

利用抗噪幂归一化倒谱系数的鸟类声音识别

颜 鑫, 李 应

(福州大学数学与计算机科学学院, 福建福州 350108)

摘 要: 针对真实环境中各种背景噪声下的鸟类声音识别问题, 提出了一种基于新型抗噪特征提取的鸟类声音识别技术. 首先, 根据适用于高度非平稳环境下的噪声估计算法求出噪声功率谱. 其次, 使用多频带谱减法对声音功率谱进行降噪处理. 接着, 结合降噪的声音功率谱提取抗噪幂归一化倒谱系数(APNCC). 最后, 采用支持向量机(SVM)分别对提取的 APNCC, 幂归一化倒谱系数(PNCC)和 Mel 频率倒谱系数(MFCC)对 34 种鸟类声音进行不同环境和信噪比情况下的对比实验. 实验表明, 提取的 APNCC 具有较好的平均识别效果及较强的噪声鲁棒性, 更适用于信噪比低于 30dB 环境下的鸟类声音识别.

关键词: 鸟类声音识别; 非平稳噪声估计; 多频带谱减法; 抗噪幂归一化倒谱系数; Mel 频率倒谱系数

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2013) 02-0295-06

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2013.02.014

Anti-Noise Power Normalized Cepstral Coefficients in Bird Sounds Recognition

YAN Xin, LI Ying

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou, Fujian 350108, China)

Abstract: In order to improve the accuracy of bird sounds recognition under different kinds of noise environments in the real world, a new bird sounds recognition technology based on the APNCC extraction was proposed. First, the noise estimation algorithm for highly non-stationary environments was used to estimate the noise power spectrum of the bird sound in the noise environment. Second, the multi-band spectral subtraction was presented to achieve the background noise reduction. Then, the estimated clean bird sound spectrum was combined with the process of the PNCC extraction to calculate the APNCC. Finally, the comparison experiments of 34 bird sounds recognition in 3 different real environments under different SNRs were constructed, based on the combination of the SVM classifier and 3 different features, namely the APNCC, PNCC and MFCC. The experimental results show that the APNCC outperforms other features in the average recognition rate and the noise robustness, especially for the conditions of all SNRs lower than 30dB.

Key words: bird sounds recognition; non-stationary noise estimation; multi-band spectral subtraction; anti-noise power normalized cepstral coefficients (APNCC); Mel-frequency cepstral coefficients (MFCC)

1 引言

生态环境下的鸟类声音包含着丰富的信息, 鸟类声音识别不仅可以分析鸟类本身的行为等特征, 而且可以对与鸟类有关的外界环境及相关影响领域做分析.

近年来, 借鉴较为成熟的语音识别技术, 鸟类声音识别领域取得了丰富的成果. 文献[1]通过分析鸟类声音音节对的柱状图来进行鸟叫声的分类. 文献[2]使用 MFCC 及高斯混合模型 (Gaussian Mixture Model, GMM) 进行不依赖于个体叫声变化的四种雀形目鸟类声音识别. 滤除背景噪声一直是信号处理领域的基本问题^[3], 在真

实环境中, 各种噪声是普遍存在的. 因此, 对于噪声环境下的鸟类声音识别具有重要的现实意义. 文献[4]通过对基于帧线性预测编码 (Linear Predictive Coding, LPC) 的聚类, 提取噪声下欧亚鸲声音段的有效音节并结合隐马尔可夫模型 (Hidden Markov Model, HMM) 进行分类. 文献[5]使用谱减法, 功率谱分析及自相关分析对噪声下的两种鸟叫进行识别.

环境中的噪声具有非平稳特性, 大多数方法主要针对特定的噪声环境进行鸟类声音识别, 并没有提出一种普遍适用于各种噪声下的鸟类声音识别方法. 因此, 本文提出一种非平稳噪声功率谱估计, 多频带谱减法与

KIM 等人提出的 PNCC 特征提取^[6]相结合的新型抗噪幂归一化倒谱系数 (APNCC) 提取方法, 并使用 SVM 进行鸟类声音识别. 实验表明, 该方法有效提高了不同信噪比环境噪声下的鸟类声音识别效果.

2 噪声功率谱估计

本文根据一种改进的针对高度非平稳噪声具有良好适应性的噪声估计算法^[7]进行噪声估计. 具体流程见图 1 的噪声功率谱估计部分.

