文章编号：0427-7104(2019)03-0269-45

理解数字声音——基于一般音频/环境声的计算机听觉综述

李 伟[1,2]，李 硕[1]

（1.复旦大学 计算机科学技术学院，上海 201203；2.复旦大学 上海市智能信息处理重点实验室，上海 200433）

摘 要：声音是人类获取信息的重要来源，对声音内容进行自动分析和理解具有重要意义.本文介绍声音的基本知识，从信号、听觉感受、声音特性等 3 个角度对声音进行分类，阐明各个分类之间的关系，明确基于一般音频/环境声的计算机听觉技术的研究对象和学科位置.之后，介绍计算机听觉技术的基本概念、原理、研究课题和技术框架.作者全面总结了计算机听觉技术在各个领域中：包括医疗卫生，安全保护，交通运输、仓储，制造业，农、林、牧、渔业，水利、环境和公共设施管理业，建筑业，其他采矿业、日常生活、身份识别、军事等的典型应用.分类总结了各领域计算机听觉应用中现有典型文献的基本原理、技术路线.最后总结计算机听觉领域存在的各方面问题，并展望未来发展趋势.

关键词：数字声音；一般音频/环境声；计算机听觉；音频信号处理；人工智能

中图分类号：TP391 文献标志码：A

DOI:10.15943/j.cnki.fdxb-jns.2019.03.001

1 声音概述

声音在现实世界中无所不在，种类繁多.有的声音由人创造，有的存在于自然界和日常生活中.听觉和视觉对于感知系统一样重要，密不可分，缺一不可.声音蕴含着极大的信息量.例如，轰隆隆的雷声预示快要下雨，动物的叫声表征其种类，人类语言可用于分辨性别甚至具体的人，交响乐队的乐器声让人知道这是一场古典音乐会，鸟叫声通常暗示周围有很多树，枪炮声代表战争场面，有经验的技师听到汽车发动机的声音就能大体判断出存在的故障，经过训练的声呐员通过声呐接收的水下声信号就可以判断水下目标的类型，诸如此类，无法尽数.因此，对声音的内容进行基于信息科技的自动分析与理解，在语言交互、数字音乐、工业、农业、生物、军事、安全等几乎所有的自然和社会领域都具有重要的现实意义.本文阐述的局限于人耳能听到的声音，人类感觉不到的超声波和次声波不在所述范围之内.

声音是一种物理波动现象，即声源振动或气动发声所产生的声波.声波通过空气、固体、液体等介质传播，并能被人或动物的听觉器官所感知.人类听到的声音基本都是在空气中传播.振动源周围空气分子的振动形成疏密相间的纵波传播机械能，一直延续到振动消失.声波具有一般波的各种特性，包括反射(Reflection)、折射(Refraction)和衍射(Diffraction)等.声音还是一种心理感受，不仅与人的生理构造和声音的物理性质有关，还受到环境和背景的影响.例如，同样的一段乐曲，轻松时听起来让人愉悦，紧张时听起来却让人烦躁.

从信号的角度看，声音可分为纯音(Pure tone)、复合音(Compound tone)和噪声(Noise).纯音和复合音都是周期性声音，波型具有一定的重复性，具有明显的音高(Pitch).纯音是只具有单一频率的正弦波，通常只能由音叉、电子器件或合成器产生，在自然环境下一般不会发生.我们在日常生活和自然界中听到的声音大多是复合音(有少量不是，例如清辅音)，由许多参数不同的正弦波分量叠加而成.复合音信号可用正弦波模型(Sinusoidal Model，SM)模拟，即任何复杂的周期振动都可以分解为多个具有不同频率、不

收稿日期：2019-01-14

基金项目：国家自然科学基金(61671156)

作者简介：李 伟(1970—)，男，教授，博士生导师，E-mail：weili-fudan@fudan.edu.cn

同强度、不同相位的正弦波的叠加,如图 1 所示,图形所示波的频率从上到下依次升高.该模型也称为傅里叶分析(Fourier Analysis, FA)或频谱分析(Spectral Analysis, SA),纯音和复合音之间可以互相合成与分解.

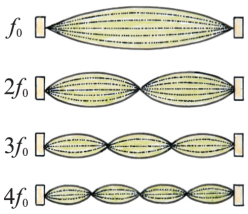

图 1 正弦波模型示意图

Fig.1 A schematic diagram of sine wave model

通常在复合音中,频率最低的正弦波(即整个波形振动的频率)称为基频(Fundamental frequency),记为 f_0,f_0 决定声音的音高.其他频率较高的的正弦分量(如 $2f_0$,2.5f_0,$3f_0$,…)称为泛音(Overtone),泛音决定声音的音色(Timbre).泛音之中频率是 f_0 整数倍的正弦分量(如 $2f_0$,$3f_0$,…)连同 f_0 统称为谐音(Harmonics).特殊情况下,在复合音中,频率最低的正弦波不是基频.例如当手机或计算机音箱播放不出低频(例如 100 Hz)以下的声音时,出现基频缺失现象.另一个相关的概念是物理上的谐波(Partial),包含 f_0 与所有泛音.在 f_0 的整数倍上谐波与谐音相同,但与泛音次数不同.如 1 次谐波/谐音定义为 f_0,2 次谐波/谐音定义为 1 次泛音,3 次谐波/谐音定义为 2 次泛音,依此类推.

声音是一种时间域(Time-domain)随机信号.声音的基本物理维度(或要素)是时间、频率(Frequency)、强度(Intensity)和相位(Phase).频率即每秒钟振动的次数,单位是赫兹(Hz),振动越快音高越高;强度与振幅的大小成正比,单位是分贝(dB),体现为声音的强弱(Dynamics);相位指特定时刻声波所处的位置,是信号波形变化的度量,以角度作为单位.两个声波相位相反会相互抵消,相位相同则相互加强.

与纯音和复合音不同,噪声是非周期性声音,由许多频率、幅度和相位各不相同的声音成分无规律地组合而成.噪声一般具有不规则的声音波形,没有明显的音高,听起来感到不舒服甚至刺耳.噪声的测量单位是分贝(dB).按照频谱的分布规律,噪声可分为白噪声(White noise)、粉红噪声(Pink noise)和褐色噪声(Brown noise)等.白噪声是指功率谱密度(Power Spectrum Density, PSD)在整个可听频域(20~20 000 Hz)内均匀分布为常数的噪声,听感上是比较刺耳的沙沙声.粉红噪声能量分布与频率成反比,主要集中于中低频带.频率每上升一个八度(Octave)能量就衰减 3 dB,所以又被称做频率反比($1/f$)噪声.粉红噪声可以模拟出自然界常见的瀑布或者下雨的声音,在人耳听感上经常会比较悦耳.褐色噪声的功率谱主要集中在低频带,能量下降曲线为 $1/f^2$.听感上有点和工厂里面轰隆隆的背景声相似.

