用多频带能量分布检测低信噪比声音事件

李 应* 吴灵菲

(福州大学数学与计算机科学学院 福州 350116)

(网络系统信息安全福建省高校重点实验室 福州 350116)

摘 要：该文针对低信噪比噪声环境下的声音事件检测问题，提出基于多频带能量分布图离散余弦变换的声音事件检测的方法。首先，将声音数据转化为gammatone频谱，并计算其多频带能量分布；接着，对多频带能量分布图进行8×8分块与离散余弦变换；然后，对8×8的离散余弦变换系数进行Zigzag扫描，抽取离散余弦变换系数的主要系数作为声音事件的特征；最后，利用随机森林分类器对特征建模与检测。实验结果表明，在低信噪比及各种噪声环境下，该文提出的方法具有良好的检测效果。

关键词：声音事件检测；多频带能量分布；随机森林；离散余弦变换

中图分类号：TP391.42 **文献标识码**：A **文章编号**：1009-5896(2018)12-2905-08

DOI: 10.11999/JEIT180180

Detection of Sound Event under Low SNR Using Multi-band Power Distribution

LI Ying WU Lingfei

(*College of Mathematics and Computer Science*, *Fuzhou University*, *Fuzhou* 350116, *China*)

(*Fujian Province Key Laboratory of Information Security of Network Systems*, *Fuzhou University*, *Fuzhou* 350116, *China*)

Abstract: As to the problem of sound event detection in low Signal-Noise-Ratio (SNR) noise environments, a method is proposed based on discrete cosine transform coefficients extracted from multi-band power distribution image. First, by using gammatone spectrogram analysis, sound signal is transformed into multi-band power distribution image. Next, 8×8 size blocking and discrete cosine transform are applied to analyze the multi-band power distribution image. Based on the main Zigzag coefficients which are scanned from the discrete cosine transform coefficients, features of sound event are constructed. Finally, features are modeled and detected through random forests classifier. The results show that the proposed method achieves a better detection performance in low SNR comparing to other methods.

Key words: Sound event detection; Multi-band power distribution; Random forests; Discrete cosine transform

1 引言

低信噪比声音事件检测，就是试图检测、分类和识别嵌入在各种噪声和混响音频信号中的相对微弱的声音对象。它对于音频取证、环境声音识别、生物声音监控、声场景分析，实时军事关注点的检测、定位跟踪和声源分类，病人监护，工业系统非正常事件监测及故障诊断，递交早期维护的关键信息等都具有重要意义。

关于声音事件检测，目前的研究包括：特定声音事件的声音信号增强方法[1]，基于声音场景分类的噪声抑制方法[2]；吵闹环境下特定的声音事件检测方法[3]，声音事件的特征[4−6]及声音事件的分类器[7,8]；特征及分类器的组合[9,10]；特定环境下特定声音的检测方法[11]等。这些方法、特征及其与分类器组合等，从不同侧面对各种声音事件的检测与分类进行了深入的研究。它们对声音事件取得了良好的检测率，然而对于低信噪比声音事件却没有明显的效果。

Feng等人[3]通过小波包滤波器选择地过滤场景声音，可以检测到特定声场景下–10 dB的特定声音

收稿日期：2018-02-09；改回日期：2018-07-09；网络出版：2018-07-26

*通信作者： 李应 fj_liying@fzu.edu.cn

基金项目：国家自然科学基金(61075022)，福建省自然科学基金(2018J01793)

Foundation Items: The National Natural Science Foundation of China (61075022), The Natural Science Foundation of Fujian Province (2018J01793)

事件。对于各种声音事件，文献[12]提出的基于子带能量分布(SPD)图及相关方法，可以在0 dB时获得接近90%的检测率。而对于更低信噪比的声音事件，这种方法的检测率却受到了限制。

针对低信噪比，如–5 dB和–10 dB各种声音事件的检测，本文提出基于多频带能量分布(Multi-Band Power Distribution, MBPD)的低信噪比声音事件检测方法。这种方法通过对MBPD图的分块离散余弦变换(DCT)，把DCT系数的Z编码的主要部分作为声音事件的特征，即MBPD-DCTZ，并用随机森林(RF)分类器对MBPD-DCTZ进行训练与检测。

2 相关方法

2.1 现有方法

对于低信噪比声音事件，Dennis等人[12,13]把SPD图特征及谱图特征用于非匹配条件下声音事件分类。这两种方法分类检测声音事件的过程如图1所示。

文献[13]的过程如图1的下半部分的细线框所示，即灰度对数频谱图、图像特征抽取，SVM分类。采用这种方法，在0 dB的情况下，检测率达到74.4%[13]。在文献[13]中，如图1中的虚线框所示，对图特征抽取是通过Jet映射，把灰度对数谱图，映射成3张子图，对每张子图进行9×9分块，再提取每一块均值与方差，即共486(2×3×9×9)维向量作为特征进行SVM的训练与分类。

