时, DD 算法会低估先验信噪比, 造成跟踪延迟和声音失真的问题<sup>[6]</sup>. Gustafsson 等人采用 Bartlett 平均周期图 (Bartlett Averaging Periodogram, BAP) 方法先将信号分段估计, 然后平均各个分段结果<sup>[7]</sup>, 以减小估计方差, 平滑功率谱图. 但是该方法造成严重的能量泄露, 影响算法性能. Welch 方法引进重叠加窗改进了 Bartlett 法, 虽然能在一定程度上抑制能量泄露, 但是窗函数的选择对估计效果影响较大, 需事先针对不同信号和不同的处理目的选择合适的窗函数. 实际上, 噪声功率谱的结构特征变化复杂, 既可能是平坦的连续谱, 如白噪声谱; 也包括离散的线谱, 如正弦信号等周期信号的功率谱; 更普遍的是既包括连续谱又包括非连续谱的复杂谱. 对于连续谱噪声, 应采用频率分辨率较低的功率谱估计以减小方差, 从而减少“音乐噪声”; 而对非连续谱噪声, 则应该采用频率分辨率较高的功率谱估计以降低噪声谱能量泄露, 减少声音失真. 因此无论哪种功率谱估计方法, 都难以在方差和频率分辨率之间取得良好的折中.

针对上述问题, 本文提出了基于噪声谱结构特征的自适应 Bartlett 平均周期图 (Adaptive Bartlett Averaging Periodogram, ABAP) 方法. 根据噪声谱的结构特征对周期图进行自适应平滑. 真实环境中的噪声对整个声音信号整个频谱的影响并不均匀<sup>[8]</sup>, 因此本文将 ABAP 应用于频段谱减法 (Multi-band Spectral Subtraction, MSS)<sup>[9]</sup> 对自然环境的鸟叫声进行降噪处理, 以减少音乐噪声和声音失真. 传统的 Mel 频率倒谱系数 (Mel Frequency Cepstrum Coefficient, MFCC) 充分利用人耳听觉感知特性, 经常被作为声音特征用于分类识别<sup>[10]</sup>, 但是 MFCC 遇到噪声性能骤降. 因此本文将含噪声的鸟叫声信号经过前端改进的降噪处理后, 在传统的 MFCC 基础上加入二维离散余弦变换<sup>[11]</sup>, 提取更具有针对性、动态性和鲁棒性的二维 Mel 频率倒谱系数 (Two Dimensional Mel-Frequency Cepstrum Coefficients, TDMFCC) 来表征鸟叫声信号, 最后结合判决分类器支持向量机 (Support Vector Machine, SVM) 来对含噪声的鸟叫声进行分类识别.

## 1 前端降噪

### 1.1 ABAP 法

假设纯净声音为  $s(n)$ , 噪声为  $d(n)$ , 则带噪声声音表示为  $y(n) = s(n) + d(n)$ . 去噪的目的就是从复杂的带噪声声音  $y(n)$  中估计出纯净声音  $s(n)$ .

$Y_N(k, l)$  为  $y(n)$  第  $l$  帧  $N$  点快速傅里叶变换 (FFT) 第  $k$  个频带分量, 其相应的周期图为<sup>[12]</sup>

$$I_{YY}^N(k, l) = |Y_N(k, l)|^2, \quad (1)$$

直接周期图法估计的功率谱方差过大, 因此引入 Bartlett 平均周期图 (BAP) 法, 对信号进行帧内分段估计, 然后取平均值.

$Y_M(k, l)$  是第  $l$  帧  $M$  点 FFT 第  $k$  个频带分量, 其相应的 Bartlett 平均周期图表示为

$$\bar{I}_{YY}^M(k, l) = \sum_{i=-\frac{L}{2}}^{\frac{L}{2}-1} |Y_M(k, l-i)|^2 / L, \quad (2)$$

其中:  $N = LM$ ,  $L$  是 BAP 的平均段数,  $M$  是每段的样本点数. 通过帧内分段平滑, BAP 有效降低功率谱方差.