从听觉感受的角度看,声音可分为乐音(Musical tone)和噪声两种.乐音是让人感觉愉悦的声音,通常由有规则的振动产生,具有明显的音高.如图 2 所示,乐音包括语音、歌声、各种管弦和弹拨类乐器(如小提琴、萨克斯、钢琴、吉他等)等发出的复合音(Compound Tone-Speech and Music, Compound Tone-SM),部分环境声中的复合音(Compound Tone-General Audio, Compound Tone-GA)如鸟叫,以及少量称为噪乐音(Noise tone)的打击类乐器(如锣、钹、鼓、沙锤、梆子、木鱼等)发出的噪声.噪声是让人听起来不悦耳的声音,通常由无规则的振动产生,没有明显的音高.去掉噪乐音之后其余的绝大部分噪声可称为一般噪声(Ordinary noise),包括自然界及日常生活中的风雨声、雷电声、海浪声、流水声、敲打声、机器轰鸣声、物体撞击声、汽车声、施工嘈杂声等.

从声音特性的角度看,声音可划分为语音(Speech)、音乐(Music)和一般音频/环境声(General audio/ambient sound)3 大类.人类的语言具有特定的词汇及语法结构,用于在人类中传递信息.语音是语言的声音载体,语音信号属于复合音,其基本要素是音高、强度、音长、音色等.音乐是人类创造的复杂的艺术形式,组成成分是上述的各种乐音,包括歌声、各种管弦和弹拨类乐器发出的复合音、少量来自环境声的复合音以及一些来自打击乐器的噪乐音.其基本要素包括节奏、旋律、和声、力度、速度、调式、曲式、织体、音色等.除了人类创造的语音和音乐,在自然界和日常生活中,还存在着其他数量巨大、种类繁多的声音,统称为一般音频或环境声.如图 2 所示,一般音频/环境声包含噪乐音、一般音频复合音、一般噪声,后两者是本文所述的内容.一般音频中的噪乐音主要对应于打击乐器等各种艺术化的噪声,其对应的主要学科领域是音乐声学(Music Acoustics,MA)和音乐信息检索技术(Music Information Retrieval, MIR)(见图 3),因此不在本文讨论的范围内.专门处理语音的学科是语音信息处理,以语言声学为基础,历史悠久,发展相对成熟,已独立成为一门学科.本文涉及的媒体是一般音频复合音与一般噪声,如图 2 中黑色加粗框所显

示，对应的学科领域则称为基于一般音频/环境声的计算机听觉(Computer Audition，CA).如图 3 所示，该学科与语音信息处理、音乐信息检索(MIR)技术高度相似，也主要使用音频信号处理及机器学习这两种技术，属于人工智能(Artificial Intelligence，AI)与音频领域的交叉学科，同时需要用到对应声音种类的声学知识.与相对成熟的语音信息处理和音乐信息检索技术相比，基于一般音频/环境声的 CA 技术由于各种原因发展更慢.

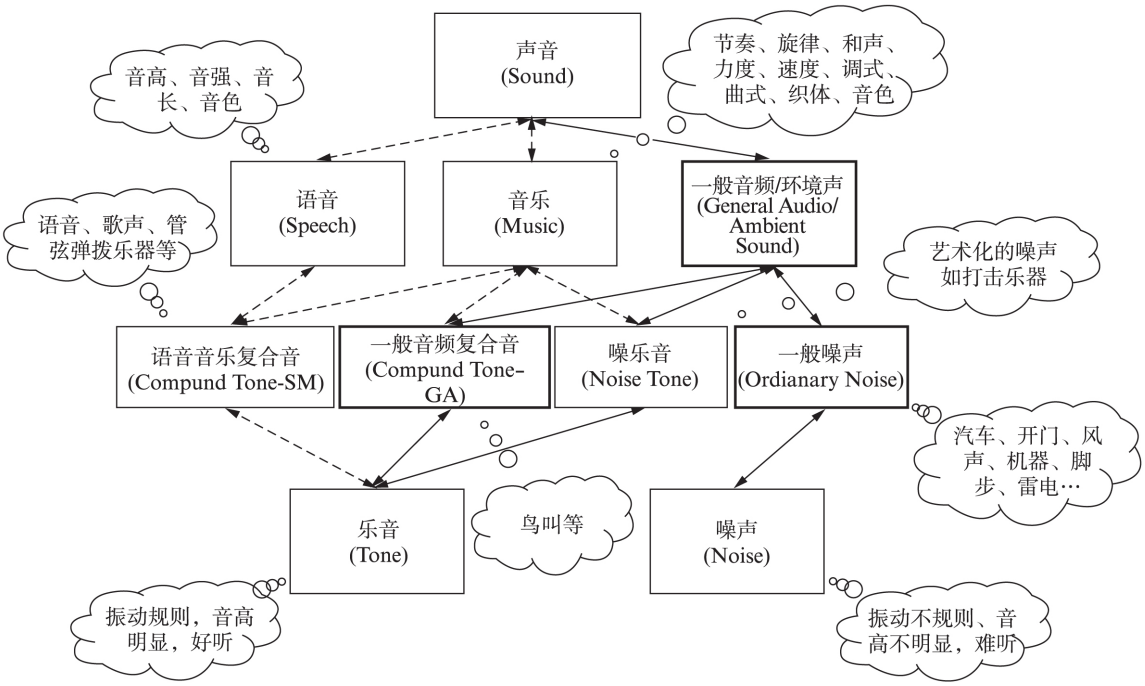

图 2　声音的种类关系图

Fig.2　A relation graph of sound type

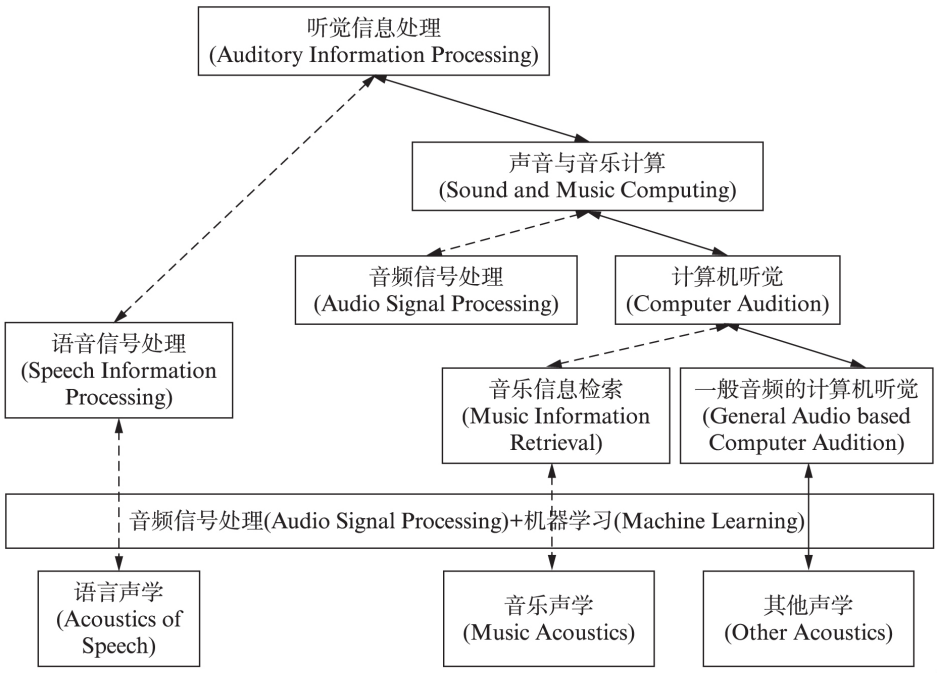

图 3　听觉信息处理各学科关系图

Fig.3　A relation graph of different disciplines about auditory information processing