以文献[13]的图像特征抽取方法为基础，如图1的上半部分的粗线框所示，文献[12]在频谱、频谱分析及分类器的选择上进行了进一步的处理。其中，频谱及分析包括：灰度gammatone谱图、子带能量分布(SPD)、对比增强形成增强的子带能量分布图。对图像特征的进一步处理包括：帧缺失掩饰估计，去除不可靠维度。然后再用基于Hellinger距离的k近邻分类器(kNN)分类。采用这种方法，在信噪比为0 dB情况下，对声音事件的检测率可以达到88%[12]。在文献[12]中，也是通过Jet把子带能量分布图映射成3张子图，对每张子图进行10×10分块，再提取均值与方差，即共600(2×3×10×10)维向量作为特征进行kNN的建模与分类。

文献[12]实现低信噪比声音事件分类的主要措施，是采用帧缺失掩饰估计与去除不可靠维度。使得声音事件保留部分的SPD只与对应的声音事件相关。因此改善了在0 dB情况下，对声音事件的检测率。

2.2 SPD与低信噪比的问题

对SPD进一步分析发现，对于更低信噪比的声音事件，如–5 dB或–10 dB采用文献[12]的方法，可能存在问题。其中包括：(1)通过统计信号中50个频率子带内的共100个不同等级能量的概率密度[12]，使得声音事件在低信噪比的情况下，高能的背景噪声将引起高能分布增值，SPD中原有的能量等级分布下移并因此减少可靠SPD成分；(2)声音事件在低信噪比的情况下，高能的背影噪声可能影响到更多的子频带，使得可能得到的SPD的可靠部分减少；(3)更低信噪比的情况下，噪声与声音事件的分界更加模糊，估计背景噪声的SPD误差增大。这些问题进而使得文献[12]的方法对低信噪比声音事件分类性能受到严重影响。

2.3 多频带能量分布的声音事件分类

针对文献[12]对低信噪比声音事件检测存在的问题，本文提出基于多频带能量分布(MBPD)的低信噪比声音事件检测方法。这种方法包括：(1)划分更细的频带，把频带数量从文献[12]的50增加到256，使得高能噪声对频带的影响细化，因此降低被影响频带的比例；(2)压缩能量等级，把能量等级从文献[12]的100减少到64，减少高能噪声引起的能量分布下移；(3) 8×8分块DCT，对MBPD图进行8×8分块DCT与Zigzag编码，捕捉MBPD图的细微变化；(4) RF分类器，采用随机森林(RF)分类器投票确定最后结果。

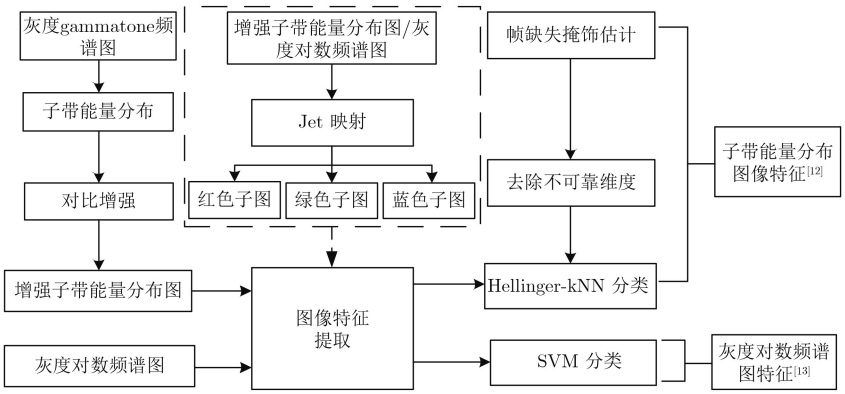


图 1 谱图特征用于非匹配条件的声音事件分类

针对更低信噪比声音事件检测，本文把文献[12]如图1所示的过程改进并简化成如图2所示的过程。其中包括灰度gammatone谱图、MBPD图、MBPD图8×8分块DCT、Zigzag编码、RF分类器检测声音事件。

3　多频带能量分布特征与检测

3.1　多频带能量分布

MBPD通过统计信号中256个频带内的64个能量等级的概率密度，将声音数据转化为MBPD图。形成MBPD图的步骤如下：

(1) Gammatone频谱图：声音信号$y(t)$通过gammatone滤波器组[14]滤波，得到$y_f[t]$。对$y_f[t]$取对数进行动态压缩，形成相应的gammatone谱图$S_g(f, t)$。图3(a)所示的是茶隼叫声对应的gammatone谱图。

$$S_g(f,t)=\lg|y_f[t]| \tag{1}$$

其中，f表示滤波器的中心频率，t表示的帧索引。

(2) 归一化能量谱：对每个声音信号的能量谱$S_g^2(f,t)$进行归一化处理，得到归一化后的能量谱$G(f, t)$。

$$G(f,t)=\frac{S_g^2(f,t)}{\max\limits_{f,t}(S_g^2(f,t))} \tag{2}$$

(3) 多频带能量分布：对$G(f, t)$的能量分布情况进行统计，得到$M(f, b)$，即图3(b)所示的MBPD图。设能量等级数目为B，采用基于统计的非参数法，对每个频率子带f的能量元素进行概率密度统计，得到特定频带在整个采样时间W上的能量分布情况。

$$M(f,b)=\frac{1}{W}\sum_{t}^{W}I_b(G(f,t)) \tag{3}$$

$$I_b(G(f,t))=\begin{cases}1, & \dfrac{b-1}{B}<G(f,t)<\dfrac{b}{B}\\ 0, & \text{其他}\end{cases} \tag{4}$$