2.1 带噪声声音平滑功率谱计算

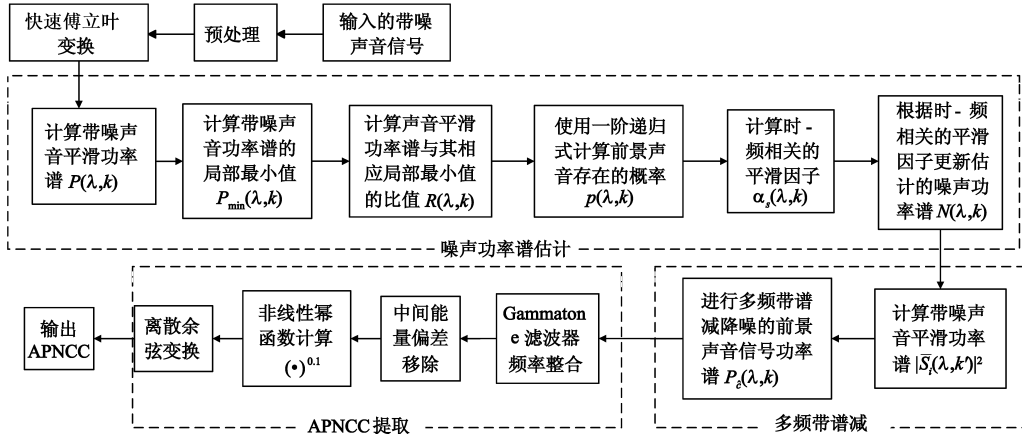


图1 APNCC提取流程图

2.2 带噪声声音平滑功率谱最小值追踪计算

通过连续的平均前面帧的功率谱值, 使用式(3)追踪计算带噪声声音功率谱的最小值.

$$P_{\min}(\lambda, k) = \begin{cases} \gamma P_{\min}(\lambda - 1, k) + \frac{1 - \gamma}{1 - \beta} (P(\lambda, k) - \beta P(\lambda - 1, k)), & P_{\min}(\lambda - 1, k) < P(\lambda, k) \\ P(\lambda, k), & \text{其他} \end{cases} \quad (3)$$

其中, $P_{\min}(\lambda, k)$ 为带噪声声音功率谱的局部最小值, β 和 γ 为常量. 前向因子 β 控制局部最小值的适应时间. 取 $\beta = 0.8$, $\gamma = 0.998$.

2.3 计算待识别前景声音存在的概率

带噪声声音功率谱与其相应局部最小值的比值可以用式(4)表示.

$$R(\lambda, k) = \frac{P(\lambda, k)}{P_{\min}(\lambda, k)} \quad (4)$$

通过该比值与频率相关阈值的比较, 当该比值比相应阈值大时, 该帧的频率窗口就被当作前景声音存在的频率窗口, 反之, 则为噪声的频率窗口. 该思想是基于这样的原则: 当不存在前景声音或者前景声音极为微弱时, 带噪声声音功率谱非常接近其相应的局部最小值. 当式(4)越小时, 该频率窗口只含噪声的可

用式(1)表示带噪声声音.

$$s(t) = c(t) + n(t) \quad (1)$$

其中, $s(t)$ 为带噪声声音, $c(t)$ 为纯净声音, 即待识别前景声音, $n(t)$ 为噪声. 对式(1)进行 (Fast Fourier Transform) 快速傅里叶变换后得到其幅度谱 $S(\lambda, k)$. 带噪声声音的平滑功率谱可以通过式(2)计算.

$$P(\lambda, k) = \eta P(\lambda - 1, k) + (1 - \eta) |S(\lambda, k)|^2 \quad (2)$$

其中, $P(\lambda, k)$ 为带噪声声音的平滑功率谱, λ 为帧索引, k 为频率索引, $|S(\lambda, k)|^2$ 为带噪声声音功率谱, 取平滑常量 $\eta = 0.7$.