2 计算机听觉简介

人类听觉系统(Human Auditory System, HAS)将外界的声音通过外耳和中耳组成的传音系统传递到内耳,在内耳将声波的机械能转变为听觉神经上的神经冲动,神经冲动传送到大脑皮层的听觉中枢,产生主观感觉.人类的听觉感知能力主要体现在通过声音特性产生主观感受(Subjective perception)、音频事件检测(Audio event detection)、声音目标识别(Acoustic target detection)、声源定位(Sound source location)等几个方面.

近 20 年来,半导体技术、互联网、音频压缩技术、录音设备及技术的共同发展使得数字格式的各种声音数量急剧增加.在人类听觉机制的启发下,诞生了一个新的学科—计算机听觉,也可称为机器听觉(Machine listening).计算机听觉是一个面向数字音频和音乐(Audio and music),研究用计算机软件(主要是信号处理及机器学习)来分析和理解海量数字音频音乐内容的算法和系统的学科.

CA 涉及乐理(Music theory)、一般声音的语义(General sound semantics)等领域知识,与音频信号处理(Audio signal processing)、音乐信息检索(MIR)、音频场景分析(Auditory science analysis)、计算音乐学(Computational musicology)、计算机音乐(Computer music)、听觉建模(Auditory modelling)、音乐感知和认知(Music perception and cognition)、模式识别(Pattern recognition)、机器学习(Machine learning)、心理学(Psychology)等学科有交叉.

从技术的角度看,CA 的研究可以被粗略地分成以下 6 个子问题.

(1) 音频时频表示(Time-frequency representation)

音频时频表示包括音频本身的表示,如信号或符号(Signal or symbolic)、单声道或双声道(Monaural or stereo)、模拟或数字(Analog or digital)、声波样本、压缩算法的参数等;音频信号的各种时频(Time-frequency, T-F)表示,如短时傅里叶变换(Short-time Fourier Transform, STFT)、小波变换(Wavelet Transform, WT)、小波包变换(Wavelet Packet Transform, WPT)、连续小波变换(Continuous Wavelet Transform, CWT)、常数 Q 变换(Constant-Q Transform, CQT)、S 变换(S-Transform, ST)、希尔伯特-黄变换(Hilbert-Huang Transform, HHT)、离散余弦变换(Discrete Cosine Transform, DCT)等;音频信号的建模表示由于种类繁多,又通常包含多个声源,无法像语音信号那样被有效地表示成某个特定的模型,如源-滤波器模型(Source-filter model),通常使用滤波器组(Filter banks)或正弦波模型来获取并捕捉多个声音参数(Sound parameters).

(2) 特征提取(Feature extraction)

音频特征是对音频内容的紧致反映,用来刻画音频信号的特定方面,有时域特征、频域谱特征、T-F 特征、统计特征、感知特征、中层特征、高层特征等数十种.典型的时域特征如过零率(Zero-Crossing Rate, ZCR)、能量(Energy),频域谱特征如谱质心(Spectral Centroid, SC)、谱通量(Spectral Flux, SF),T-F 特征如基于频谱图的 Zernike 矩、基于频谱图的(Scale Invariant Feature Transform, SIFT)描述子,统计特征如峰度(Kurtosis)、均值(Mean),感知特征如梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficients, MFCC)、线性预测倒谱系数(Linear Predictive Cepatral Coefficient, LPCC),中层特征如半音类(Chroma),高层特征如旋律(Melody)、节奏(Rhythm)、频率颤音(Vibrato)等.

(3) 声音相似性(Sound similarity)

两段音频之间或者一段音频内部各子序列(Subsequence)之间的相似性一般通过计算音频特征之间的各种距离(Distance)来度量.距离越小,相似度越高.在某些时域(Temporal)信息很重要的场合,通常使用动态时间规整(Dynamic Time Warping, DTW)来计算相似度.也可通过机器学习方法进行音频相似性计算.

(4) 声源分离(Sound Source Separation, SSS)

与通常只有一个声源的语音信号不同,现实声音场景中的环境声及音乐的一个基本特性就是包含多个同时发声的声源,因此 SSS 问题成为一个极其重要的技术难点.音乐中的各种乐器及歌声按照旋律、和声及节奏耦合起来,对其进行分离比分离环境声中各种基本不相关的声源要更加困难,至今没有方法能很好地解决这个问题.

(5) **听觉感知**(Auditory cognition)

人类欣赏音乐时引起的的情感效应(Emotional effect)**以及人类和动物对于声音传递的信息的理解，都需要从心理和生理**(Psycho-physiological)**的角度加以研究理解，不能只依赖于特定的声音特性和机器学习方法.**

(6) **多模态分析**(Multi-modal analysis)

人类对世界的感知都是结合各个信息源综合得到的.因此，对数字音频和音乐进行内容分析理解时，理想情况下也需要结合文本、视频、图像等多种媒体进行多模态的跨媒体研究.

3　计算机听觉通用技术框架及典型算法

从实际应用的角度出发，一个完整的 CA **算法系统应该包括的几个步骤如图** 4 **所示.首先使用麦克风**(Microphone)/**声音传感器**(Acoustic sensor)**采集声音数据；之后进行预处理(例如将多声道音频转换为单声道、重采样、解压缩等)；音频是长时间的流媒体，需要将有用的部分分割出来，即进行音频事件检测**(Audio Event Detection, AED)**或端点检测**(Endpoint Detection, ED)**；采集的数据经常是多个声源混杂在一起，还需进行声源分离，将有用的信号分离提取出来，或至少消除部分噪声，进行有用信号增强；然后根据具体声音的特性提取各种时域、频域、T-F 域音频特征，进行特征选择**(Feature selection)**或特征抽取**(Feature extraction)**，或采用深度学习**(Deep Learning, DL)**进行自动特征学习**(Feature learning)**；最后送入浅层统计分类器或深度学习模型进行声景**(Sound scape)**分类、声音目标识别或声音目标定位.机器学习模型通常采用有监督学习**(Supervised learning)**，需要事先用标注好的已知数据进行训练.本文所述的基于一般音频/环境声的** CA **算法设计与语音信息处理及音乐信息检索**(MIR)**技术高度类似，区别在于声音的本质不同，需要更有针对性的设计各个步骤的算法，另外需要某种特定声音的领域知识.**