对较为平稳的声音功率谱, BAP 法有良好的效果, 能有效弥补直接周期图法功率谱估计存在的方差过大的缺陷. 而对于波峰较窄的信号, 当在频域上变化较大时, 所估计 BAP 分辨率过低, 能量泄露较为严重, 不能准确反应声音信号. 因此在 BAP 法的基础上, 引入自适应频带结构特性平滑机制, 形成自适应 Bartlett 平均周期图 (ABAP) 法来进行噪声功率谱估计. 分析噪声谱的结构特性, 根据频带间连续程度, 将其区分为连续谱和非连续谱. 判断噪声谱是否频带间连续, 如果是, 则在谱线之间做平滑; 否则, 不做平滑.

在频域上将功率谱分成  $L$  个频带, 对  $I_{YY}^N(k, l)$  进行频带间自适应平滑.

$$\bar{I}_{YY}^N(k, l) = \sum_{i=-\frac{L}{2}}^{\frac{L}{2}-1} w_i I_{YY}^N(k-i, l), \quad (3)$$

其中, 权值  $w_i$  满足:

$$w_i \geq 0, \sum_{i=-\frac{L}{2}}^{\frac{L}{2}-1} w_i = 1, \quad (4)$$

引入当前频带与相邻频带的功率谱期望比值, 作为判断噪声谱是否连续的依据, 即:

$$\phi(k, i) = \frac{E(P_D(k))}{E(P_D(k+i))}, \quad i = \pm 1, \pm 2, \dots, \pm K_f. \quad (5)$$

其中  $E(P_D(k))$  为噪声功率谱期望. 当  $\phi(k, i)$  越接近 1, 表示频带  $k$  与频带  $k+i$  的期望越接近, 其谱结构越相似, 因此这两频带间可以进行平滑以减少方差.

但是判断每个频带与周围  $2K_f$  个频带的结构连续性, 所需要的运算量过大, 无法满足声音分类实时性的要求. 因此, 考虑频带  $k$  与相邻  $2K_f+1$  个频带功率谱均值的比作为判决依据.

$$\lambda(k) = \frac{P_D(k)}{\sum_{i=-K_f}^{K_f} P_D(k+i) / (2K_f+1)}. \quad (6)$$

为了防止异常噪声点大值的影响, 利用调和平均数来代替功率谱均值的计算.

$$\hat{\lambda}(k) = \frac{P_D(k)}{(2K_f+1) / \sum_{i=-K_f}^{K_f} \frac{1}{P_D(k+i)}}, \quad (7)$$

其中, 实验中为  $P_D(k+i)$  设置下限值 0.5, 以排除出现零点时均值过小的误差.

当  $\hat{\lambda}(k) \geq \lambda_{th}$  时, 判定频带  $k$  为非连续谱, 当前频带不能与邻近频带平滑; 反之, 当  $\hat{\lambda}(k) < \lambda_{th}$  时, 则判决当前频带为连续谱, 可以与邻近频带做平滑处理. 其中  $\lambda_{th}$  为判决门限, 可通过虚警概率计算得到, 典型取值为 3.

计算判决依据, 并确定当前频带的权值  $w_k$ ,

$$w_k = \begin{cases} 1 & \hat{\lambda}(k) \leq \lambda_{th} \\ 0 & \text{else.} \end{cases} \quad (8)$$

计算平滑周期图  $\bar{I}_{YY}^N(k)$  和平滑的噪声谱  $\bar{P}_D(k)$ ,

$$\bar{I}_{YY}^N(k) = \begin{cases} \sum_{i=-k_f}^{k_f} w_{k+i} I_{YY}^N(k+i) / \sum_{i=-k_f}^{k_f} w_{k+i} \mu w_k \neq 0, \\ I_{YY}^N(k), & \text{else,} \end{cases} \quad (9)$$

$$\bar{P}_D(k) = \begin{cases} \sum_{i=-k_f}^{k_f} w_{k+i} P_D(k+i) / \sum_{i=-k_f}^{k_f} w_{k+i} \mu w_k \neq 0, \\ P_D(k), & \text{else.} \end{cases} \quad (10)$$