其中，W为声音片段的长度，$M(f, b)$表示在频带f中能量等级为b的元素占该频带元素总数的比例($0\le M(f, b)\le 1$)；$I_b(G(f, t))$为指示函数，当$G(f, t)$属于能量等级b时，其值为1，否则为0。

3.2　离散余弦变换与多频带能量分布图

对一幅图像进行离散余弦变换(DCT)，可以将图像的重要可视信息都集中到DCT的少部分系数中[15]。一般情况下，DCT系数矩阵中，沿左上至右下的方向，DCT系数大小是依次递减的。左上角的第1个系数，被称为DCT的直流系数，是图像像素的均值。其它系数被称为交流(AC)系数。越靠近左上角，AC系数包含着越多的图像信息。

利用图像DCT的这些特性，对MBPD图进行分块，然后对子块进行DCT。

受图像处理中8×8的子块编码具有较高的效率的启发，本文对64×256大小的MBPD图进行8×8分块，即把图4(a)分为256个8×8子块。其中，每个子块携带着声音数据在相应频带及能量等级的分布情况，如图4(b)对应的4(a)中黑框子块，是MBPD图中频带从96至103、能量等级从25至32的能量分布。对子块进行DCT后，可以得到如图4(c)所示的8×8的DCT系数。

3.3　Zigzag扫描

为了有效地将DCT系数按重要程度排序，本文采用Zigzag行程扫描[16]，其路径如图4(c)的折线及箭头所示。8×8的DCT系数经过Zigzag扫描，可以得到如图4(d)的64个DCT系数的1维排列。在提取特征参数时，如图4(e)所示，只取1维排列的前部分数据即可表征图像的主要特征。通过综合实验分析，本文中，只取64个DCT系数的1维Zigzag排列的前5个系数作为该8×8图像块的特征值。这个特征值，即为MBPD子块的DCT系数经Zigzag扫描的特征，简称为MBPD-DCTZ。

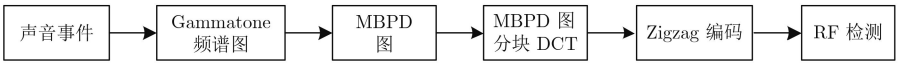
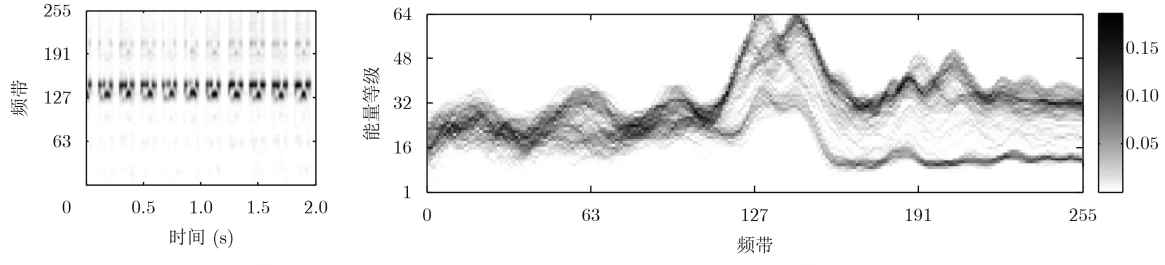
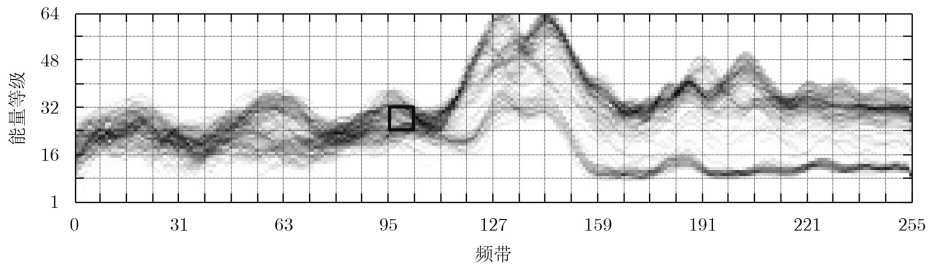


图 2 基于MBPD图的低信噪比声音事件检测

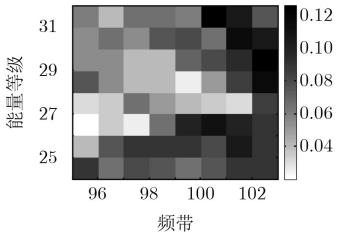


(a) Gammatone频谱图　　(b) MBPD频谱图

图 3 茶隼叫声的gammatone频谱图及MBPD



(a) MBPD 图的 8×8 分块



(b) 黑框子块

0.5631	0.0073	0.0727	0.0177	-0.0379	0.0220	-0.0028	0.0147
-0.1228	0.0223	0.0092	0.0230	0.0459	0.0100	-0.0348	-0.0185
0.0328	-0.0275	-0.0308	0.0619	0.0291	-0.0089	-0.0018	0.0163
-0.0105	-0.0163	0.0246	-0.0167	-0.0098	-0.0113	0.0451	0.0253
7.9781e-18	0.0058	-0.0249	0.0157	-5.5199e-18	0.0198	-6.3477e-04	0.0036
0.0078	-0.0062	0.0268	2.2202e-04	0.0201	-0.0067	-0.0086	0.0130
6.8678e-05	-0.0030	-0.0018	-0.0082	0.0372	-0.0030	0.0156	-0.0192
-9.3247e-04	0.0057	0.0036	0.0086	-0.0083	0.0064	-0.0092	-0.0064