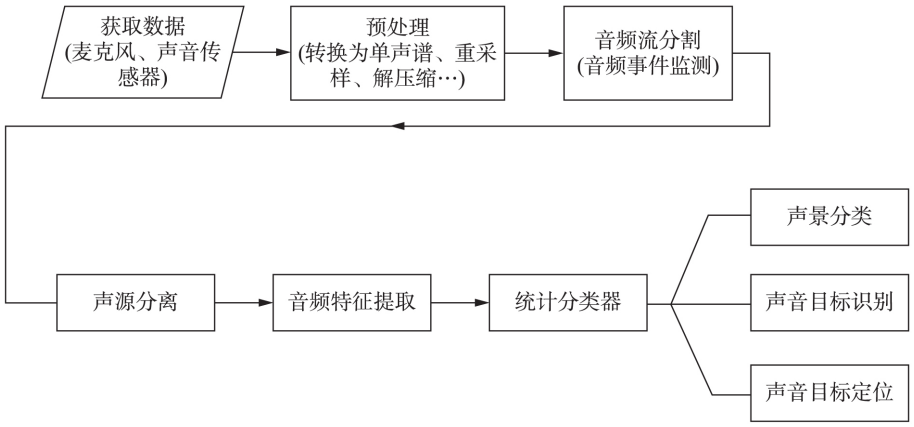

图 4　**计算机听觉技术算法系统的框架图**

Fig.4　A frame diagram of computer audition algorithm system

3.1　音频事件检测

音频事件(Audio event)**指一段具有特定意义的连续声音，时间可长可短，例如笑声、鼓掌声、枪声、犬吠、警笛声等，也可称为音频镜头**(Audio shot)**.音频事件检测**(AED)**，亦称声音事件检测**(Sound Event Detection, SED)**、环境声音识别**(Environmental Sound Recognition, ESR)**，旨在识别音频流中事件的起止时间**(Event onsets and offsets)**和类型**[1-2]**，有时还包括其重要性**(Saliency)**[2]**.面向实际系统的** AED **需要在各种背景声音的干扰下，在连续音频流中找到声音事件的边界再进行分类，比单纯的分类问题要更困难**[3]**.虽然声音识别的研究在传统上侧重于语音和音乐信号，但面向一般音频/环境声的声音识别问题早在** 1999 **年即已开始**[4]**，而且近年来得到了越来越多的关注**[5]**.AED 应用范围广泛，典型的如多媒体分析，对人类甚至动物生活的监控，枪声识别**(Gunshot recognition)**[6]**，声音监控**(Acoustic surveillance)**和智能家居**(Smart home automation)**[7]**、犯罪调查等安全系统**[8]**，行车环境的音频监控**[9]**，推断人类活动和位置**[10]**等.**

环境声音是非结构化的(Unstructured),类似于噪声[8].麦克风是最常见的声音采集设备,从单麦克风[11]到双麦克风[7]甚至 4 个麦克风[6].声源往往来自不同声学环境下的未知距离,混有噪声,并且是混响(Reverberant).例如,在家庭环境的噪声中,最难处理的是非平稳干扰如电视、收音机或音乐 TV[7].物联网(Internet of Things, IoT)平台有大量的分布式麦克风可用,能够将来自多个传感器的信息进行融合,从而使各麦克风组成多麦克风系统,可提高 AED 系统的识别精度[12].一个很具有挑战性的任务是从单通道(Single channel)音频中同时识别出重叠的音频事件(Overlapping sound events)[13].

传统的基于帧(Frame-based)的方法不太适合环境声音识别,因为每个时间帧都混合了来自多个声源的信息[13].基于声音场景或事件(Acoustic scenes or events)分割更适合于识别.场景具有明确的语义,适用于预先知道目标类别的应用.事件适用于监督程度较低的情况,通常在基本音频流分割单元上聚类得到[2,14].文献[14]使用基于经验模式分解(Empirical Mode Decomposition, EMD)产生的第 1 到第 6 个本征模态函数(Intrinsic Mode Functions, IMF)的投票(Voting)方法来检测音频事件的端点,进行盲分割.环境声音在日常生活中经常重复,音频分割的一个特例就是环境声音的重复识别(Repeat recognition),对于这些声音的紧致表示(Compact representation)和预测至关重要.文献[15]根据能量包络的形状将输入的环境声信号分成几个单元,计算每对单元之间的听觉距离(Auditory distance),然后利用近似匹配算法(Approximate matching algorithm)检测重复的部分.

在实际情况下,各种干扰噪声和背景声音与感兴趣的音频事件同时存在,滤波等传统降噪方法完全无效[16].文献[17]采用概率潜在成分分析(Probabilistic Latent Component Analysis, PLCA)进行噪声分离(Noise separation).为了减轻声源分离引入的人工痕迹(Artifacts),应用一系列频谱加权(Spectral weightings)技术来提高声谱(Audio spectra)的可靠性.文献[7,16]使用一种新型的基于回归的噪声消除(Regression-based Noise Cancellation, RNC)技术以减少干扰.对于残留噪声,采用频带功率分布的图像特征(Subband Power Distribution-Image Feature, SPD-IF)增强框架,将噪声和信号定位到不同的区域.然后对可靠部分进行缺失特征分类,利用频带上的时间信息来估计频带功率分布.

在非平稳(Non-stationary)环境中,T-F 表示是一种强大的分析工具,可进行信号的分类或检测[18].常见的如 Gabor 变换[19],EMD[14]等.EMD 将信号表示为一组 IMFs,然后将这些 IMFs 的动态表示为线性动态系统(Linear dynamical system),采用线性和非线性技术来学习系统动态,可以区分不同类别的声音纹理(Sound textures)[20].非线性时序分析技术在处理环境声音方面具有较大潜力[21].

音频特征影响 AED 系统的性能[22].最近的研究集中在非平稳特性的新特征,力求将与信号的时间和频谱特征有关的信息(Temporal and spectral characteristics)内容最大化[5].使用过的音频特征有 MFCC[10,23-26]及其变种 Binaural MFCC[23]、log MFCC[23]、小波(Wavelet)系数[24]、使用 OpenSMILE 提取的两个不同的大规模时间池特征(Large-scale temporal pooling features)[23]、mile983(983 维)、Smile6k(6 573 维)[25]、线性预测系数(Linear Prediction Coefficient, LPC)、匹配追踪(Matching Pursuit, MP)[8]、伽玛通倒谱系数(Gammatone Cepstral Coefficients, GCC)[27]、降维对数谱特征(Log-spectral features)[28]、STE[26]、SE[26]、ZCR[26]、SC[26]、SBW[26]、f_0[26]、为结合 CNN 使用的低级空间特征(Low-level spatial features)[29]、频谱图(Spectrogram)[25]等.文献[30]认为背景声比前景声更具鲁棒性,在复杂的声音环境中可以从背景声中提取音频特征.文献[16]提出一种基于类补偿(Class-Based Compensation, CBC)的方法,基本思想是为分类器的每一个类学习一组过滤器,将较高的权重分配给最能区分类信息的频率成分,以增强特征的区分能力.