能性就越大, 反之亦然. 用式(5)对前景声音是否存在进行判断.

$$I(\lambda, k) = \begin{cases} 1, & \text{前景声音存在, } R(\lambda, k) > \delta(k) \\ 0, & \text{只含背景噪声, 其他} \end{cases} \quad (5)$$

其中, $\delta(k)$ 为频率相关常量, 用式(6)表示.

$$\delta(k) = \begin{cases} 1.5, & 1\text{Hz} \leq k \leq LF \\ 2, & LF < k \leq MF \\ 5, & MF < k \leq 0.5 F_s \end{cases} \quad (6)$$

本文中鸟类声音频率范围大致为 1kHz ~ 3.5kHz, 结合所采用的马路附近, 雨天湖边和山间流水噪声的功率谱分析, 得出这 3 种环境下噪声功率谱大部分集中在低频区域, 故取 $LF = 1\text{kHz}$, $MF = 3.5\text{kHz}$. 采样频率 $F_s = 8\text{kHz}$. 根据式(4)和式(5)的计算, 前景声音存在的概率 $P(\lambda, k)$ 可以通过式(7)进行更新.

$$p(\lambda, k) = \alpha_p p(\lambda - 1, k) + (1 - \alpha_p) I(\lambda, k) \quad (7)$$

其中, α_p 为平滑常量, 取 $\alpha_p = 0.2$.

2.4 时-频相关平滑常量的计算

结合前景声音存在概率 $p(\lambda, k)$, 时-频相关平滑因子定义为式(8).

$$\alpha_s(\lambda, k) = \alpha_d + (1 - \alpha_d) p(\lambda, k) \quad (8)$$

其中, α_d 为常量, 取 $\alpha_d = 0.85$. $\alpha_s(\lambda, k)$ 的取值范围为 $\alpha_d \leq \alpha_s(\lambda, k) \leq 1$.

2.5 噪声功率谱估计的更新

最后,通过式(8)时-频相关平滑因子 $\alpha_s(\lambda, k)$ 的计算,噪声功率谱估计的更新可以用式(9)表示.

$$N(\lambda, k) = \alpha_s(\lambda, k) N(\lambda - 1, k) + (1 - \alpha_s(\lambda, k)) |S(\lambda, k)|^2 \quad (9)$$

其中, $N(\lambda, k)$ 为估计的噪声功率谱.

3 多频带谱减法

真实环境中的噪声不具有平谱特性且对声音的整个频谱影响不均匀,因此,针对非平稳噪声,一般的谱减法在对声音进行降噪的同时也引入了所谓的“音乐噪声”.为了减少残差噪声及声音失真,本文使用多频带谱减法^[8]进行降噪.具体流程见图1的多频带谱减部分.

3.1 平滑带噪声幅度谱

本文实验表明,平滑估计的噪声功率谱对减少残差噪声是无效的.然而,对带噪声幅度谱进行平滑可以减小前景声音幅度谱的方差并且提高其质量,平滑的幅度谱可以用式(10)表示.

$$\bar{S}(\lambda, k) = \sum_{l=-M}^M W_l S(\lambda - l, k) \quad (10)$$

其中, $\bar{S}(\lambda, k)$ 为平滑的带噪声幅度谱, M 为相邻的帧数,取 $M=2$ 能避免产生频谱泄漏与谱间干扰.权重 W_l 设置为 $W_l = [0.09, 0.25, 0.32, 0.25, 0.09]$, $-2 \leq l \leq 2$.

3.2 多频带谱减降噪

式(11)为带噪声功率谱的近似取值.

$$|S(\lambda, k)|^2 \approx |C(\lambda, k)|^2 + |N(\lambda, k)|^2 \quad (11)$$

其中, $|C(\lambda, k)|^2$ 为前景声音功率谱, $|N(\lambda, k)|^2$ 为估计的噪声功率谱.

由于非平稳的有色噪声对于功率谱的各个频率范围的影响是不同的,因此,声音功率谱被划分为 N 个互不重叠的均匀线性子频带, $1 \leq N \leq 8$,根据鸟类声音频率分布,取 $N=6$.第 i 个子频带的估计前景声音功率谱表示为式(12).

$$|\hat{C}_i(\lambda, k')|^2 = |\bar{S}_i(\lambda, k')|^2 - \alpha_i(\lambda) \delta_i(\lambda) |N_i(\lambda, k')|^2, \quad b_i \leq k' \leq e_i \quad (12)$$

其中, $|\bar{S}_i(\lambda, k')|^2$ 为平滑和平均后的 $\bar{S}(\lambda, k)$ 再根据多频带划分后得出的第 i 个子频带的带噪声功率谱, $|N_i(\lambda, k')|^2$ 为第 i 个子频带的估计噪声功率谱, b_i 和 e_i 分别为第 i 个子频带频率窗口的开始值和结束值.

