

时频图像特征用于声场景分类

高 敏¹, 尹雪飞¹, 陈克安²

(1. 西北工业大学电子信息学院, 陕西西安 710129; 2. 西北工业大学航海学院, 陕西西安 710072)

摘要: 为解决根据音频流识别声场景的问题, 对音频信号进行恒 Q 变换, 得到其时频表达图像, 然后进行滤波平滑等处理, 随之提取能够表述信号谱能量变化方向信息的梯度直方图特征, 以及能够捕捉信号谱纹理信息的局部二值模式特征, 输入具有线性核函数的支持向量机分类器, 对不同声场景数据进行分类实验。结果表明, 相对于传统的时频域特征和梅尔频率倒谱系数特征, 所提出的特征基本能够捕捉到给定声场景具有区分度的信息, 所得分类率更高, 且两者的互补作用使得联合特征分类效果达到最优, 该方法为声信号特征提取贡献了一种新思路。

关键词: 声场景; 恒 Q 变换; 梯度直方图; 局部二值模式

中图分类号: TN911.72

文献标识码: A

文章编号: 1000-3630(2017)-05-0399-06

DOI 编码: 10.16300/j.cnki.1000-3630.2017.05.001

Time-frequency representation based feature extraction for audio scene classification

GAO Min¹, YIN Xue-fei¹, CHEN Ke-an²

(1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, Shaanxi, China;
2. School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China)

Abstract: To recognize audio scene in a complex environment according to an audio stream, a constant-Q transform is chosen to obtain the time-frequency representation (TFR) of the signal. Due to the lack of prior knowledge on the signal and noise, a mean filtering is used to smooth the TFR image, then the features based on the histogram of gradients (HOG) of the TFR image are extracted, which can reflect the local direction of variation (both in time and frequency) of the signal power spectrum. Consequently the Local Binary Pattern (LBP) feature is considered, which captures the texture information of the signal. As for the classification algorithm, support vector machine with linear kernel function is used. Classification experiment has been done on the data of different acoustic scenes. Compared with the classical audio features such as MFCCs, the proposed features capture the discriminative power of a given audio scene to show good performance in classification, and the combined features achieve the best results. It is valuable in the field of feature extraction of acoustic signal.

Key words: acoustic scene classification; constant-Q transform; histogram of oriented gradient; local binary pattern

0 引 言

将语义标签和音频流相关联以标识产生声音的声环境, 此问题被称为声场景分类(Acoustic Scene Classification, ASC)^[1], 它是计算听觉场景分析背景下最困难的任务之一。此种分类任务在机器听声领域非常重要, 其应用包括语境识别服务、智能可穿戴设备、机器人导航系统和音频的归档管理等。

声场景分类是一个相当复杂的问题, 因为一个

特定场景的录音可能由大量的单个声事件组成, 但其中只有少数声事件提供了有关该场景的信息^[2]。因此, 现在关于 ASC 的工作主要集中在提取声信号特征的方法上^[3-6], 人们期望, 所提取的特征能够捕捉到一些给定声事件具有区分度的信息。单独的时域或频域特征识别效果较差, 而时频表达可将声信号在二维平面上可视化, 反映了其时频结构信息, 有利于克服通过特征融合途径获得时频联合信息的弊端, 避免了融合过程中特征间的彼此抑制。因此, 可以借鉴图像处理领域发展较为成熟的识别方法, 对声音时频表达图像进行识别, 从而达到声识别的目的。图像处理中的梯度直方图(Histogram of Oriented Gradient, HOG)特征, 可以描述时频表达图像的形状, 捕捉声音谱能量变化的方向信息; 局部二值模式(Local Binary Pattern, LBP)可以描述其局部纹理特征, 捕捉谱能量的缓慢变化或周期性

收稿日期: 2016-11-04; 修回日期: 2017-03-15

基金项目: 国家自然科学基金资助项目(11574249、11074202)

作者简介: 高敏(1991—), 女, 山西运城人, 硕士研究生, 研究方向为信号与信息处理。

通讯作者: 高敏, E-mail: 253191300@mail.nwpu.edu.cn

变化信息。两者组合更具有互补作用,有利于进一步提高特征的效能。

本文选取不同声场景的录音作为研究对象,用恒 Q 变换来表达信号,并将 HOG、LBP 以及两者的联合特征应用于声信号的时频表达图像,输入到多类别支持向量机分类器进行分类实验,最后和传统的时频域及梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficients, MFCCs)特征分类结果进行比较,并对结果进行分析。