图 1 中 (a)、(b)、(c) 分别展示了周期信号加带通白噪声的 3 种功率谱估计方法的效果。(a) 中直接周期图法的帧长  $N$  取 32 ms 的采样点数, (b) 中 BAP 法将  $N$  划分为  $L$  段, 每段的  $M$  取 4 ms 的采样点数。

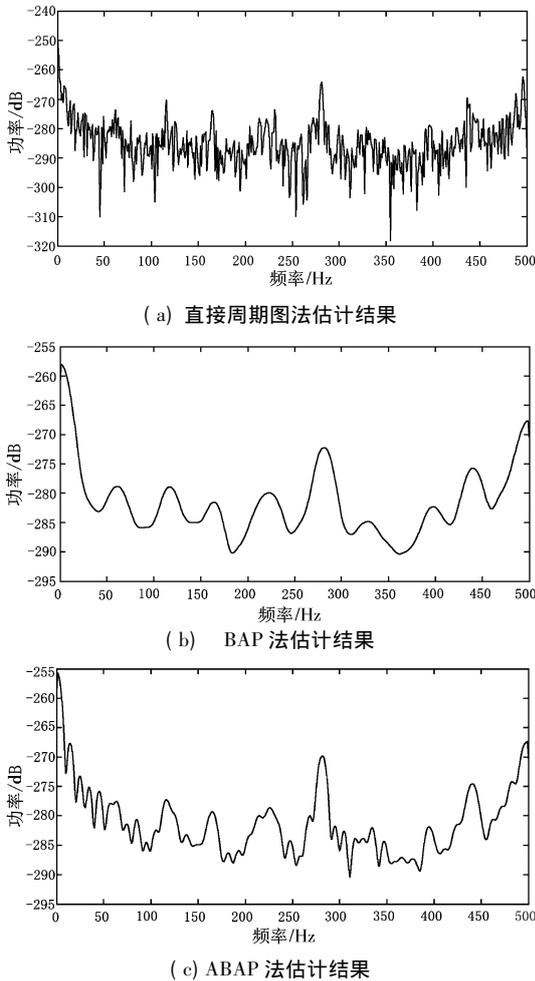


图 1 周期信号加带通白噪声的噪声功率谱估计结果

Fig. 1 Noise power spectrum estimation result of periodic signal with band-pass white noise

由图 1 的 3 个对比图可知, (a) 中的直接周期图法有较高的分辨率, 能量泄露较小, 对周期信号具有良好的估计特性, 能有效估计窄带噪声, 但是对像白噪声这样的宽带噪声, 估计的方差很大; (b) 中的 BAP 法虽然方差较小, 但能量泄露比较严重, 几乎看不到谱线结构, 只适用于连续谱估计; (c) 中即本文提出的 ABAP 法, 能够根据频带间谱结构的连续性来自适应地平滑周期图, 较好地缓解了传统周期图估计中频率分辨率与能量泄露之间的矛盾。

### 1.2 分频段 ABAP 谱减法

自然环境的有色背景噪声(风雨声、流水声等)大多是非平稳信号, 其均值、功率谱、相关函数等统计量具有随机性, 对声音信号功率谱各频段的影响也具有不确定性。因此本文针对不同频率段中噪声谱的影响程度, 引入分频段谱减法, 并将 ABAP 的噪声估计方法应用于其中, 形成分频段 ABAP 谱减法来对带噪声进行降噪, 以此提高声音质量。

将带噪声声音信号的功率谱分成  $N$  个不重叠的子频带  $i$ , 分频带作差异谱减, 谱减后第  $i$  子频带的前景鸟叫声功率谱估计  $|\hat{S}_i(k, l)|^2$  可表示成

$$|\hat{S}_i(k, l)|^2 = |Y_i(k, l)|^2 - \alpha_i(l) \delta_i(l) |D_i(k, l)|^2, \quad b_i \leq k \leq e_i, \quad (11)$$