$\alpha_i(\lambda)$ 为第 λ 帧的第 i 个子频带的过减因子,用于控制第 λ 帧第 i 个子频带的减噪程度,可以通过式(13)计算第 λ 帧第 i 个子频带的功率谱信噪比 $\text{SNR}_i(\lambda)$ 后得出.

$$\text{SNR}_i(\lambda) = 10 \log_{10} \left(\frac{\sum_{k'=b_i}^{e_i} |S_i(\lambda, k')|^2}{\sum_{k'=b_i}^{e_i} |N_i(\lambda, k')|^2} \right) \quad (13)$$

其中, $|S_i(\lambda, k')|^2$ 为非平滑的第 i 个子频带的带噪声功率谱,当子频带信噪比高时,噪声影响比较小,所以设定较低的谱减权重,而子频带的信噪比低时,设定较高的谱减权重,设定的 $\alpha_i(\lambda)$ 可以用式(14)表示.

$$\alpha_i(\lambda) = \begin{cases} 4.75, & \text{SNR}_i(\lambda) < -5 \\ 4 - 1.25 \text{SNR}_i(\lambda), & -5 \leq \text{SNR}_i(\lambda) \leq 20 \\ 1, & \text{SNR}_i(\lambda) > 20 \end{cases} \quad (14)$$

$\delta_i(\lambda)$ 为第 λ 帧第 i 个子频带的子带附加控制相减因子,用式(15)表示.

$$\delta_i(\lambda) = \begin{cases} 2, & f_i(\lambda) \leq 1 \text{ kHz} \\ 1, & 1 \text{ kHz} < f_i(\lambda) \leq 0.5 F_s - 0.5 \text{ kHz} \\ 1.5, & f_i(\lambda) > 0.5 F_s - 0.5 \text{ kHz} \end{cases} \quad (15)$$

其中, $f_i(\lambda)$ 为第 λ 帧第 i 个子频带的频率上限.根据鸟类声音频率范围 $1 \text{ kHz} \sim 3.5 \text{ kHz}$,分别对 $0 \text{ Hz} \sim 1 \text{ kHz}$, $1 \text{ kHz} \sim 3.5 \text{ kHz}$ 和 $3.5 \text{ kHz} \sim 4 \text{ kHz}$ 这3个频带设定 $\delta_i(\lambda)$.

在式(12)中,当估计的前景声音功率谱出现负值时,采用式(16)进行更新.

$$|\hat{C}_i(\lambda, k')|^2 = \begin{cases} |\hat{C}_i(\lambda, k')|^2, & |\hat{C}_i(\lambda, k')|^2 > 0 \\ \varphi |\bar{S}_i(\lambda, k')|^2, & \text{其他} \end{cases} \quad (16)$$

其中,功率谱下限参数 $\varphi = 0.002$.

为了掩盖去噪后可能残留的音乐噪声并避免出现声音的过度失真,用式(17)往估计的前景声音功率谱上添加5%的原始带噪声功率谱.

$$|\hat{C}_i(\lambda, k')|^2 = |\hat{C}_i(\lambda, k')|^2 + 0.05 |S_i(\lambda, k')|^2 \quad (17)$$

因此,整个频带的估计的前景声音功率谱表示为式(18).

$$P_C(\lambda, k) = |\hat{C}_i(\lambda, k')|^2, 1 \leq i \leq 6 \quad (18)$$

图2表明,10dB山间流水情况下,降噪后公画眉声谱图相比纯净公画眉声谱图,虽然前景信号有所衰减,但是较为彻底地移除了背景噪声部分.因此,结合非平稳噪声估计的多频带谱减法可以有效地进行降噪.

4 抗噪 APNCC 特征提取

相比 MFCC 与感知线性预测系数(Perceptual Linear Prediction, PLP), PNCC 具有较好的抗噪效果^[6].根据得到的估计的前景声音功率谱(以下简称声音功率谱) $P_C(\lambda, k)$,进行 APNCC 特征提取.具体流程见图1的 APNCC 提取部分.