与以上声音特征不同,从频谱图中提取的声音子空间(Acoustic subspaces)矩阵可以作为识别的基本元素,有效地描述了频谱图的时间-谱模式(Temporal-spectral patterns)[17,19].文献[19]通过从 Gabor 频谱图中提取子空间,进一步对低秩(Low-rank)的突出的(Prominent)T-F 模式进行编码.子空间特征需要通过两步得到:首先,在复杂向量空间中通过目标事件分析建立子空间库(Subspace bank);然后,通过将观测向量(Observation vectors)投影到子空间库上,可以减少噪声效应(Noise effect),生成源自不同事件子空间(Event subspaces)的判别字符(Discriminant characters)[31].

受图像处理技术启发,在 2 维 T-F 频谱图上计算 LBP,提取频谱相关的局部特征,可以更好地描述音

频[32]，而且通常认为局部特性比全局特性更重要[8].文献[33]将本地的统计数据、均值、标准偏差结合在一起，建立了鲁棒的 LBP.文献[13]提出一种基于局部频谱图特征(Local Spectrogram Features，LSF)的方法，找出频谱图中稀疏的、有区分性的峰值作为关键点，在围绕关键点的 2 维区域内提取局部频谱信息.通过一组具有代表性的 LSF 簇(Clusters)和它们在频谱图中的出现时间(Occurrences)来模拟音频事件.

音频片段长度即粒度(Granularity)对分类识别结果有影响.文献[8]使用较长持续时间(6 s)，比使用较短持续时间(1 s)显著提高了分类精度，而没有增加额外开销.较大的训练和标签集也有益于分类任务[34].文献[11]也表明分类准确度受分类粒度的影响.文献[8]研究了关于分类准确性与窗口大小和采样率(Sampling rate)的关系，以找出每个因素的合适的值，还研究了这些因素的所有组合.

在很多的候选特征中需确定最佳特征(Optimal feature)组合并进行特征融合.文献[35]通过因子分析(Factor analysis)研究特征的性能，并确定特征组合.文献[36]利用进化算法(Evolutional algorithm)中的粒子群优化(Particle Swarm Optimization，PSO)算法和遗传算法(Genetic Algorithm，GA)从大量音频特征中选择最重要的声音特征.

选取最佳特征集后，有时还需进行后处理(Post-processing)，增强区分能力和鲁棒性.文献[33]采用 L2-Hellinger 归一化(Normalization)技术.文献[37]在给定的时间窗口中，计算内部所有帧的心理声学(Psychoacoustic)特征，即梅尔和伽玛通频率倒谱系数(Mel and Gammatone-Frequency Cepstral Coefficients，MGFCC).按照学习好的码本(Codebook)将特征量化为音频词袋(Bag of Audio Words，BoAW)，即直方图(Histogram).特征袋方法计算成本低，对于在线处理特别有用.文献[38-41]也采用了类似的音频词袋方法.文献[29]扩展了 CNN，分别学习多通道特征.该网络不是将各个通道的特征连接到一个单独的特征向量中，而是将多声道音频中的音频事件作为单独的卷积层来更好地学习.

音频事件通常发生在非结构化的环境中，频率内容和时间结构都有很大的变化.早期的算法通常基于手工制作(Hand-crafted)特征.随着 DL 的流行，大量基于 DL 的算法被用于自动特征学习.CNN 能够提取反映本质内容的特征，并且对局部频谱和时间变化不敏感[42].文献[43]提出一种使用 CNN 的新型端到端(End-to-end)的 ESC 系统，直接从原始波形(Raw waveforms)中学习特征用于分类.因为缺乏明确的语义单元，对音频事件进行端到端的识别通常需要较长的时间片段，文献[38]引入了具有更大输入域(Input field)的 CNN.文献[22]使用多流分层深度神经网络(Multi-stream Hierarchical Deep Neural Network，MS-H-DNN)提取音频深度特征(Deep feature)，融合了多个输入特性流的潜在互补信息，更具区分性.基于极端学习机的自动编码器(Extreme Learning Machine-based Auto-Encoder，ELM-AE)是一种新的 DL 算法，具有优异的表现性能和快速的训练过程.文献[44]提出一种双线性多列(Bilinear Multi-column ELM-AE，B-MC-ELM-AE)算法，以提高原始 ELM-AE 算法的鲁棒性、稳定性和特征表示能力，学习声信号的特征表示.

简单的音频事件种类识别可采用核 Fisher 判别(Kernel Fisher Discriminant，KFD)分析法[19]，正则化核 Fisher 判别(Regularized KFD)分析法[17]，DTW[24]，矢量量化(Vector Quantization，VQ)[24].但更多的采用统计分类器，如 K 近邻(K-Nearest Neighbors，KNN)[8,36]，GMM[23,25,45]，随机森林(Random Forest，RF)[14]，支持向量机(Support Vector Machine，SVM)[16,25-26]，HMM[28]，人工神经网络(Artificial Neural Network，ANN)[24,46]，DNN[23,25]，RNN[23,25]，CNN[23,25]，RDNN[25]，I-Vector[23]，EC[47]等.文献[46]在相同数据集上对两种不同的神经网络(Neural Network，NN)进行分析，后向传播神经网络(Back-Propagation Neural Network，BPNN)与径向基函数神经网络(Radial-Basis Function Neural Network，RBFNN)相比，识别结果具有显著性和有效性.文献[34]研究了几个深度 NN 架构，包括全连接 DNN(Fully-connected DNN)、CNN-AlexNet、CNN-VGG、CNN-GoogLeNet Incepetion 和 CNN-ResNet，发现 CNN 类网络表现良好.文献[25]全面研究各种统计分类器后，发现深度学习模型与传统浅层模型相比具有一定的优越性，但没有一个模型能在所有数据集上优于所有其他模型，说明模型的性能随着特征的不同而有很大差异.文献[48]的研究也表明，在 AED 任务上，基于 DNN 的系统比使用 GFB 特征与多类 GMM-HMM 相结合的系统识别精度要差.

序列学习(Sequential learning)方法被用来捕捉环境声音的长期变化[5].RNN 擅长学习音频信号的

长时上下文信息,而 CNN 在分类任务上表现良好,文献[42]将这两种方法结合形成 CRNN (Convolutional Recurrent Neural Network),性能在日常复合音频事件(Polyphonic sound event detection)检测任务中有很大的改进.但在文献[23]和[25]的实验中,表现最好的模型是非时态(Non-temporal)DNN,表明 DCASE(IEEE Challenge on Detection and Classification of Acoustic Scenes and Events)挑战中的声音不会表现出强烈的时间动态(Temporal dynamics),这与文献[42]的结论相反.关于时序信息对于音频事件检测的作用还有待进一步研究.