其中:  $|Y_i(k, l)|^2$  为原始带噪声声音功率谱;  $|D_i(k, l)|^2$  为用 1.1 的 ABAP 法所估计出的噪声功率谱;  $b_i$  和  $e_i$  分别为第  $i$  子频带上频率窗口的初始和结束;  $k$  表示频率索引;  $l$  表示帧索引;  $\alpha_i(l)$  表示第  $l$  帧第  $i$  频带的过估计系数, 其值通过式 (14) 中第  $l$  帧第  $i$  子带的信噪比  $SNR_i(l)$  来分段确定。

$$SNR_i(l) = 10 \lg \left( \frac{\sum_{k=b_i}^{e_i} |Y_i(k, l)|^2}{\sum_{k=b_i}^{e_i} |D_i(k, l)|^2} \right). \quad (12)$$

再根据  $SNR_i(l)$  和实际实验将  $\alpha_i(l)$  设定为:

$$\alpha_i(l) = \begin{cases} 5, & SNR_i(l) < -5, \\ 4 - 0.15 SNR_i(l), & -5 \leq SNR_i(l) \leq 20, \\ 1, & SNR_i(l) > 20. \end{cases} \quad (13)$$

$\alpha_i(l)$  随着信噪比非线性变化, 对于信噪比较低的频段  $\alpha_i(l)$  的取值较大, 这使得谱减过程中的噪声干扰被有效抑制。

式 (11) 中的  $\delta_i(l)$  代表用来设定噪声移除特性的补偿因子, 根据声音的实际情况和经验设置为:

$$\delta_i(l) = \begin{cases} 2.5, & f_i(l) \leq 1 \text{ kHz}, \\ 1.5, & 1 \text{ kHz} < f_i(l) \leq \frac{F_s}{2} - 3 \text{ kHz}, \\ 1.2, & \frac{F_s}{2} - 3 \text{ kHz} < f_i(l) < \frac{F_s}{2} - 2 \text{ kHz}, \\ 0.8, & f_i(l) > \frac{F_s}{2} - 2 \text{ kHz}. \end{cases} \quad (14)$$

其中:  $f_i(l)$  为第  $l$  帧的第  $i$  个子频带的最高频率,  $F_s$  为声音信号的采样率。

最后, 对上述谱减后出现负值的情况进行下限修改,

$$|\hat{S}_i(k, l)|^2 = \begin{cases} |\hat{S}_i(k, l)|^2, \\ |\hat{S}_i(k, l)|^2 \geq S_{\min}(k, l), \\ S_{\min}, & \text{else.} \end{cases} \quad (15)$$

$$S_{\min}(k, l) = S_{\min} - (1 - w_k) S_{\max}, \quad (16)$$

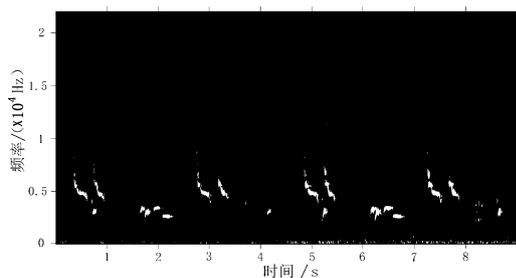
其中:  $S_{\min}$  取  $0.1 |Y_i(k, l)|^2$ ,  $S_{\max}$  取  $0.9 S_{\min}$ 。

采用式 (15)、(16) 主要基于以下两点考虑: 第一, 对于连续谱, 保留残留噪声不变, 利用其宽带特性掩蔽“音乐噪声”; 第二, 与连续谱相比, 非连续谱往往具有更大的能量, 因此需要对其进行更多抑制, 使得残留噪声谱更加平坦, 减

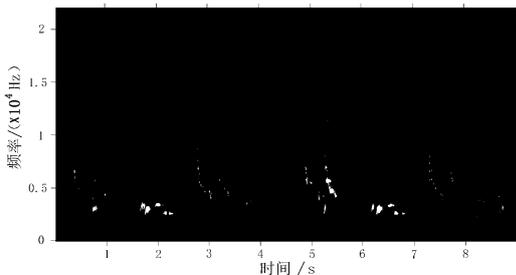
少音调音出现.