4.1 Gammatone 滤波

Gammatone(简称GT)滤波器是基于人耳耳蜗听觉

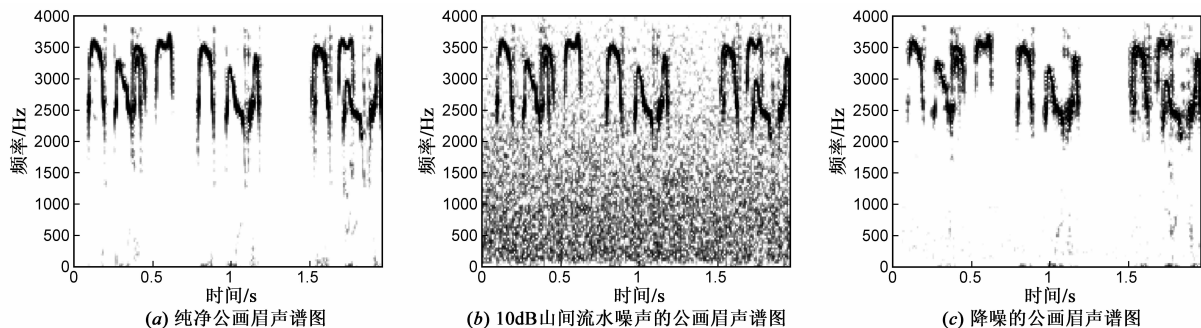


图2 结合非平稳噪声估计的多频带谱减法的降噪情况

特性的滤波器模型^[9],具有尖锐的频率选择特性,与 Mel 三角滤波器相比,其更加集中了临界频带中心频率附近的声音特征,且中心频率的两边过渡圆滑并能降低滤波器组相邻频带之间频谱能量的泄漏。

根据采样频率,经对比实验后取 32 阶 GT 滤波器进行滤波,且设定 GT 滤波器组的各个等价矩形带宽的中心频率在 100Hz ~ 4kHz 均匀分布.使用 Auditory toolbox^[10]工具箱进行 GT 滤波,滤波后的声音能量(简称 GT 能量)用式(19)表示。

$$X(\lambda, g), 1 \leq g \leq 32 \quad (19)$$

其中, g 为 GT 滤波后的等价矩形带宽(简称 GT 频带)的中心频率索引。

4.2 中间能量偏差移除

实验分析表明,背景噪声的 GT 能量与前景声音的 GT 能量相比,其变化速度通常较缓慢.为了移除背景噪声的 GT 能量引起的能量偏差,通过式(20)估计出每帧的中值 GT 能量。

$$Q(\lambda, g) = \frac{1}{2M+1} \sum_{\lambda'=\lambda-M}^{\lambda+M} X(\lambda', g) \quad (20)$$

其中,取 $M=3$,相当于取 7 个连续帧进行中值 GT 能量计算。

对于每个 GT 频带,通过计算其所有帧的算术平均值与几何平均值的比值(Arithmetic Mean to Geometric Mean Ratio, AM/GM),可以估计前景声音的明显程度.为了降低计算量,用式(21)表示 AM/GM。

$$G(g) = \ln \left[\sum_{\lambda=0}^{J-1} \max(Q(\lambda, g), \epsilon) \right] - \frac{1}{J} \sum_{\lambda=0}^{J-1} \ln[\max(Q(\lambda, g), \epsilon)] \quad (21)$$

其中, J 为声音段的总分帧数.为避免比值出现负无穷的情况,添加极小值 $\epsilon = 2.2204 \times 10^{-16}$ 。

经过实验对比分析得出,随着声音信噪比的下降,AM/GM 跟着下降.因此,得出中间能量偏差移除的思想:通过对中间能量偏差量的不断移除,使带噪声声音的 AM/GM 不断上升,直到逼近纯净声音的 AM/GM 为止。

进行中间能量偏差移除,首先要得到每个 GT 频带

中间能量偏差估计值 $B^*(g)$ 的逼近值 $B(g)$.对 $B(g)$ 的取值进行动态估计,从相对于第 g 个 GT 频带的平均能量的信噪比为 -50dB 开始进行 $B(g)$ 的取值,然后,该信噪比不断增加 1dB 进行 $B(g)$ 值的更新直至最后中间能量偏差完全移除。

由中间能量偏差估计值的逼近值 $B(g)$ 可以计算出移除中间能量偏差的中值 GT 能量,用式(22)表示。

$$\tilde{Q}(\lambda, g | B(g)) = \max(Q(\lambda, g) - B(g), d_0 Q(\lambda, g)) \quad (22)$$

其中, d_0 为用于避免 $\tilde{Q}(\lambda, g | B(g))$ 出现负值情况的常量,取 $d_0 = 10^{-3}$ 。

根据式(21)和式(22),可以计算出移除中间能量偏差后的 AM/GM,用式(23)表示。

$$\begin{aligned} \tilde{G}(g | B(g)) = & \log_{10} \left[\sum_{\lambda=0}^{J-1} \max(\tilde{Q}(\lambda, g | B(g)), c_f(g)) \right] \\ & - \frac{1}{J} \sum_{\lambda=0}^{J-1} \log_{10} [\max(\tilde{Q}(\lambda, g | B(g)), \\ & c_f(g))] \end{aligned} \quad (23)$$