1 数据来源

鉴于声场景数据集缺乏共享机制,公开数据集十分有限,本文所用数据一是来自于文献[7]所提供的 D-case 数据,该文献提供了一套由专业录音师进行录制的高质量标准化、室内外场景数目均衡的开源录音样本,一共包含 10 种不同声场景,录音采样率为 44.1 kHz,总共有 3 000 s,每 5 s 的录音作为一个样本,一共有 600 个样本。数据二是由文献[8]公开提供的 EA 数据,它是由 Ma 等人于 2000 年在 East Anglia 大学收集的,一共包含 10 种声场景,采样频率为 22.1 kHz,总共 2 400 s,每 5 s 录音作为一个样本,一共有 480 个样本。

数据三是由实验获取。选取 15 种日常生活中最常见的声场景,分别为咖啡厅、餐厅、地铁、高铁、公共汽车、公园、候车室、繁华街区、课堂、篮球场、超市、商场、火车、图书室、校园,使用专业录音设备,采样频率为 44.1 kHz,陆续进行录音时间持续约一年,对于一种给定的场景,在该时间段的不同日期分多次进行录音,并保证其录制在温和的天气下进行,不受其他环境因素干扰,为了减少数据集的时间依赖性,每段录音通常持续 1 min,但是在某些地点,持续时间会达到 10 min。总体来说,一共录制了约 1 000 min 的声场景音频,最终把所有录音片段进行整合,并挑选出那些不受传声器噪声及其他因素干扰的音频组合成最终的数据库,其中每个样本持续时间为无重叠的 5 s,对于每一类声场景选取了 300 个样本进行实验,总共时长为(5×300×15) s。

2 信号的时频表达

由于声场景信号的非平稳特性,希望其时频表达体现出短时局部窗函数内信号的功率谱,通常考虑基于小波或基于短时傅里叶变换的方式,本文利

用恒 Q 变换^[9-10](Constant-Q Transform, CQT)表达信号,该方法在 1990 年被提出,一般用于语音和音乐信号的分析 and 处理。与短时傅里叶变换不同,该变换用时变的窗函数在对数刻度上进行频率分析,频率分辨率与人的听觉系统几乎一致,时频局部化描述能力更强,文献[6]已经证明它对于声场景分类中基于图像的特征提取方法更为有效。从滤波器的角度分析,CQT 是中心频率与带宽的比为定值 Q 的一组滤波器,可以在低频获得较高的频率分辨率,在高频获得较高的时间分辨率。

离散时间信号 $x(n)$ 的 CQT 变换 $X(k,n)$ 定义^[9]为

$$X(k,n) = \sum_{m=0}^N x(m) a_k^*(m-n) \quad (1)$$

其中: k 和 n 分别代表频率和时间指数; N 是输入信号 $x(n)$ 的长度; $a_k^*(t)$ 是调制窗函数 $a_k(m)$ 的共轭复数。 $a_k(m)$ 的定义^[9]为

$$a_k(m) = g_k(m) e^{i2\pi m f_k / f_s}, \quad m \in Z \quad (2)$$

其中: $g_k(m)$ 是窗函数; f_s 是采样频率; f_k 为各频带的中心频率。 f_k 可表示为^[9]

$$f_k = f_0 2^{k/b}, \quad k=0,1,\dots,K,K-1 \quad (3)$$

其中: b 是一个八度内划分的音符的个数,一般取 12 或 24, f_0 是初始最低频率, K 是总的频带数目。在 CQT 中,中心频率和带宽的比值定义为常数 Q ^[9]:

$$Q = (2^{1/b} - 1)^{-1} \quad (4)$$

一个八度内的频率分布并非均匀,而是呈指数分布。

对信号进行恒 Q 变换后,为了获得不依赖于信号长度和采样频率的特征,对 CQT 矩阵进行双三次插值处理,调整得到像素为 512×512 的时频表达图像,该图像保留了声场景的时频结构信息。其次,由于对信号噪声缺少先验知识,所以利用均值滤波来平滑时频表达图像,其目的是减小图像中局部的强变化。图 1 及图 2 所示为地铁声场景的时域波形和处理后的 CQT 时频图。