### 1.3 分频段 ABAP 谱减法性能测试

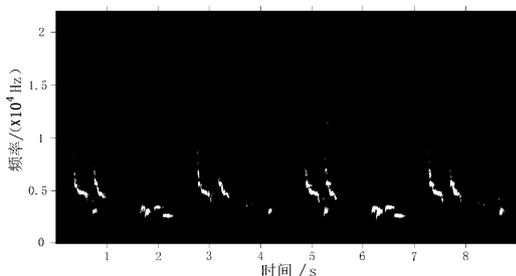
为了测试分频段 ABAP 谱减法对自然环境噪声移除的有效性, 将直接周期图估计谱减法、BAP 估计谱减法与本文提出的分频段 ABAP 谱减法进行降噪性能对比实验. 对添加 10 dB 溪流噪声的画眉鸟叫声, 分别使用上述 3 种不同方法进行降噪处理, 其效果如图 2(a)、(b)、(c) 的 3 种声谱图所示. 由图 2 可知: (1) 使用直接周期图估计谱减法对带溪流噪声的画眉鸟叫声降噪后, 依然存在大量音乐噪声. (2) 使用 BAP 估计谱减法对画眉鸟叫声降噪后, 虽然没有音乐噪声, 但是低频段声音被严重抑制, 信号能量泄露严重, 声音失真较为严重. (3) 使用分频段 ABAP 谱减法降噪后, 不但没有残留音乐噪声, 而且低频段的声音失真也较少, 降噪效果较好.



(a) 直接周期图估计谱减法降噪后的画眉声谱图



(b) BAP 估计谱减法降噪后的画眉声谱图



(c) 分频段 ABAP 谱减法降噪后的画眉声谱图

图 2 不同降噪方法对画眉鸟叫声的降噪效果

Fig. 2 The effect of different noise reduction methods for wood thrush sound

综上所述, 本文提出的分频段 ABAP 谱减法既能解决噪声估计方差太大的问题, 减少音乐噪声, 又能有效解决因能量泄露而导致的失真问题.

## 2 TDMFCC 特征提取

多频段的频谱结构变化包含了声音信号的动态特征, 为了更加充分地利用声音信号的动态和静态特征, 本文引入二维离散余弦变换<sup>[13-14]</sup>的 Mel 频率倒谱系数 (Two Dimensional Mel-Frequency Cepstrum Coefficients, TDMFCC).

Mel 频率倒谱系数<sup>[15-16]</sup>是目前广泛使用的声音识别特征, 信号的第  $t$  帧第  $q$  个 MFCC 系数可以表示为:

$$c_i(q) = \sum_{b=0}^{B-1} \log(E_i(b)) \cos\left(\frac{(2b+1)q\pi}{2B}\right), \quad (17)$$

其中:  $E_i(b)$  表示通过第  $t$  帧第  $b$  个 Mel 滤波器的能量,  $B$  是 Mel 滤波器的总数量. 经过二次离散余弦变换 (Discrete Cosine Transform, DCT) 后的 TDMFCC 矩阵可以表示为:

$$c(q, n) = \sum_{t=0}^{L-1} c_i(q) \cos\left(\frac{(2t+1)n\pi}{2L}\right), \quad 0 \leq q < B, 0 \leq n < L. \quad (18)$$

其中:  $n$  是调制频率的索引,  $L$  是声音信号的总帧数.

TDMFCC 的结构图如图 3 所示, 通过对连续对数频谱序列进行横向二次 DCT 变换, 可以得到包含各序列间频谱变化信息的 TDMFCC 矩阵. 但是由于二次 DCT 变换使得频谱信息的存储量增加, 因此需要考虑算法的计算效率.

低维的频率参数和时间参数能为声音分类提供更多的全局变化信息. 图 3(3) 中的 A、D 区域主要包含随着时间变化的局部对数谱均值信息, C 区域主要体现频谱结构的细化信息, 而有利于声音分类的对数谱全局变化情况, 主要体现在 B 区域. 为了使用较少的空间获得较多的频谱变化信息, 本文选取 B 区域的 15 行 5 列 TDMFCC 子矩阵作为声音信号的主特征, 用于分类识别, 其表达式如下:

$$\text{TDMFCC} = [c(0, 0), \dots, c(0, A), c(1, 0), \dots, c(1, A), \dots, c(14, 0), \dots, c(14, A)]^T. \quad (19)$$