其中,式(24)为下限参数 $c_f(g)$ 的定义。

$$c_f(g) = d_1 \left(\frac{1}{J} \sum_{\lambda=0}^{J-1} Q(\lambda, g) \right) \quad (24)$$

实验表明,引入 $c_f(g)$ 可以获得更好的中间能量偏差移除效果.取 $d_1 = 10^{-3}$,表示 $c_f(g)$ 相对于第 g 个 GT 频带的平均能量的信噪比为 -30dB。

根据中间能量偏差估计值的逼近值 $B(g)$ 从小到大依次代入式(25)进行计算,可以得出中间能量偏差估计值。

$$B^*(g) = \min\{B(g) | \tilde{G}(g | B(g)) \geq G_d(g)\} \quad (25)$$

其中, $G_d(g)$ 为纯净前景声音的 AM/GM。

计算出 $B^*(g)$ 后,根据式(20)和式(22),GT 频带的能量归一化增益可用式(26)表示。

$$w(\lambda, g) = \frac{\tilde{Q}(\lambda, g | B^*(g))}{Q(\lambda, g)} \quad (26)$$

用式(27)对 GT 能量进行平滑可得移除中间能量偏差后声音的 GT 能量。

$$\tilde{X}(\lambda, g) = (\frac{1}{2N+1} \sum_{g'=\max(g-N,1)}^{\min(g+N,C)} w(\lambda, g')) X(\lambda, g) \quad (27)$$

其中, C 为 GT 频带数, 取 $C = 32, N = 5$.

4.3 提取 APNCC

与 MFCC 提取中使用对数函数不同, 使用式(28)对 $\tilde{X}(\lambda, g)$ 进行非线性幂函数变换.

$$Y(\lambda, g) = (\tilde{X}(\lambda, g))^{0.1} \quad (28)$$

最后, 根据式(29)对 $Y(\lambda, g)$ 进行离散余弦变换 (Discrete Cosine Transform, DCT) 得到 APNCC.

$$\text{APNCC}(\lambda, i) = \sum_{g=1}^C Y(\lambda, g) \cos\{(g - \frac{1}{2}) \frac{i\pi}{C}\} \quad (29)$$

其中, $i = 1, 2, \dots, I, 1 \leq I \leq C$, 取 $I = 16$.

5 实验与结果分析

5.1 声音样本集

实验用到的 34 种纯净鸟叫声大部分来自 Freesound^[11]声音数据库, 分别为白面鸡、董鸡、公鹧鸪、母鹧鸪、黑水鸡、红毛鸡、山斑鸠、朱颈斑鸠、雉鸡、秧鸡、褐雨燕、绣眼、赭红尾鸲、鹭、麻雀、乌鸦、杜鹃、鸽子、鸭、鹊、猫头鹰、欧亚鸪、海鸥、天鹅、燕鸥、公画眉、母画眉、母画眉狂叫、母绣眼、公绣眼、母竹鸡、竹鸡破声、竹鸡仔、竹鸡求伴. 实验用到的 3 种背景噪声为使用 SONY ICD-UX512F 录音棒以 44.1kHz 的采样频率录取的真实环境声音, 分别为马路附近噪声、雨天湖边噪声和山间流水噪声. 以上声音文件的编码格式和长度没有统一规范, 经声谱图分析, 34 种鸟叫声可以根据声谱图频率低于 4kHz 的部分进行区分. 因此, 本文中所有的声音都转换为采样频率为 8kHz、采样位数为 16b, 时长为 2s 的单声道 WAV 格式声音段. 经处理后, 每种纯净鸟叫声都含有 36 个声音段, 总共 1224 个声音段. 然后对每个声音段按信噪比为 40dB、30dB、20dB、10dB、0dB 和 -10dB 分别添加 3 种背景噪声.

5.2 实验设计

对声音段先进行预处理, 首先对声音段归一化处理使用一阶 FIR 滤波器进行预加重, $H(z) = 1 - \mu z^{-1}$, 取 $\mu = 0.9375$. 然后进行分帧, 并加 Hamming 窗, 帧长 32ms (256 个采样点), 帧移 16ms (128 个采样点). 接着, 对每帧进行 FFT. 最后, 分别提取了 APNCC、PNCC 及 MFCC 这 3 种特征结合 LIBSVM^[12]工具箱进行实验效果对比. 本实验采用基于 RBF (Radius Basis Function) 核的 SVM. 进行 SVM 建模时, 对于所有的 2s 的训练声音段, 都取前 100 帧的特征进行惩罚因子 c 及核参数 g 的自动寻优, 结果 $c = 2, g = 2.8284$.