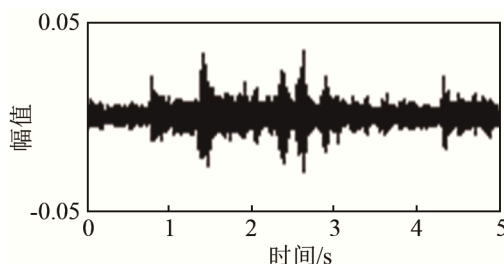


图 1 地铁声场景中信号波形图
Fig1 Waveform of the signal from Tube audio scene

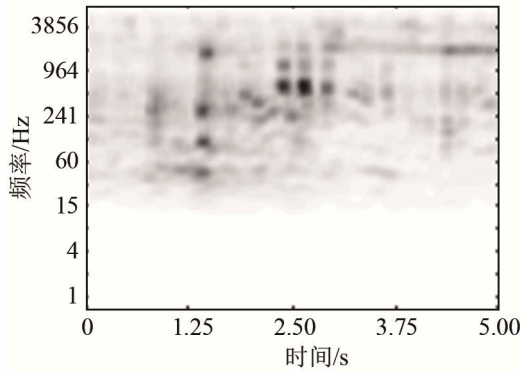


图 2 地铁声场景中信号 CQT 时频图
Fig.2 CQT representation of the signal from Tube audio scene

3 特征提取

3.1 梯度直方图

特征提取的主要目标是，捕捉时频结构的形状信息，并期望捕捉到的时频结构信息和其所属声场景特性有关。计算视觉领域^[5]的研究表明，局部形状信息可以通过梯度密度和方向来表示。梯度直方图基本上给出了图像局部区域关于梯度方向出现次数的信息，因此，它们能够描述该区域的形状。

计算图像的 HOG 主要基于以下步骤^[11]：

(1) 计算时频图像中每个像素 $f(i, j)$ 的梯度幅值和方向：

$$Mag(i, j) = \{ [f(i+1, j) - f(i-1, j)]^2 + [f(i, j+1) - f(i, j-1)]^2 \}^{1/2} \quad (5)$$

$$\theta(i, j) = \arctan \left[\frac{f(i, j+1) - f(i, j-1)}{f(i+1, j) - f(i-1, j)} \right] \quad (6)$$

(2) 将图像分成无重叠的单元(cell)。

(3) 统计给定单元中各梯度方向的出现次数。

(4) 最终根据相邻单元直方图范数，对每个单元直方图进行归一化。

对于上述 512*512 时频图，计算每个 32*32 像素单元，具有 8 个方向的梯度直方图特征，结果得到特征向量的维数是 $16^2 \times 8 = 2048$ 。如果减小单元的尺寸或增加直方图计算的方向数目，这个维数会进一步增加。由于声场景实例的数目较大，可以通过合并梯度直方图的方法降低特征向量的维度。

合并的根本思想是将局部区域的特征进行合并，变为另一个更低维的特征，但它仍保持了与邻近区域的相关性。这种合并有助于获得更稳健的信息。主要考虑以下的合并方法^[12]：

(1) 随时间边缘化合并：平均时频表达中沿时间轴的所有直方图，其结果是在特征向量中丧失了所有的时域信息。

(2) 随频率边缘化合并：平均时频表达中沿频率轴的所有直方图，其结果是在特征向量中丧失了所有的频域信息。

(3) 分块合并：对相邻的单元进行分块，合并其中所有的特征，块尺寸的大小由用户自己定义。

对于上述图 2 所示的时频图，计算每个 8*8 像素单元，具有 8 个方向的梯度直方图，结果如图 3 所示。可以看出，HOG 正确捕捉到了信号功率谱变化的方向。

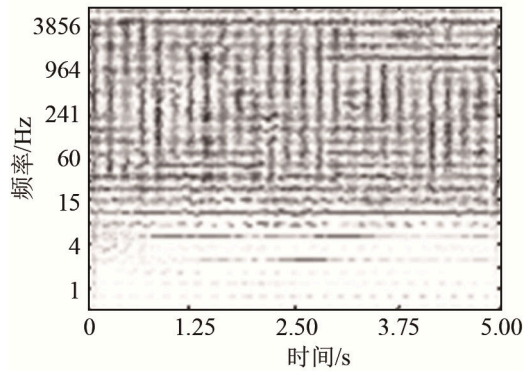


图 3 地铁声场景中信号梯度直方图
Fig.3 HOG of the signal from Tube audio scene

3.2 局部二值模式

局部二值模式用于描述图像的局部纹理特征，在时频图像中表现为捕捉谱能量的缓慢变化或周期性变化信息^[13]。其核心思想是，设定一个像素窗口，用窗口内中心像素的灰度值作为阈值，与其邻域相比较，得到的二进制码称为一种模式并作为该局部的纹理特征。