(c) 黑框子块的 DCT 系数及 Zigzag 扫描

0.5631	0.0073	-0.1228	…	-0.0192	-0.0092	-0.0064

(d) 64个DCT 系数的1维 Zigzag 排列

0.5631	0.0073	-0.1228	0.0328	0.0223

(e) 前 5 个 Zigzag 系数

图 4 图像分块及DCT系数

3.4 随机森林分类器

随机森林(RF)分类器[17]是一种利用多个决策树分类器来对数据进行判别的集成分类器算法。随机森林检测声音事件的步骤为：(1)将待测声音数据的MBPD-DCTZ特征置于随机森林中所有n棵决策树的根节点处。(2)按照决策树的分类规则，由根节点依次向下传递直到到达某一叶节点；该叶节点对应类标签便是这棵决策树对MBPD-DCTZ特征所属类别所做的投票。(3)随机森林的n决策树均对每一个待测声音信号的MBPD-DCTZ特征的类别进行了投票，统计森林中n棵决策树投票；其中票数最多的类标签便是最终待测样本对应的类标。

4 试验

4.1 实验数据

实验数据使用两个数据集，动物声音事件集和办公室声音事件集。其中，50种动物声音事件来自Freesound[18]声音数据库，包括不同鸟鸣声和哺乳动物叫声；每种声音事件有30个样本。11种办公室中常见的声音事件，来自DCASE2016 Task2[19]；每一类声音事件共有20个样本。实验中，以50种动物声音事件为主，11种办公室声音事件作为进一步的辅助验证。实验用到的6种噪声环境可分为两类，即平稳噪声和非平稳噪声。平稳噪声为粉噪声(pink)，非平稳噪声包括模拟真实场景声音的流水声、风声、公路声、海浪声和雨声。噪声样本与声音事件的格式为单声道“.wav”格式，采样频率为44.1 kHz。

4.2 实验设计

实验中，gammatone滤波器参数为：帧长为25 ms，帧移为10 ms，滤波器组数目为256；取随机森林分类器中决策树的个数k=500，决策树中非叶节点分裂时预选特征成分的数量m=5。为了验证本文方法的检测性能，共进行5个实验。其中包括：(1)MBPD-DCTZ特征的参数设置；(2)MBPD-DCTZ特征与RF分类器相结合的性能检测；(3)分类器性能的比较；(4)MBPD-DCTZ特征与常用特征性能的比较；(5)MBPD-DCTZ-RF与现有方法的比较。

4.3 结果与分析

(1) MBPD特征提取的是DCT系数Zigzag排列的主要系数。以动物声音事件集为基础，分别选择MBPD-DCT的Zigzag排列的前面1～10个系数进行实验。在–10 dB，–5 dB和0 dB 3种信噪比的6种噪声环境下随机森林的平均检测结果如图5所示。

由图5可知，对DCT系数进行Zigzag排列提取Z个重要系数在一定程度上均能提高DCT系数对声音事件的表征性能。当Z=4和Z=5时，对–10 dB, –5 dB和0 dB检测率达到最佳。相对而言，Z=5时平均检测率，略高于Z=4时的平均检测率。因此我们在后面的实验中，取Z=5。

(2) 为说明MBPD-DCTZ与RF分类器结合的有效性，本文对动物声音事件集进行了交叉验证实验。在–10 dB, –5 dB, –0 dB和5 dB等4种不同信噪比，及流水、粉噪声、风声、海浪、公路和雨声等

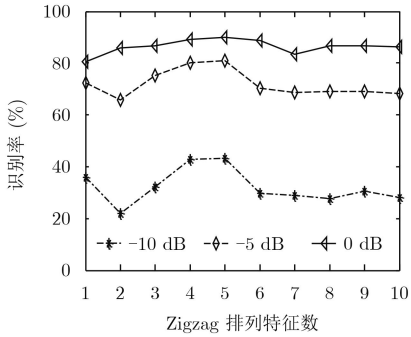


图 5 不同Z值的检测率

6种背景噪声条件下，3次交叉验证实验的平均检测结果如表1所示。由表1可知，不论在平稳噪声条件还是非平稳噪声条件下，MBPD-DCTZ特征都表现出了良好的性能，在–5 dB低信噪比时，达到平均87±3%的平均检测率。