在决策阶段,文献[23]对多个分类器的结果采用后期融合方法(Late-fusion approach).文献[13]使用广义霍夫变换(Generalized Hough Transform, GHT)投票系统,对许多独立的关键点的信息进行汇总,产生起始假设(Onset hypotheses),可以检测到频谱图中任何音频事件的任意组合.对每个假设进行评分,以识别频谱图中的重叠音频事件.

训练统计模型必须具备较大的数据量,完全监督的训练数据需要在一个音频片段中只清楚地包含某个特定的音频事件.所需时间及人力、经济代价巨大,经常还需要各类声音的领域知识.为使收集大量训练声音数据的过程更容易,文献[49]设计了基于游戏的环境声音采集框架"Sonic home".为降低训练数据量的要求,通常使用主动学习(Active learning)或半监督学习(Semi-supervised learning)技术[50].文献[51]提出一种新的主动学习方法.首先在未标记的声音片段上进行 K-medoids 聚类,并将簇的中心点(Medoids)呈现给标注者进行标记,中心点带标注的标签用于派生其他簇成员的预测标签.该方法优于对所有数据进行标注的传统主动学习法如随机抽样(Random sampling)、基于确定性的主动学习(Certainty-based active learning)和半监督学习.在保持相同识别准确率的同时,可节省 50%~60%训练音频事件分类器的标注工作量.文献[52]使用一个基于全卷积神经网络(Fully Convolutional Networks, FCN)的模型,基于弱监督学习(Wakly-supervised learning)识别音频事件,而且能够在只有片段级别(Clip-level)没有帧级别(Frame-level)标注的训练下进行音频事件定位.文献[53]提出一个与文献[52]类似的 FCN 结构,从 YouTube 上的弱标记数据识别音频事件.该网络有 5 个卷积层,后边没有采用最常见的全连接层(Fully connected dense layers),而是采用了另外 2 个卷积层,最后是一个全局最大池化层(Global max-pooling layer),形成了一个全卷积的 CNN 架构.与将时间域信息全部混合起来得到最后结果的全连接架构不同,使用全局最大池化层可以在时间轴上选择最有效的片段输出最后的预测结果.因此,在训练和测试中能有效处理可变长度的输入音频,不需要进行固定分割的前处理过程,可进行粗略的音频事件定位.文献[54]结合带标记的音频训练数据集和互联网上的未标记音频进行自训练(Self-training)来改进声音模型.首先在带标记音频上训练,然后在 YouTube 下载的音频上测试.当检测器以较高的置信度识别出任何已知的声音事件时,就把这个未标记的音频加入到训练集进行重新训练.

弥补目标域(Target domain)训练样本的不足还可以采用迁移学习(Transfer learning),调用在其他具有类似特点的大型数据库已预先训练好的模型[55].该技术旨在将数据和知识从源域(Source domain)转移到目标域,即使源和目标具有不同的特性分布和标签集[56].基于 DNN 的迁移学习已经被证明在视觉对象分类(Visual Object Classification, VOC)中是有效的,文献[55]利用 VOC-DNN 在其训练环境之外的学习能力,迁移到 AED 领域.文献[56]假设所有的音频事件都有相同的基本声音构件(Basic acoustic building blocks)集合,只是在这些声音构件的时间顺序上存在差异.构造一个 DNN,它具有一个卷积层来提取声音构件,和一个递归层(Recurrent layer)来捕获时间顺序(Temporal order).在上述假设下,通过将卷积层从源域(合成源数据库)转移到目标域(DCASE 2016 的目标数据库),实现从源域转换到具有不同声音构件及顺序的目标域的迁移学习.注意,递归层是直接从目标域学习的,无法通过转移来检测与源领域中声音构件不同的事件.

训练数据的多样性对于防止过拟合(Overfitting),获得鲁棒的模型具有关键作用.文献[38]提出一种新的数据增强(Data augmentation)方法来引入数据变化,以充分利用 CNN 网络的建模能力.文献[57]在训练过程中使用模拟仿真,将目标声音(Target sounds)与各种环境声音按照不同的角度配置(Angular source configuration)和信噪比(Signal-to-Noise Ratio, SNR)叠加在一起,增强其泛化性能,称为多条件训练(Multi-conditional training).

环境声的种类无法尽数，在研究中只能选择个别类型作为例子.文献[47]使用了两个基准数据集：RWCP(Real World Computing Partnership)数据库和 Sound Dataset.文献[23]使用了最大的数据集之一——DCASE 2016，将声音分类为 15 种常见的室内和室外声音场景，如公共汽车(Bus)、咖啡馆(Cafe)、汽车(Car)、市中心(City center)、森林道路(Forest path)、图书馆(Library)、火车(Train)等，共 13 h 的立体声录音.文献[26]将环境声分为 6 类，即车鸣声、钟声、风声、冰块声、机床声、雨声.文献[28]包含男性演讲(Male speech)、女性演讲(Female speech)、音乐(Music)、动物声音(Animal sounds)等.文献[6]则专门识别燃放鞭炮(Firecracker)、9 mm 和 44 mm 口径发令枪(Starter pistol)、爆炸(Explosion)、射击(Firing)等冲击型声音.文献[36]将声音分为 6 类：语音(Speech)、音乐(Music)、噪声(Noise)、掌声(Applause)、笑声(Laughing)、哭声(Crying).文献[20]录制 5 种声音组成了一个数据集，包括噼啪的火焰声(Crackling fire)、打字声(Typewriter action)、暴雨声(Rainstorms)、碳酸饮料声(Carbonated beverages)和观众的掌声(Crowd applause).网络视频提供了一个几乎无限的音频来源，文献[58]在 100 万部 YouTube 视频中提取 45 kh 的音频，构成一个多样化语料库.文献[59]建立的 ESRD03 数据库从 21 张音效 CD 和 RWCP 数据库中收集数据，包括 16 000 多个音轨，大部分发生在家庭环境中.

AED 还可用于自动和快速标记音频记录(Audio tagging).这是一项具有挑战性的任务，音频事件变化无穷，对应的标签数量众多，不同的标注者可能提供不完整或不明确的标签.为了处理这些问题，文献[60]使用一个共同正则化(Co-regularization)方法来学习一对声音和文本上的分类器.第一个分类器将低级音频特性映射到真正的标签列表，第二个分类器将损坏的标签映射到真正的标签，减少了由第一个分类器中的低级声学变化引起的不正确映射，并用额外的相关标签进行扩充.音频信息还可以辅助进行视频事件检测(Video Event Detection，VED).文献[61]提出一种音频算法，基于 STE、ZCR、MFCC、基于统计特性的改进特征、HMM，对视频中的尖叫片段进行检测.