计算图像的 LBP 特征^[14]主要基于以下步骤：

(1) 对整个时频图使用 LBP 算子标记所有的像素。

(2) 将时频图划分成若干相等的单元。

(3) 统计每个单元 LBP 值出现的概率以得到直方图。

(4) 最后将每个单元的统计直方图连接成为一个特征向量，也就是整幅图的 LBP 纹理特征向量。

LBP 等价模式算子见公式(7)

$$U(G_p) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (7)$$

其中： $s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$ ， $g_p (p=0, \dots, c, \dots, P-1)$ 对应

p 处像素的灰度值， g_c 是中心像素灰度值， P 为窗口内像素总个数。若某种模式的计算结果 $U(G_p)$ 小于或等于 2，则将其归为等价模式类，除此之外归为混合模式类，改进后的模式数量由 2^P 降为

$P(P-1)+3$ 种。

本文对上述 512×512 的时频图, 单元划分为 64×64 大小, 采用 LBP 等价模式算子, 通过 3×3 邻域内的 8 个采样点计算得到 $8 \times 8 \times 59 = 3776$ 维特征向量, 其中某一单元的 LBP 直方图如图 4 所示。

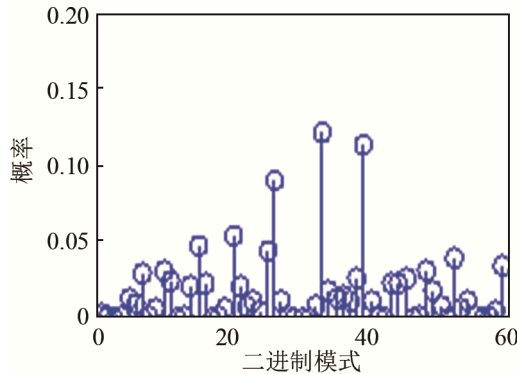


图 4 地铁声场景中信号 LBP 直方图

Fig.4 LBP histogram of the signal from Tube audio scene

4 实验结果比对及分析

支持向量机(Support Vector Machine, SVM)已普遍用于各种数据的分类, 且表现出良好的分类性能^[15], 用具有高斯核函数和线性核函数的 SVM 算法, 每类声场景选取 1/2 的训练样本进行训练, 其余用做测试, 采用十折交叉验证方式进行分类实验。

HOG 特征的梯度方向考虑有符号和无符号两种, 沿频率和时间方向相邻的 64 个单元进行合并可得到 1536 维特征向量, 使用具有线性核函数的 SVM 在本文数据集上进行分类, 结果如表 1 所示。不同的合并方式对分类结果有较大影响, 随频率边缘化合并的效果最差, 因为该方法丢失了关于频谱内容的所有信息; 随时间边缘化合并特征中缺乏时间信息, 但获得了较好的分类精度, 这是因为声场景的频谱内容比时间内容更具有区分性, 大多数的声场景总体上是时不变的, 尽管某些短时声事件, 可能携带了具有区分性的信息, 但多数声场景的周

表 1 不同合并方式的分类效果

Table 1 The classification effects of different pooling methods

单元合并方向/个		分类率/%
沿频率	沿时间	
1	64	61.33
2	32	72.31
4	16	74.89
8	8	83.16
16	4	89.33
32	2	93.47
64	1	92.00

期性模式可进行全局性分析^[5], 促进分类效果; 分块合并方式可以达到最高分类精度, 该方式的显著趋势是: 随着频率边缘化合并减小, 时间边缘化合并增加, 分类精度先提高, 后降低。在分块大小为 32×2 时, 分类精度达到最高。