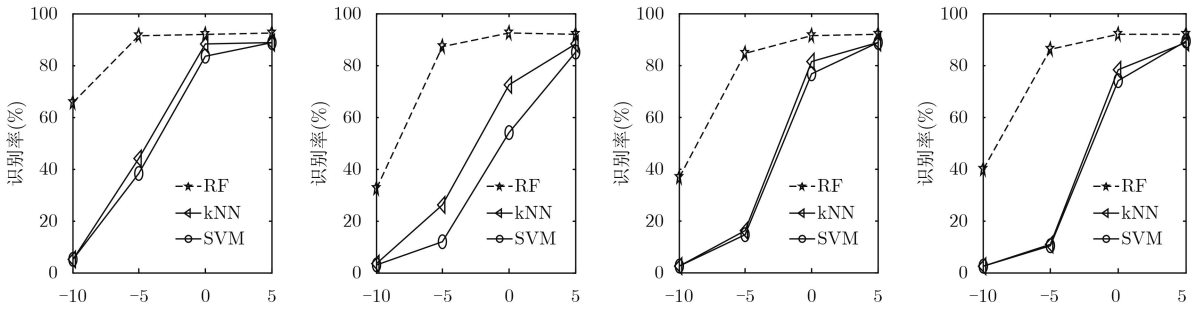
(3) 比较MBPD-DCTZ特征分别在随机森林(RF)、支持向量机(SVM)[20]和k近邻(kNN)[21]中的检测效果。以动物声音事件集为例，将不同信噪比的不同环境下的MBPD-DCTZ特征，分别送入RF, SVM, kNN分类器进行检测。4种信噪比在粉噪声、风声、雨声和流水噪声条件下的比较结果如图6所示。由图6可知，在不同信噪比及不同背景噪声条件下，本文提出的MBPD-DCTZ特征比较适合于用RF分类器进行分类检测。因此，本文在对MBPD-DCTZ特征的检测过程中选用RF分类器。

(4) 为了进一步说明MBPD-DCTZ特征表征低信噪比声音事件的性能，本文采用RF分类器进行MBPD-DCTZ特征与MFCC[22], PNCC[23], GLCM-SDH[24], LBP[25], HOG[26]等几种特征的比较。其中，MFCC特征采用24个滤波器的三角滤波器组，提取12维DCT系数；PNCC特征采用32阶的gammatone滤波器，DCT时取12维系数。

在流水、粉噪声、风声、海浪、公路和雨声等六种背景噪声下，几种特征对动物声音事件的检测结果如表2所示。而对于办公室声音事件的检测结果如表3，采用MBPD-DCTZ特征，对原声及5 dB,

表 1　MBPD-DCTZ特征的交叉验证结果(%)

信噪比(dB)	噪声环境						
	流水	粉噪声	风声	海浪	公路	雨声	平均
–10	40.0±0.7	65.7±5.1	32.5±3.8	44.7±0.9	52.6±3.8	36.5±3.2	45.3±11.1
–5	86.1±3.4	91.1±1.7	87.0±3.2	82.9±1.9	91.2±2.1	84.7±2.5	87.2±3.1
0	91.7±1.9	91.8±1.9	92.3±1.9	91.6±1.4	92.01±2.2	91.5±1.9	91.8±0.3
5	91.9±1.9	92.2±1.9	92.1±2.3	92.2±1.8	92.3±2.1	92.0±1.9	92.1±0.1



(a) 粉噪声　(b) 风声　(c) 雨声　(d) 流水

图 6 MBPD-DCTZ特征在不同分类器下的检测率

表 2　6种噪声环境下不同特征对动物声音事件的平均检测率(%)

特征	信噪比(dB)			
	5	0	–5	–10
LBP	64.3±14.3	16.6±10.5	2.8±0.8	2.4±0.9
GLCM-SDH	41.4±3.5	36.0±4.3	14.6±9.5	4.2±1.7
HOG	68.9±5.4	28.8±10.5	7.4±5.2	4.1±1.8
MFCC	17.5±4.8	9.5±2.5	4.7±0.7	3.0±0.8
PNCC	28.0±0.9	20.0±0.9	9.1±2.0	2.5±0.8
MBPD-DCTZ	**92.1±0.1**	**91.8±0.3**	**87.2±3.1**	**45.3±11.1**

0 dB, –5 dB信噪比粉噪声的检测结果，也明显优于现有的特征。

(5) 不同信噪比及不同环境下，本文方法与MFCC-SVM[22]，SIF-SVM[13]，MP-SVM[10]和SPD-KNN[12]方法的比较。 由表4和表5可知，在低信噪比的情况下，本文所提的MBPD-DCTZ与RF结合的方法均能保持较好的检测性能，且受噪声环境影响较小。尤其在低信噪比时，本文所提的方法大幅度优于其它几种方法。

表 3 不同特征对办公室声音事件的检测率(%)

特征	办公室声音事件	粉噪声信噪比(dB)		
		5	0	–5
LBP	69.7±2.3	70.9±5.1	35.2±0.9	16.4±2.6
GLCM-SDH	47.3±5.4	44.2±7.5	45.5±5.4	38.8±4.8
HOG	70.3±5.2	40.6±4.8	33.9±3.1	32.1±2.3
MFCC	43.7±0.7	27.2±4.7	22.1±4.5	17.6±3.4
PNCC	47.2±1.9	34.3±2.0	28.1±2.3	22.1±1.8
MBPD-DCTZ	**75.2±0.6**	**75.2±1.7**	**75.8±4.3**	**54.6±5.4**

表 4 6种噪声环境下不同方法对动物声音事件的平均检测率(%)