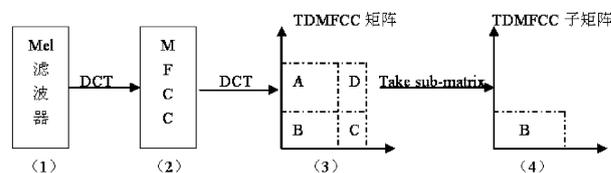


图 3 TDMFCC 矩阵

Fig. 3 The TDMFCC matrix

## 3 结果与分析

实验中用到的鸟叫声来自 Freesound<sup>[17]</sup> 声音数据库, 具体类别如表 1 所示. 每类声音有 50 个声音样本, 每个样本的采样率为 11 025 kHz, 量化精度为 16 bits, 持续时间长度为 4~6 s, 格式为 wav, 单声道. 随机选择 20 个样本用于训练, 另外 30 个样本用于测试. 为了模拟实际噪声场景, 将测试声音样本分别与不同背景噪声分别按 40、30、20、10、0 dB 的信噪比合成. 实验所用的背景噪声分别为高斯白噪声、海

浪噪声、溪流噪声,其中高斯白噪声用噪声发生器生成,海浪噪声及溪流噪声从实地采集。

实验直接使用 LIBSVM<sup>[18]</sup> 工具箱,选用径向基核函数,采用自动寻优方式设置惩罚因子等参数,利用“一对一”的多分类策略,来对声音信号进行 SVM 建模分类,得出识别结果。

表 1 实验数据

Tab. 1 Experimental data

声音类型	具体类别
前景鸟叫声	公画眉,水秧鸡,野山鸡母,毛鸡,朱斑鸠,山斑鸠,绣眼,董鸡,黑水鸡,白面鸡,母鹌鹑,公鹌鹑,竹鸡母,公野鸡,小毛鸡,鸚鵡,燕子,西雀,布谷鸟,海鸥,八哥,鸽子,天鹅,麻雀,雨燕,孔雀,金丝雀,火鸡,翠鸟,知更鸟
背景噪声	高斯白噪声,海浪噪声,溪流噪声

实验旨在验证本文所提方法对噪声环境下的鸟叫声分类的有效性,因此分别对 MFCC、TDMFCC 以及用文中所提的分频段 ABAP 谱减法降噪后的 MFCC(标记为 ABAP\_MSS + MFCC)、分频段 ABAP 谱减法降噪后的 TDMFCC(标记为 ABAP\_MSS + TDMFCC) 进行 SVM 建模分类的声音识别实验,前 3 个为对照实验,第 4 个为目标实验。针对高斯白噪声、海浪噪声、溪流噪声这 3 种不同噪声下不同信噪比的平均分类结果如表 2 所示。而针对这 3 种不同噪声类型和不同信噪比所对应的实验结果如图 4 所示。

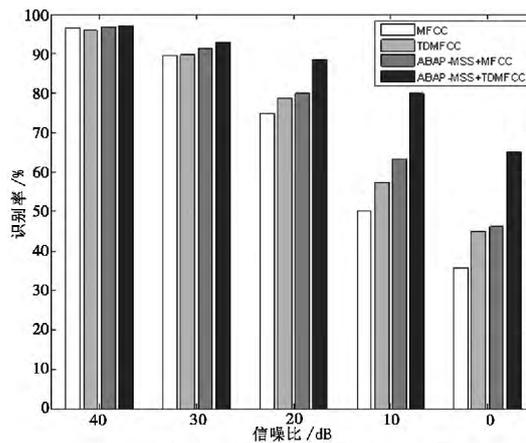
表 2 不同噪声类型下不同信噪比的平均分类结果

Tab. 2 The average classification rates of different SNRs under different types of noises