使用上述分块方式效果最好的 HOG 特征、LBP 特征及两者联合特征在三个数据集上进行实验, 另外为了评估本文所用算法的识别性能, 还用对信号进行分帧后得到的时频域特征^[16-17], 将其标记为 TFF, 包括零交点比率(Zero-Crossing Rate, ZCR)、谱质心(Spectral Centroid, SC)、谱下降值(Spectral Roll-Off, SRO)、谱通量(Spectral Flux, SF)、线性预测倒谱系数(Linear Prediction Cepstrum Coefficient, LPCC), 一共是 $1+1+1+1+12=16$ 维, 其次考虑到应用于声场景识别效果较为突出的特征之一是 MFCC^[18], 本文将 MFCC 及其一阶二阶差分进行平均, 得到每帧信号 39 维特征向量也用作进行比较的基准特征, 这些特征能够很好地描述信号的动态性能。

用以上特征在三个数据集上分别进行实验, 所得识别率如表 2 所示。观察可知不同核函数得到的识别率不同, 总体来看线性核函数效果较好, 更适用于本文所提取的特征向量。联合特征得到的识别率最高, 其次是 HOG 特征和 LBP 特征, 时频域特征效果最差。分析其原因, 一方面, 传统的时频域特征不能很好地捕捉到声音信号时频结构中关于形状和演变的相关特征, MFCC 的本质是捕捉了信号功率谱中的非线性信息^[5], 该信息并非声场景中具有区分性的信息, 而时频表达的 HOG 和 LBP 特征却可以提供具有强区分性的信号谱能量变化的局部方向信息及周期性变化信息, 且两者具有互补作用, 使得联合特征识别率更高。另一方面, 由于大多数声场景具有一定的周期性模式, 因此可以忽略其中短时的单个声事件所携带的信息而进行全局性分析, HOG 和 LBP 特征计算过程中将图像分割成单元的步骤, 恰恰使得它们对于小的时间和频率平移是不变的, 这有助于识别效果。

表 2 不同特征的分类效果(%)

Table 2 The classification accuracies of different feature sets(%)

		TFF	MFCC	LBP	HOG	LBP_HOG
D-case	线性	51.33	58.87	64.53	81.67	84.23
	数据	高斯	46.37	56.80	62.50	67.63
EA	线性	70.42	77.08	85.83	96.25	97.08
	数据	高斯	62.50	77.08	83.79	95.25
本文	线性	71.33	76.53	87.63	93.47	95.64
	数据	高斯	70.33	77.52	81.64	85.35

另外本文数据的实验结果优于公开的 D-case 数据，其主要原因：一是录制同类场景音频所选取的场景个数有限(场景的多样性不如 D-case 数据)，例如公园场景一共选取了四个不同的城市公园进行录制，这可能导致场景类内差距较小；二是测试集和训练集所包含的不同音频片段可能来自于同一个录音文件，它们之间的时间相关性使得分类率有所提高。

用 Matlab 8.0 版本进行仿真实验，对每个样本得到 TFF 特征的维数是 16*332，MFCC 特征的维数是 39*332(其中 332 表示帧个数)，LBP 特征的维数是 3 776，以及 HOG 特征的维数是 1 536，因此相对而言，后两种特征所占用的内存空间更小。在 D-case 数据集上计算每种特征所消耗的时间如表 3 所示，由表 3 可知，本文所提出的三种特征由于计算过程复杂，因此时间复杂度较高，可见识别率的提高是以增加计算时间为代价。

表 3 几种特征计算时间对比

Table 3 The computing time of several feature sets

特征	TFF	MFCC	LBP	HOG	LBP_HOG
时间/s	0.084	0.233	1.008	1.622	1.376

在 D-case 数据集上用 LBP_HOG 特征对声场景进行分类，得到的识别率混淆矩阵如表 4 所示，其中横向为预测类标签，纵向为实际类标签，从中可注意到，与其他场景具有显著差异的公共汽车和繁

表 4 分类结果混淆矩阵

Table 4 The confusion matrix of classification results(%)