对于每种鸟叫声, 每次随机选取 6 个纯净的声音段作为训练样本, 其余 30 个纯净声音段及其相应的不同背景噪声下不同信噪比的声音段分别作为各自情况下

的测试样本. 进行 10 次随机选取并对这 10 次的识别率取平均值后, 得到最终识别率.

5.3 结果与分析

APNCC、PNCC 和 MFCC 对于 3 种环境下不同信噪比的识别率如表 1 所示. 由实验结果可得出, 对于这 3 种环境下不同信噪比的识别率均值, APNCC 相比 PNCC 和 MFCC 都有较大的提高, 这说明了 APNCC 更适用于噪声下的鸟类声音识别, 具有较好的噪声鲁棒性.

对于纯净的鸟类声音识别, 3 种特征的识别效果都很理想, 识别率都在 94% 以上, 其中, MFCC 获得最高的识别率, 这说明 MFCC 更适用于无噪声的鸟类声音识别. 在信噪比为 40dB 的 3 种微小噪声下, 使用这 3 种特征的识别效果都略有下降, 但也都在 90% 以上, 而 PNCC 识别率最高. 这说明了在噪声极小的情况下, APNCC 提取过程中包含了 2 个降噪步骤: 结合非平稳噪声估计的多频带谱减去噪和中间能量偏差移除去噪. 因此, 在降噪过程中出现了由过分去噪带来的较小信号失真并对识别效果产生一定影响. 随着噪声的增大, 在信噪比低于 30dB 的 3 种噪声下, 使用 APNCC 的识别率都为最高, 而 MFCC 的识别率都为最低. 这说明了 APNCC 对于不同的非平稳环境噪声具有较好的鲁棒性, PNCC 本身具有一定的抗噪性, 而 MFCC 对噪声敏感且抗噪性差.

图 3 表明, 对于这 3 种噪声, 在信噪比逐渐降低到 10dB 的过程中, APNCC 的识别率下降速度最平缓, 说明了该信噪比范围内 APNCC 有良好的噪声鲁棒性, 而 MFCC 受噪声的影响最大, 识别率迅速下降. 在信噪比降低到 0dB 时, APNCC 的识别率虽然出现了大幅度下降且低于 50%, 但也显著高于 PNCC 与 MFCC. 在信噪比降低到 -10dB 时, 对于马路附近噪声, 由于其主要集中在低频区域, 对鸟类声音的频率影响较小, 故 APNCC 的识别率将近 30%, 而 APNCC 针对其他 2 种背景噪声的识别率则降低到 10% 左右.

表 1 这 3 种特征对于 3 种环境下不同信噪比的识别率 (%)

特征 \ 环境	APNCC			PNCC			MFCC		
	山间 流水	马路 附近	雨天 湖边	山间 流水	马路 附近	雨天 湖边	山间 流水	马路 附近	雨天 湖边
纯净	94.12	94.12	94.12	94.41	94.41	94.41	97.35	97.35	97.35
40dB	92.94	93.24	92.06	94.12	94.12	93.82	91.18	90.88	90.29
30dB	90.29	92.94	91.76	88.24	91.18	89.12	85.59	88.53	86.76
20dB	85.00	88.24	82.06	72.94	80.59	74.71	72.35	72.94	72.94
10dB	64.41	73.24	74.41	43.82	58.82	49.71	36.76	38.24	40.88
0dB	21.47	43.53	39.14	12.35	29.12	21.76	5.59	8.82	8.24
-10dB	8.82	29.41	12.06	4.41	8.82	4.12	2.06	3.53	4.12
均值	65.29	73.53	69.37	58.61	65.29	61.09	55.84	57.18	57.23