3.2　音频场景识别

音频场景(Audio scenes)是一个保持语义相关或一致性(Semantic consistant)的声音片段，通常由多个音频事件组成.例如，一段包含枪声、炮声、呐喊声、爆炸声等声音事件的音频很可能对应一个战争场景.对于实际应用中的连续音频流，音频场景识别(Audio Scene Recognition，ASR)首先进行时间轴语义分割，得到音频场景的起止时间即边界(Audio scene cut)，再进行音频场景分类(Audio Scene Classification，ASC).ASR 是提取音频结构和内容语义的重要手段，是基于内容的音频、视频检索和分析的基础[26,62].目前场景检测(Scene detection)的研究主要基于图像和视频.音频同样具有丰富的场景信息，基于音频既可独立进行场景分析，也可以辅助视频场景分析，以获得更为准确的场景检测和分割.音频场景的类别并没有固定的定义，依赖于具体应用场景.在电影等视频中，可粗略分为语音、音乐、歌曲、环境音、带音乐伴奏的语音等几类[62].环境音还可以进行更细粒度的划分.基于音频分析的方法用户容易接受，计算量也比较少[63-64].

音频场景由主要的几个声源所刻画.换句话说，音频场景可以定义为一个包含多个声源的集合[65].当大多数声源变化时，就会发生场景变化.基于一个模拟人类听觉的具有时间两个参数(Attention-span 和 Memory)的模型[66]，文献[65]逐块提取能量、过零率、谱特征、倒谱特征等多个音频特征，对每个特征拟合最佳包络线，通过计算包络线之间的相关度，基于阈值进行边界分割.参数 Attention-span 增加时性能提升.文献[67]假设大多数广播包含语音、音乐、掌声、欢呼声等声音类别，将每秒音频包含的分类构成直方图形式的纹理(Texture)表示，基于纹理的变化进行场景变化检测.文献[68]首先使用模糊 C 均值聚类(Fuzzy C-means)算法检测 Audio shot cuts，之后计算音频镜头之间的语义相关性，语义相关的音频镜头被合并为音频场景.文献[69]基于音频事件进行音频场景检测，符合人类的思维习惯.与文本信息检索中的罕见词和常见词类似，给更能反映音频内容主题(Topic)的音频事件赋予更大的权重，而给在多个主题中出现的常见音频事件赋予较小的权重，会有助于音频场景的检测.

声音特征的确定是音频场景自动识别中的一个重要问题，提取正确的特性集是获得系统高性能的关键.设计选择音频特征与对应的音频场景有很强的相关性.例如，在文献[70]的水声、风声、鸟叫声、城市声音等 4 种类型的声音中，一般来说，水和风的声音都有较低的音高值和音高强度；鸟叫声有很高的音高值

和音高强度;城市的声音有很低的音高值和相对广泛的音高强度.

人们已经提出了各种各样的音频特征,但过去的绝大多数工作都利用结构化数据(如语音和音乐)的特性,并假定这种关联会自然地传递到非结构化的声音[71].ASR 使用的特征有 MFCC[25-26,53,72],短时能量(Short-Time Energy, STE)[26],频带能量(Subband Energy, SE)[26],ZCR[26],f_0[26],SC[26,72],频谱带宽(Spectral Band Width, SBW)[72],MPEG-7 特征[26,39,73],基于幅度调制滤波器组(Amplitude modulation filterbank)与 Gabor 滤波器组(Gabor Filterbank, GFB)的特征[48].文献[70]使用音高特征(Pitch features),包括音高值、音高强度、可听音高随时间变化的百分比.文献[74]通过线性正交变换的主成分分析(Principal Component Analysis, PCA)将多通道观测幅度的对数转换为特征向量.文献[71]基于匹配追踪进行环境声音的特征提取.利用字典来选择特征,得到灵活、直观、物理可解释的表示形式,对噪声的敏感度较低,能够有效地代表来自不同声源和不同频率范围的声音.通常特征向量只描述单个帧(Frame)的信息,但与时间动态(Temporal dynamics)相关的局部特征会有益于环境声信号的分析.文献[72]将帧级的 MFCC 特征视为 2 维图像,采用局部二进制模式(Local Binary Pattern, LBP)来描述时间动态的隐藏(Latent)信息,并使用 LBP 对演化(Evolution)过程进行编码.由于音频场景有丰富的内容,多个特征的组合将是获得良好性能的关键.

与传统的手工特征相比,矩阵分解(Matrix factorization)类的非监督学习方法包括稀疏性(Sparsity)、基于内核(Kernel-based)、卷积(Convolutive)、PCA 的新方法,可以自动从 T-F 表示中学习场景的更好表示[75].文献[76]通过有监督的非负矩阵分解(Supervised Non-negative Matrix Factorization, NMF)进行矩阵分解,研究了使用监督特征学习方法从声场记录中提取具有相关性和区分性(Relevant and discriminative)特征的方法.文献[77]使用卷积神经网络(Convolutional Neural Networks, CNN)作为特征提取器,从标签树嵌入图像(Label-tree embedding image)中自动学习对分类任务有用的特征模板.文献[74]通过 PCA 得到的线性正交变换将多通道观测幅度的对数转换为特征向量.

ASR 使用的模型包括高斯混合模型(Gaussian Mixture Model, GMM)[25,48],隐马尔可夫模型(Hidden Markov Model, HMM)[48],SVM[25-26,78-79],I-Vector[53],集成分类器(Ensemble Classifier, EC)[72],深度神经网路(Deep Neural Network, DNN)[25,48]、递归神经网络(Recurrent Neural Network, RNN)[25,48]、递归深度神经网络(Recurrent Deep Neural Network, RDNN)[25]、CNN[25]等.文献[48]采用能够像 RNN 一样分析长期上下文信息(Long contextual information),且训练代价与传统 DNN 类似的时延神经网络(Time-Delay Neural Network, TDNN)系统.

声音与视觉信息互为补充是人类感知环境的重要方式[25].音频场景分析被大量用于辅助视频场景分析、检测和分割,提高对视频内容的识别准确率,解决诸如图像变化而实际场景并未变化的困难,且整体运算复杂度更低[64].音频场景分析可应用于视频内容监控及特定视频片段的检索与分割[78],即使在视频数据丢失的情况下,也能检测到目标声源的活动[80].文献[81]使用声音识别广播新闻中说话人的变化位置,定位每一个主题的开始,实现快速自动浏览.文献[82]结合音、视频特点,对足球视频进行基于进球语义事件的检索,满足观众的个性化检索要求.为满足网络视频的监管需求,文献[39]提取音频流的 MPEG-7 低层(SC、SBW)和高层音频特征(音频签名),采用独特的权重分配机制形成音频词袋特征,输入 SVM 对暴力和非暴力视频进行分类.文献[40]结合视频静图特征、运动特征以及声音特征,建立一个多模态色情视频检测算法.文献[79]首先用两层(粗/细)SVM 识别爆炸/类似爆炸的音频区间,得到爆炸的备选场景.对这些备选场景再判断其对应的视觉特征是否发生剧烈突变,得到最后的识别结果.

4 各领域基于一般音频/环境声的计算机听觉算法概述

如前所述,CA 是一个运用音频信号处理、机器学习等方法对数字音频和音乐进行内容分析理解的学科.其中音乐部分的技术综述参见文献[83],本文面向一般音频/环境声,以国民经济行业分类国家标准[84]中的各个领域为主线,总结已有的 CA 技术的典型算法.