	公共汽车	繁华街区	办公室	露天市场	公园
公共汽车	100.00	0	0	0	0
繁华街区	3.33	96.67	0	0	0
办公室	0	0	90.00	0	3.33
露天市场	0	0	0	86.67	0
公园	0	0	6.67	0	80.00
宁静街区	0	0	6.67	0	16.67
餐厅	0	0	0	10.00	0
超市	3.33	0	6.67	0	0
地铁	3.33	0	0	0	0
地铁站	0	0	0	6.67	0
	宁静街区	餐厅	超市	地铁	地铁站
公共汽车	0	0	0	0	0
繁华街区	0	0	0	0	0
办公室	6.67	0	0	0	0
露天市场	0	6.67	6.67	0	0
公园	10.00	0	3.33	0	0
宁静街区	70.00	0	3.33	0	3.33
餐厅	3.33	80.00	6.67	0	0
超市	0	6.67	76.67	0	6.67
地铁	0	3.33	0	83.33	10.00
地铁站	0	3.33	0	16.67	73.33

华街区能够被精确地识别，较为相似的场景如公园和宁静街区，地铁和地铁站台出现一些混淆现象，但总体的识别率达到 83.67%，比文献[1]所提出的定量递归分析联合 MFCC 特征所获得的最高识别率 83.2%高出 0.47%，平均识别率达到 84.23%，高出 1%左右，说明本文所提特征能够很好地捕捉到不同声场景中的区分性信息。

5 总结

本文首先对不同声场景的声信号进行恒 Q 变换得到其时频表达，在此基础上用图像处理中的梯度直方图特征和局部二值模式提取信号谱能量变化的局部信息，输入支持向量机分类器，与传统的时频特征和 MFCC 特征进行比较，结果证明，所提出的基于声音信号时频表达图像所提取的特征能够很好地捕捉到声场景中具有区分度的信息，且两者联合特征具有互补作用，效果更好。

参 考 文 献

- [1] Barchiesi D, Giannoulis D, Dan S, et al. Acoustic scene classification: classifying environments from the sounds they produce[J]. IEEE Signal Processing Magazine, 2015, 32(3): 16-34.
- [2] Stowell D, Giannoulis D, Benetos E, et al. Detection and classification of acoustic scenes and events[J]. IEEE Transactions on Multimedia, 2015, 17(10): 1733-1746.
- [3] Ghoraani B, Krishnan S. Time-frequency matrix feature extraction and classification of environmental audio signals[J]. IEEE Transactions on Audio Speech & Language Processing, 2011, 19(7): 2197-2209.
- [4] Cotton C V, Ellis D P W. Spectral vs. spectro-temporal features for acoustic event detection[C]//Applications of Signal Processing to Audio and Acoustics, IEEE Workshop on. IEEE, 2011, 69-72.
- [5] Roma G, Nogueira W, Herrera P. Recurrence quantification analysis features for environmental sound recognition[J]. BMC Public Health, 2013, 9(22): 1-4.
- [6] Bisot V, Serizel R, Essid S, et al. Acoustic scene classification with matrix factorization for unsupervised feature learning[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, 2016, 6445-6449.
- [7] Giannoulis D, Stowell D, Benetos E, et al. A database and challenge for acoustic scene classification and event detection[C]// European Signal Processing Conference, 2013, 1-5.
- [8] Ma L, Smith D J, Milner B P. Context awareness using environmental noise classification[C]//European Conference on Speech Communication and Technology, Eurospeech, 2003, 1-4.
- [9] Schörkhuber C, Klapuri A, Holighaus N, et al. A matlab toolbox for efficient perfect reconstruction time-frequency transforms with log-frequency resolution[C]//Aes Conference on Semantic Audio, 2014, 1-8.
- [10] Schörkhuber C, Klapuri A, Sontacchi A. Audio pitch shifting using the constant-Q transform[J]. Journal of the Audio Engineering Society, 2013, 61(7/8): 562-572.
- [11] Minetto R, Thome N, Cord M, et al. An effective gradient-based