方法	信噪比(dB)			
	5	0	–5	–10
本文方法	**92.1±0.1**	**91.8±0.3**	**87.2±3.1**	**45.3±11.1**
MFCC-SVM[22]	25.2±6.0	13.8±4.8	5.7±3.1	3.7±2.0
MP-SVM[10]	30.0±2.5	16.4±4.0	8.2±2.4	4.6±0.9
SIF-SVM[13]	61.4±8.5	40.3±12.1	18.9±13.4	9.7±7.7
SPD-KNN[12]	87.9±1.8	82.7±3.9	45.4±22.1	9.9±8.8

表 5 不同方法对办公室声音事件的检测率(%)

方法	办公室声音事件	粉噪声信噪比(dB)		
		5	0	–5
本文方法	75.2±0.9	**75.2±1.7**	**75.8±4.3**	**54.6±5.4**
MFCC-SVM[22]	16.4±1.8	15.8±1.7	17.6±0.9	16.4±3.0
MP-SVM[10]	62.7±4.2	45.4±2.1	26.0±0.9	14.0±1.4
SIF-SVM[13]	**75.2±2.3**	40.6±6.2	31.5±8.2	25.5±1.5
SPD-KNN[12]	36.4±13.6	28.5±4.8	25.5±5.4	21.8±5.4

5 讨论

从上述实验及结果可知，本文方法对于低信噪比声音事件，具有较好的检测能力。但从表4与表5中，我们同时也看到，当信噪比低至–10 dB时，对动物声音事件的检测率只有45%左右；当信噪比低至–5 dB时，对办公室声音事件的检测率只有55%左右。为此，我们对相关的机理进行讨论。

5.1 MBPD图与低信噪比

图7给出风声环境下–10 dB茶隼叫声、纯净茶隼叫声和风声的波形图，同时也分别给出与它们对应的gammatone频谱图和MBPD图。低信噪比时，声音事件的MBPD图存在4个方面的变化。

(1) 声音事件的MBPD能量等级下移与压缩。如图7(d)所示，黑框内的两个高能量等级峰值，在图7(b)中发生了下移与压缩。这是由于图7(b)混合了图7(d)与7(f)的能量而引起。在MBPD中，我们把能量等级分成了64级，如果环境噪声的能量等级高于声音事件的能量等级，如图7(b)、7(f)虚线框部分，最高能量等级被分配到第64级。这样，图7(d)中声音事件的最高等级的黑框部分，不再是最高等级，而被按比例压缩并下移至图7(b)所示黑框部分。

(2) 声音事件的MBPD被背景噪声的MBPD掩盖。当声音事件的信号变化与背景噪声同频带、同时刻出现时，MBPD图中，原先处在该频带上的能量等级及比例发生变化。变化大小，与声音事件和背景噪声在该频带处的能量有关。当背景噪声能量大时，如图7(d)黑线所处的纯净茶隼叫声的能量分布情况，受到图7(f)高能量等级风声在同一时刻对这一频带的影响，使得–10 dB的茶隼叫声的能量等级分布在该频带出现如图7(b)所示的被背景噪声的MBPD掩盖。当声音事件能量大时，可以归结为图7(b)、7(d)的黑框的情况。

(3) 声音事件的MBPD被加强。当声音事件的信号变化与背景噪声同频率、不同时刻出现时，MBPD图中该频带该能量等级颜色加深，即该频带部分能量等级的比例增加。这种情况如图7(f)、7(d)箭头所指的位置，声音事件与背景噪声在不同时刻都有第27频带的信号变化。因此图7(b)在该频带的位置出现颜色较深部分。

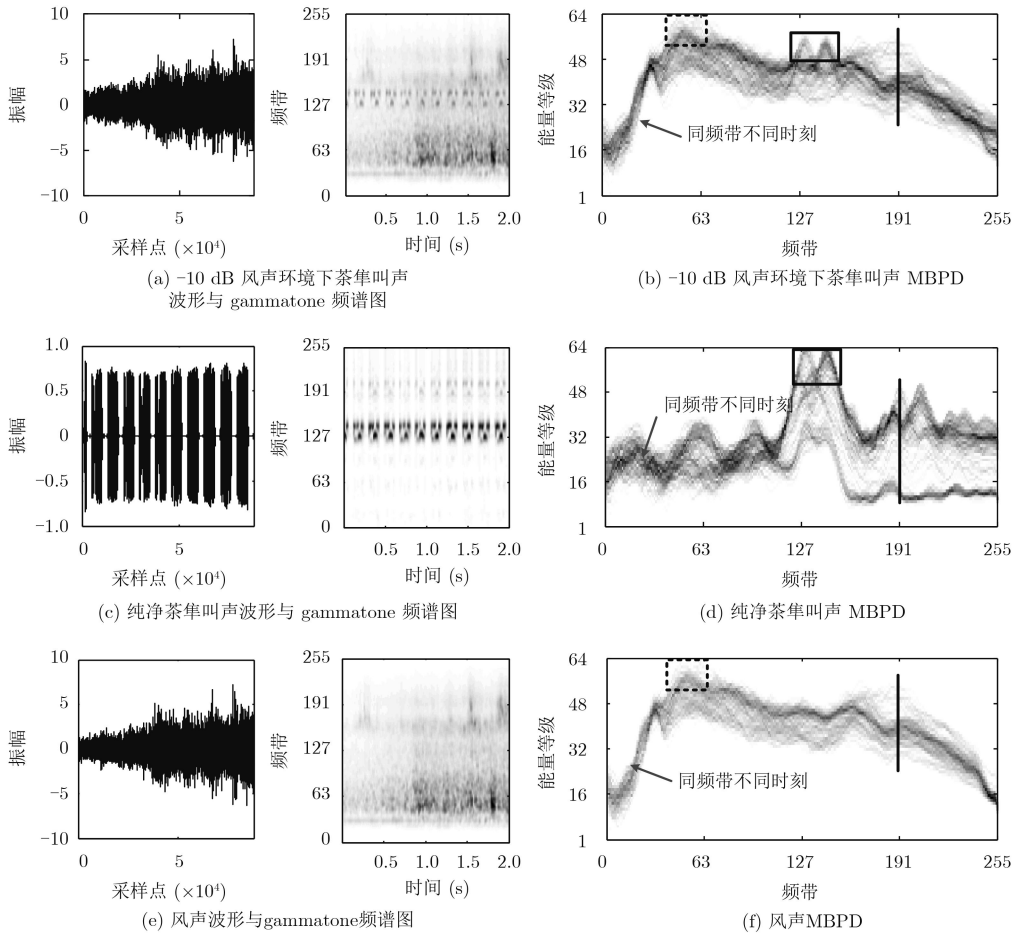
(4) 声音事件MBPD相对稳定成分。当声音事件的信号变化与背景噪声不同频带发生时，低信噪比声音事件MBPD成分的频带及能量等级分布将显示出与纯净声音事件的MBPD相同或相近之处。这种情况，也可以归结为如图7(b)、7(d)的黑框部分。