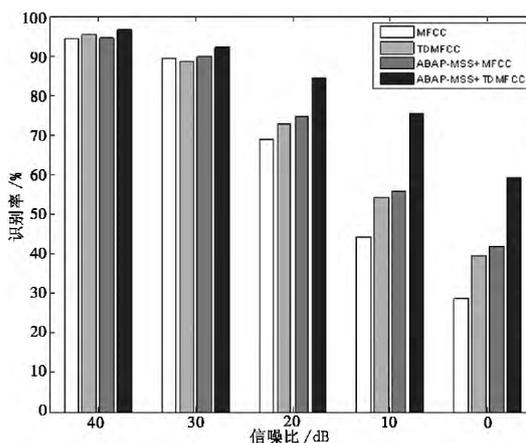
特征类别	识别率/%		
	高斯白噪声	海浪噪声	溪流噪声
MFCC	69.37	65.18	67.46
TDMFCC	73.76	70.60	71.81
ABAP_MSS + MFCC	75.44	71.53	74.69
ABAP_MSS + TDMFCC	85.25	81.72	84.14

由表 2 的实验结果可以看出,虽然是在高斯白噪声、海浪噪声和溪流噪声这 3 种不同背景噪声类型下,但是其鸟叫声的平均识别率具有相类似的规律:MFCC 的平均识别率最低,而 ABAP\_MSS + TDMFCC 是四个特征中识别率最高的。根据该表的数据可得,TDMFCC 比 MFCC 的分类性能提高了至少 4.35%;使用本文提出的 ABAP\_MSS 方法降噪后的 MFCC 和 TDMFCC 比不使用该方法的 MFCC 和 TDMFCC 的分类效果更好,其性能分别提高了至少 6.07% 和 11.12%。由此可见,本文提出的经过 ABAP\_MSS 方法降噪后的 TDMFCC 识别性能更好,对不同类型的噪声更具有鲁棒性。

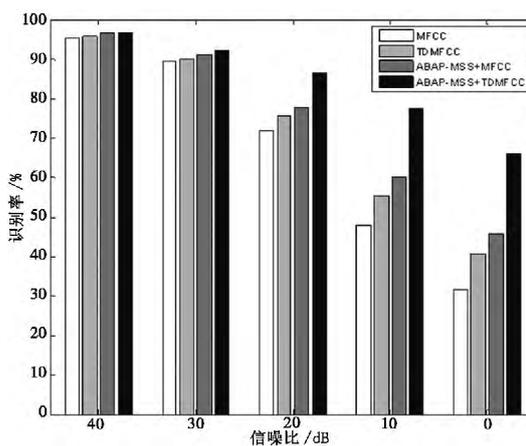
由图 4(a)、(b)、(c) 的实验结果柱状图可以看出,无论是平谱噪声还是有色噪声环境下 4 种特征的识别率都



(a) 高斯白噪声下 4 种特征的分类结果



(b) 海浪噪声下 4 种特征的分类结果



(c) 溪流噪声下 4 种特征的分类结果

图 4 不同噪声下不同信噪比的分类结果

Fig. 4 The classification results at different SNRs under different noises

随着信噪比的减小而降低,但是下降的速度有所不同。MFCC 下降的速率最快,TDMFCC 次之,而经过降噪后的 ABAP\_MSS + MFCC 和 ABAP\_MSS + TDMFCC 的识别率下降速度分别为次慢和最慢。说明前面两者受噪声干扰较大,而后面两者对噪声具有一定的鲁棒性。当信噪比为 40 dB 时,4 种特征的识别率都较高,在 90% 以上。而当信噪比下降到 20

dB 以下时, ABAP\_MSS + TDMFCC 的识别率与其他 3 种特征相比, 具有显著提高, 性能明显优于其他. 说明本文提出的 ABAP\_MSS + TDMFCC 具有较好的抗噪功能, 尤其是在低信噪比时, 比如小于 20 dB, 效果显著.

#### 4 结论

针对噪声环境下的鸟叫声分类问题, 提出一种经过自适应 Bartlett 平均周期图谱减法( ABAP\_MSS) 降噪后提取 TDMFCC 特征结合 SVM 分类器对带噪鸟叫声进行分类的方法. 不同噪声类型下和不同信噪比下的实验结果证明, 该方法的分类性能较现有的 MFCC 与降噪后的 MF-

CC 更优. ABAP\_MSS 降噪方法能够有效减少音乐噪声和声音失真, 削弱噪声干扰. 因此, 当信噪比较高时, ABAP\_MSS + TDMFCC 的分类效果略微优于其他方法; 但是当信噪比较低时, 该方法的分类效果具有显著提高. 由此说明, 本文提出的 ABAP\_MSS + TDMFCC 具有良好抗噪能力, 且适用于带噪鸟叫声的分类. 但是, 当信噪比处于 0 dB 以下时, 降噪方法可能带来音乐噪声和失真问题, 导致分类效果变差. 更低信噪比下的分类问题, 以及将分类方法推广至其他应用中的研究工作天在进行中.