- descriptor for single line text regions[J]. Pattern Recognition, 2013, 46(3): 1078-1090.
- [12] Boureau Y L, Ponce J, Lecun Y. A theoretical analysis of feature pooling in vision algorithms[C]//Proc. International Conference on Machine Learning, 2010, 328-33.
- [13] Kobayashi T, Ye J. Acoustic feature extraction by statistics based local binary pattern for environmental sound classification[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, 2014, 3052-3056.
- [14] Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Software Engineering, 2010, 32(9): 1627-45.
- [15] Tan L N, Alwan A, Kossan G, et al. Dynamic time warping and sparse representation classification for birdsong phrase classification using limited training data[J]. J. Acoust. Soc. Am., 2015, 137(3): 1069-80.
- [16] Karbasi M, Ahadi S M, Bahmanian M. Environmental sound classification using spectral dynamic features[C]//IEEE Communications and Signal Processing, 2011, 1-5.
- [17] 陈克安. 环境声的听觉感知与自动识别[M]. 北京: 科学出版社, 2014.
CHEN Kean. Auditory perception and automatic recognition of environmental sound[M]. Beijing: Science Press, 2014.
- [18] Chakrabarty D, Elhilali M. Exploring the role of temporal dynamics in acoustic scene classification[J]. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2015, 10(11): 1-5.

• 简 讯 •

中国的球面波聚焦场自易法校准技术成为 IEC 技术规范的标准方法

2017年9月14日国际电工委员会(IEC)超声学技术委员会(TC87)主席 Volker Wilkens 电邮告知: 由中国 IEC 委员会申请、上海交通大学生物医学工程学院寿文德教授任领导人兼执笔、江苏省医疗器械检验所胡济民高级工程师参加, 德国、意大利、韩国及美国专家参与编写的 IEC TS 62903 Ultrasonics-Measurements of electroacoustical parameters and output acoustic power of spherically curved transducer using self-reciprocity method 的终审技术规范(TS)文本已送达 IEC 中央办公室(进入最终编辑、修图、排版、法文版译编等工作程序)。预计在完成期限(2018年7月)前率先用英法文双语出版, 至此我国起草工作任务基本完成。一个基于我国创立的球形曲面聚焦换能器的绝对校准方法——自易法的 IEC 国际标准方法即将诞生!

球面波聚焦场自易法校准球缺面换能器技术在 2002 年由寿文德教授等首先建立。他们正确地定义了一对球面聚焦超声换能器的自由场发射电流响应和接收电压灵敏度, 把电声互易定理推广到球面聚焦声场。在几何声学和镜像原理模式下, 导出了自由场球面聚焦场互易参量和相应的电声参数的表达式。再引入实际校准中存在的衍射修正系数、反射器的平均反射系数、声衰减修正系数后, 得到了精确的计算公式, 为本方法奠定了理论基础。寿文德教授组织了全国 11 个专业单位 14 位专家成立工作组, 开展测量、验证和不确定度估计等大量工作, 起草了国家标准 GB/T32522-2016 《声学 压电球面聚焦超声换能器的电声特性及其测量》, 从此自易法成为我国测量球缺面换能器电声参数的标准方法。

2015 年我国成功地申请了基于自易法的 IEC 技术规范 TS62903 项目。在项目起草过程中, 国际超声理论权威之一、德国联邦技术物理研究院(PTB)的 Klaus Beissner 曾两次对我国草案中的自易法和声功率测量计算公式提出过否定性质疑, 指出了标准草案中的若干叙述和用词的错误, 同时也为项目组提供了 4 篇最新的相关文章和理论新结论供参考, 据此提出 Rayleigh 积分的有效空间的有限性需要定量界定。此外 Klaus Beissner 还要求补充全频率范围内适用性的实验依据等严格的要求, 帮助草案文件的修改。

为此, 寿文德教授仔细研究了所提供的新论点和论据, 修改了原来国标中存在的一些错误论述和一种推导路径, 吸收了 Klaus Beissner 提出的创新理论成果, 用充分的实验数据和严格的不确定度估计, 对规范草案进行了补充, 从而使自易校准理论和实验更加正确完善。曾两次专门做了解释性的幻灯片和文字, 在会议上进行演示, 说明了改进后的方法的合理性和可行性, 最终获得了国际同行的完全赞同。这个技术规范是一个高水平、高质量的创新性技术规范, 是国际合作交流的成功典范。

球缺面聚焦换能器的自易法校准, 通过一次测量就可测得球缺形聚焦换能器的发射响应、接收电压灵敏度、发射功率、辐射电导和电声效率等一系列参数。测量准确, 属于一级绝对校准方法。它不但填补了电声换能器互易校准的理论和实验的空白, 而且抗干扰能力强, 信噪比高, 设备简单易行, 操作简便, 有广泛而实际的用途。对高频、低占空比、大功率超声聚焦换能器的电声参数和声功率测量有独特的优点。