因此，图7(b)的黑框部分是低信噪比声音事件检测的关键依据。

5.2 提高检测率

如图7(b)黑框中的两个峰值所示，在音频数据中，声音事件只要有不同于背景噪声的频带及能量存在，在MBPD图中都可以体现出来。MBPD图分块及其DCT，就是通过适当地划分频带，细化噪声的影响，突出声音事件高能等级频带成分及其特征。

本文把MBPD划分成256个频带、64个能量等



(a) −10 dB 风声环境下茶集叫声波形与 gammatone 频谱图　(b) −10 dB 风声环境下茶集叫声 MBPD

(c) 纯净茶集叫声波形与 gammatone 频谱图　(d) 纯净茶集叫声 MBPD

(e) 风声波形与gammatone频谱图　(f) 风声MBPD

图 7 风声环境下−10 dB茶集叫声、纯净茶集叫声以及风声的波形图、gammatone频谱图和MBPD

级和对MBPD图8×8分块。在实际应用中可以根据具体待测声音事件的音域变化做调整。如表4与表5所示，对于信噪比−10 dB的动物声音事件，音域较宽而且时域变化较复杂的信噪比−5 dB的办公室声音事件，检测率不理想。这种情况，可以通过对特定的声音事件，进行更细致的频带及能量等级划分，提取更有效的MBPD-DCTZ来改善检测率。

6　结束语

本文针对各种环境下低信噪比声音事件检测问题，以自然环境下的各种动物声音事件和办公室声音事件为例，提出一种基于MBPD-DCTZ特征的随机森林检测方法。该方法在平稳噪声条件及5种非平稳噪声条件下的检测率都明显优于现有的方法，尤其是信噪比低于0 dB的情况下，其优势更加突出。我们认为，基于本文的MBPD及MBPD-DCTZ方法，还可以把MBPD图、MBPD图的映射以及MBPD-DCTZ与深度学习技术相结合，进一步改善与提高低信噪比下的各种声音事件的检测率。

参 考 文 献

[1] 米建伟, 方晓莉, 仇原鹰. 非平稳背景噪声下声音信号增强技术[J]. 仪器仪表学报, 2017, 38(1): 17–22. doi: 10.3969/j.issn.0254-3087.2017.01.003.
MI Jianwei, FANG Xiaoli, and QIU Yuanying. Enhancement technology for the audio signal with nonstationary background noise[J]. *Chinese Journal of Scientific Instrument*, 2017, 38(1): 17–22. doi: 10.3969/j.issn.0254-3087.2017.01.003.

[2] 汪家冬, 邹采荣, 蒋本聪, 等. 基于数字助听器声音场景分类的噪声抑制算法[J]. 数据采集与处理, 2017, 32(4): 825–830. doi: 10.16337/j.1004-9037.2017.04.021.
WANG Jiadong, ZOU Cairong, JIANG Bencong, *et al*. Noise reduction algorithm based on acoustic scene classification in digital hearing aids[J]. *Journal of Data Acquisition and Processing*, 2017, 32(4): 825–830. doi: 10.16337/j.1004-9037.2017.04.021.

[3] FENG Zuren, ZHOU Qing, ZHANG Jun, *et al*. A target guided subband filter for acoustic event detection in noisy environments using wavelet packets[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2015, 23(2): 361–372. doi: 10.1109/TASLP.2014.2381871.

[4] GRZESZICK R, PLINGE A, and FINK G A. Bag-of-features methods for acoustic event detection and

classification[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017, 25(6): 1242–1252. doi: 10.1109/TASLP.2017.2690574.