#### 参考文献:

- [1] Yamashita K, Shimamura T. Nonstationary noise estimation using low-frequency regions for spectral subtraction [J]. *IEEE Signal Processing Letters*, 2005, 12(6): 465-468.
- [2] Martin R. Spectral subtraction based on minimum statistics [C]// *Proceedings of the 7th European Signal Processing Conference*. Toronto, Ontario: Springer Press, 1994: 1182-1185.
- [3] Kuldip P, Belinda S. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain [J]. *Speech Communication*, 2010, 52(5): 450-465.
- [4] Hendriks R C, Heusdens R, Jensen J. Adaptive time segmentation for improved speech enhancement [J]. *IEEE Transactions on SAP*, 2006, 14(6): 2064-2074.
- [5] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator [J]. *IEEE Transactions on ASSP*, 1985, 33(2): 443-445.
- [6] 郑成诗, 周名远, 李晓东. 基于联合语音出现概率的先验信噪比估计算法 [J]. *电子与信息学报*, 2008, 30(7): 1689-1683.
- [7] Gustafsson H, Nordholm S E, Claesson I. Spectral subtraction using reduced delay convolution and adaptive averaging [J]. *IEEE Transactions on SAP*, 2001, 9(8): 799-807.
- [8] 颜鑫, 李应. 利用抗噪零归一化倒谱系数的鸟类声音识别 [J]. *电子学报*, 2013, 41(2): 295-300.
- [9] Wan Y L, Zhang T Q, Wang Z C, et al. Robust speech recognition based on multi-band spectral subtraction [C]// *Proceedings of the 6th International Congress on Image and Signal Processing*. Hangzhou, China: IEEE Computer Society, 2013: 36-40.
- [10] Kinnunen T, Saeidi R, Sedlak F, et al. Low-variance multitaper MFCC features: a case study in robust speaker verification [J]. *IEEE transactions on Audio, Speech, and Language Processing*, 2012, 20(7): 1990-2001.
- [11] Lee C H, Han C C, Chuang C C. Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients [J]. *IEEE Transactions on Audio, Speech, Language Processing*, 2008, 16(8): 1541-1550.
- [12] Martin R. Bias compensation methods for minimum statistics noise power spectral density estimation [J]. *Signal Processing*, 2006, 86: 1215-1229.
- [13] Kim M J, Kim Y, Lim J D, et al. Automatic detection of malicious sound using segmental two-dimensional Mel-frequency cepstral coefficients and histogram of oriented gradients [C]// *Proceedings of the ACM International Conference on Multimedia*. Firenze, Italy: ACM, 2010: 887-890.
- [14] Kim B W, Choi D L, Lim J D, et al. Audio-based automatic detection of objectionable contents in noisy conditions using normalized segmental two-dimensional MFCC [C]// *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Kyoto, Japan: IEEE Computer Society, 2012: 481-484.
- [15] Wu J Q, Yu J J. An improved arithmetic of MFCC in speech recognition system [C]// *Proceedings of the International Conference on Electronics, Communications and Control*. Zhejiang, China: IEEE Computer Society, 2011: 719-722.
- [16] Wang Y T, Li B, Jiang X Q, et al. Speaker recognition based on dynamic MFCC parameters [C]// *Proceedings of the International Conference on Image Analysis and Signal Processing*. Taizhou, China: IEEE Computer Society, 2009: 406-409.
- [17] Universitat Pompeu Fabra. Repository of sound under the creative commons license, Freesound.org [DB/OL]. [2013-12-25]. <http://www.freesound.org/>.
- [18] Chang C C, Lin C J. LIBSVM: a Library for Support Vector Machines [CP/OL]. [2013-12-25]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/libsvm-3.17.zip>.

责任编辑: 任长江