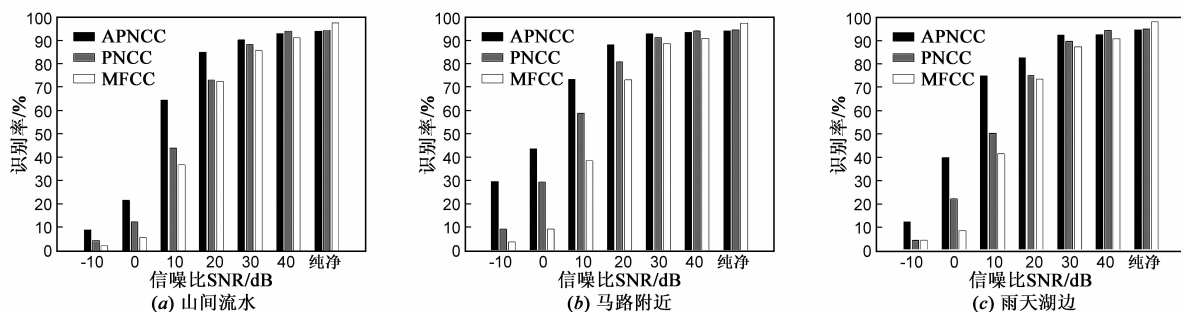


图3 这3种特征对于3种环境下不同信噪比的识别效果图

6 结论

针对非平稳的真实环境噪声下的鸟类声音识别问题,提出了一种新型的经两阶段去噪的 APNCC 特征与 SVM 分类器相结合的识别方法.实验表明,对于 3 种不同的非平稳环境噪声存在的情况,APNCC 的平均识别效果明显优于具有一定抗噪性的 PNCC 和噪声敏感的 MFCC,尤其在信噪比低于 30dB 的环境下,APNCC 的优势更加突出.因此,APNCC 具有较好的非平稳噪声鲁棒性且适用于真实环境噪声下的鸟类声音识别.

参考文献

- [1] Somervuo P, Hama A. Bird song recognition based on syllable pair histograms[A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Montreal, Canada: IEEE Press, 2004: 825 – 828.
- [2] Cheng J, Sun Y, Ji L. A call-independent and automatic acoustic system for the individual recognition of animals: a novel model using four passerines[J]. Pattern Recognition, 2010, 43 (11): 3846 – 3852.
- [3] 冯霞, 龚晓峰, 张利丹, 武瑞娟. 基于纹理特征的背景噪声提取的应用研究[J]. 电子学报, 2009, 37(9): 2092 – 2095.
Feng Xia, Gong Xiao-feng, Zhang Li-dan, Wu Rui-juan. Research of background noise extraction based on texture feature [J]. Acta Electronica Sinica, 2009, 37(9): 2092 – 2095. (in Chinese)
- [4] Chu W, et al. Noise robust bird song detection using syllable pattern-based hidden markov models [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Prague, Czech Republic: IEEE Press, 2011: 345 – 348.
- [5] Bardeli R, Wolff D, Kurth F, et al. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring[J]. Pattern Recognition Letters, 2010, 31(12): 1524 – 1534.
- [6] Kim C, Stern R. Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Dallas, TX: IEEE Press, 2010. 4574 – 4577.
- [7] Rangachari S, Loizou P C. A noise estimation algorithm for highly non-stationary environments [J]. Speech Communication, 2006, 48(2): 220 – 231.
- [8] Kamath S, et al. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Orlando, FL: IEEE Press, 2002. IV-4164-IV-4164.
- [9] 王 ■, 钱志鸿, 王雪, 程光明. 基于伽马通滤波器组的听觉特征提取算法研究[J]. 电子学报, 2010, 38(3): 525 – 528.
Wang Yue, Qian Zhi-hong, Wang Xue, Cheng Guang-ming. An auditory feature extraction algorithm based on γ -tone filter-banks [J]. Acta Electronica Sinica, 2010, 38(3): 525 – 528. (in Chinese)
- [10] Slaney M. Auditory toolbox version 2 [CP/OL]. <https://engineering.purdue.edu/~malcolm/interval/1998-010/AuditoryToolbox.zip>, 2012-5-14.
- [11] Universitat Pompeu Fabra. Repository of sound under the creative commons license, Freesound. org [DB/OL]. <http://www.freesound.org>, 2012-5-14.
- [12] Chang C C, Lin C J. Libsvm version 3.12 [CP/OL]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/libsvm-3.12.zip>, 2012-5-14.

作者简介



颜鑫男, 1988 年出生于福建省泉州市. 现为福州大学数学与计算机科学学院硕士研究生, 主要研究方向为声音识别.
E-mail: yanxin124@126.com



李应男, 1964 年出生于福建省闽清县. 现为福州大学数学与计算机科学学院教授, 主要研究方向为环境声音识别、信息安全. E-mail: fj_liying@fzu.edu.cn