[5] REN Jianfeng, JIANG Xudong, YUAN Junsong, *et al*. Sound-event classification using robust texture features for robot hearing[J]. *IEEE Transactions on Multimedia*, 2017, 19(3): 447–458. doi: 10.1109/TMM.2016.2618218.

[6] YE Jiaxing, KOBAYASHI T, and MURAKAWA M. Urban sound event classification based on local and global features aggregation[J]. *Applied Acoustics*, 2017, 117: 246–256. doi: 10.1016/j.apacoust.2016.08.002.

[7] CAKIR E, PARASCANDOLO G, HEITTOLA T, *et al*. Convolutional recurrent neural networks for polyphonic sound event detection[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017, 25(6): 1291–1303. doi: 10.1109/TASLP.2017.2690575.

[8] SHARAN R V and MOIR T J. Robust acoustic event classification using deep neural networks[J]. *Information Sciences*, 2017, 396: 24–32. doi: 10.1016/j.ins.2017.02.013.

[9] OZER I, OZER Z, and FINDIK O. Noise robust sound event classification with convolutional neural network[J]. *Neurocomputing*, 2018, 272: 505–512. doi: 10.1016/j.neucom.2017.07.021.

[10] WANG Jiaching, LIN Changhong, and CHEN Bowei. Gabor-based nonuniform scale-frequency map for environmental sound classification in home automation[J]. *IEEE Transactions on Automation Science and Engineering*, 2014, 11(2): 607–613. doi: 10.1109/TASE.2013.2285131.

[11] SHARMA A and KAUL S. Two-stage supervised learning-based method to detect screams and cries in urban environments[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2016, 24(2): 290–299. doi: 10.1109/TASLP.2015.2506264.

[12] DENNIS J, TRAN H D, and CHNG E S. Image feature representation of the subband power distribution for robust sound event classification[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2013, 21(2): 367–377. doi: 10.1109/TASL.2012.2226160.

[13] DENNIS J, TRAN H D, and LI Haizhou. Spectrogram image feature for sound event classification in mismatched conditions[J]. *IEEE Signal Processing Letters*, 2011, 18(2): 130–133. doi: 10.1109/LSP.2010.2100380.

[14] SLANEY M. An efficient implementation of the Patterson-Holdsworth auditory filter bank[R]. Apple Computer Technical Report, 1993.

[15] PAPAKOSTAS G A, KOULOURIOTIS D E, and KARAKASIS E G. Efficient 2-D DCT Computation from An Image Representation Point of View[M]. London, UK, Intch Open, 2009: 21–34. doi: 10.5772/7043.

[16] LAY J A and GUAN Ling. Image retrieval based on energy histograms of the low frequency DCT coefficients[C]. IEEE International Conference on Acoustic, Speech and Signal Processing, Arizona, USA, 1999: 3009–3012. doi: 10.1109/ICASSP.1999.757474.

[17] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5–32. doi: 10.1023/A:1010933404324.

[18] Universitat Pompeu Fabra. Repository of sound under the creative commons license, Freesound. org[OL]. http://www.freesound.org, 2012.5.14.

[19] IEEE Signal Processing Society, Tampere University of Technology, Queen Mary University of London, *et al*. IEEE DCASE 2016 Challenge[OL]. http://www.cs.tut.fi/sgn/arg/dcase2016/, 2016.

[20] CHANG Chihchung and LIN Chihjen. LIBSVM: A library for support vector machines[J]. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2(3): 1–27. doi: 10.1145/1961189.1961199.

[21] COVER T and HART P. Nearest neighbor pattern classification[J]. *IEEE Transactions on Information Theory*, 1967, 13(1): 21–27. doi: 10.1109/TIT.1967.1053964.

[22] ZHENG Fang, ZHANG Guoliang, and SONG Zhanjiang. Comparison of different implementations of MFCC[J]. *Journal of Computer Science and Technology*, 2001, 16(6): 582–589. doi: 10.1007/BF02943243.

[23] KIM C and STERN R M. Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring[C]. IEEE International Conference on Acoustic, Speech and Signal Processing, Dallas, USA, 2010: 4574–4577. doi: 10.1109/ICASSP.2010.5495570.

[24] 魏静明, 李应. 利用抗噪纹理特征的快速鸟鸣声识别[J]. 电子学报, 2015, 43(1): 185–190. doi: 10.3969/j.issn.0372-2112.2015.01.029.

WEI Jingming and LI Ying. Rapid bird sound recognition using anti-noise texture features[J]. *Acta Electronica Sinica*, 2015, 43(1): 185–190. doi: 10.3969/j.issn.0372-2112.2015.01.029.

[25] KOBAYASHI T and YE J. Acoustic feature extraction by statictics based local binary pattern for environmental sound classification[C]. IEEE International Conference on Acoustic, Speech and Signal Processing, Florence, Italy, 2014: 3052–3056. doi: 10.1109/ICASSP.2014.6854161.

[26] RAKOTOMAMONJY A and GASSO G. Histogram of gradients of time-frequency representations for audio scene classification[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2015, 23(1): 142–153. doi: 10.1109/TASLP.2014.2375575.

李 应：男，1964年生，教授，研究方向为信息安全、多媒体数据检索.

吴灵菲：女，1994年生，硕士生，研究方向为信息安全、模式识别.