4.1 医疗卫生

人的身体本身和许多疾病,都会产生各种各样的声音.借助 CA 进行辅助诊断与治疗,既可部分减轻

医生的负担，又可普惠广大消费者，是智慧医疗的重要方面.

4.1.1　呼吸系统疾病

常见的与病人呼吸系统相关的音频事件有咳嗽、打鼾、言语、喘息、呼吸等.监控病人状态，在发生特定音频事件时触发警报以提醒护士或家人具有重要意义[85].听诊器是诊断呼吸系统疾病的常规设备，文献[86]研制光电型智能听诊器，能存储和回放声音，显示声音波形并比对，同时对声音进行智能分析，给医生诊断提供参考.

咳嗽(Cough)是人体的一种应激性的反射保护机制，可以有效清除位于呼吸系统内的异物.但是，频繁、剧烈和持久的咳嗽也会给人体造成伤害，是呼吸系统疾病(Respiratory disease)的常见症状.不同呼吸疾病可能具有不同的咳嗽特征.目前对咳嗽的判断主要依靠病人的主观描述，医生的人工评估过程繁琐、主观，不适合长期记录，还有传染危险.鉴于主观判断的不足，研究客观测量及定量评估咳嗽频率(Cough frequency)、强度(Cough intensity)等特性的咳嗽音自动识别与分析系统，为临床诊断提供信息，就非常必要[87-88].有时还需要专门针对儿科人群(Pediatric population)的技术[89].

文献[90]通过临床实验测试了人类根据听觉和视觉来识别和计算咳嗽的准确性，还评估了一个全自动咳嗽监视器(Pulmotrack).被试依靠听觉可以很好地识别咳嗽，视觉数据对于咳嗽计数也有显著影响.虽然 Pulmotrack 自动测试的咳嗽频率和人类结果有较大差距，但文献[91]研发的基于音频的自动咳嗽检测(Audio-based automatic cough detection)优于使用 4 个传感器的商用系统，说明了这种技术具有一定的可行性.

从含有背景噪声的音频流中识别咳嗽音频事件(Cough events)的技术框架与上述 AED 相同，只是集中于识别分类为咳嗽声的音频片段.最简单的端点检测是分帧[92]，并对疑似咳嗽的片段进行初步筛选.文献[88]和[93]基于 STE 和 ZCR 的双门限检测算法对咳嗽信号进行端点检测.文献[88]研究了基于 WT 的含噪咳嗽信号降噪方法，通过实验确定小波函数和分解层数、阈值等.在已有工作中，几乎所有的咳嗽声音特征提取方法都来自语音或音乐领域，如 LPC[88]，MFCC[88,92-93]，香农熵(Shannon entropy)[89]，倒谱系数(Cepstral coefficients)[89]，线性预测倒谱系数(Linear Predictive Cepstral Coefficient，LPCC)[88]，结合 WPT 和 MFCC 的 WPT-MFCC 特征[88]等.从咳嗽的生理学特性和声学特点可知，咳嗽声属于典型的非平稳信号，具有突发性.在咳嗽频谱(Cough spectrum)中能量是高度分散的，与语音和音乐信号明显不同.为提取更符合咳嗽的声音特性，文献[87]基于 Gammatone 滤波器组在部分频带提取音频特征.在咳嗽声分类识别阶段，文献[92]使用 DTW 将咳嗽疑似帧的 MFCC 特征和模板库进行基于距离的匹配.文献[87]使用 SVM、KNN 和 RF 分别训练和测试，集成各种输出做出最终决策.文献[92]使用 ANN，文献[93]使用 HMM，文献[88]使用 GMM 对咳嗽片段进行分类.在咳嗽声录音里经常出现的声音种类一般还有说话声、笑声、清喉音、音乐声等[88].

在 CA 的医学应用领域，目前各项研究都是用自行搜集的临床数据.文献[87]收集了 18 个呼吸系统疾病患者的真实数据，并由人类专家进行了标注.文献[89]搜集了 14 个受试者的数据，录音长度 840 min.在识别咳嗽音频事件的基础上，如果集成更多咳嗽方面的专家知识，可以更精确地帮助提高疾病类型临床诊断的精确度[92].

肺的状况直接影响肺音(Lung sound).肺音包含丰富的肺生理(Physiological)和病理(Pathological)信息，在听诊(Auscultation)过程中对肺部噪声振动频率(Lung noise vibration frequency)、声波振幅(Amplitude)和振幅波动梯度(Amplitude fluctuation gradient)等特征进行分析来判断病因.研究尘肺患者肺部声音的改变，可以探索听声辨病的可行性[94].文献[95]对 30 多份相同类型的肺音进行小波分解，每个频带小波系数加权优化后，通过 BPNN 对大型、中型和小型湿罗音(Wet rale)和喘息声(Wheezing sound)进行分类识别.文献[96]采集肺音信号，使用 WT 滤波抑制噪声获得更纯净的肺音，然后使用 WT 进行分析，将肺音信号分解为 7 层，并从频带中提取一组统计特征输入 BPNN，分类识别为正常和肺炎两种结果.

阻塞性睡眠呼吸暂停(Obstructive Sleep Apnea，OSA)是一种常见的睡眠障碍，伴随打鼾，在睡眠时上呼吸道(Upper airway)有反复的阻塞，发生在夜间不易被发现，对人身健康造成极大的危害，对其进行

预防与诊断十分重要.此疾病监测要对患者的身体安装许多附件来追踪呼吸和生理变化,让患者感到不适,并影响睡眠.目前使用的诊断设备-多导睡眠仪需要患者整夜待在睡眠实验室,连接大量的生理电极,无法普及到家庭.鼾声信号的声音分析方法具有非侵入式、廉价易用的特点,在诊断 OSA 上表现出极大的潜力.

鼾声信号采集通常使用放于枕头两端的声音传感器[97].整夜鼾声音频记录持续时间较长,而且伴有其他非鼾声信号.首先需进行端点检测,如文献[98]采用集成经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)算法,文献[99]采用更加适合鼾声这种非线性、非平稳声信号的自适应纵向盒算法,文献[100]采用基于 STE、ZCR 的时域自相关算法.文献[101]通过整夜鼾声声压级(响度)、鼾声暂停间隔等特征,得到区分单纯鼾症(Simple Snoring, SS)与 OSA 患者的简便筛查方法.文献[100]通过数字滤波器、快速傅里叶变换(Fast Fourier Transform, FFT)、线性预测分析等技术提取呼吸音相关特征,并用 DTW 算法进行匹配识别.文献[102]采用由 f_0、SC、谱扩散(Spectral spread)、谱平坦度(Spectral flatness)组成的对噪声具有一定鲁棒性的特征集,以及 SVM 分类器,对笑声、尖叫声(Scream)、打喷嚏(Sneeze)和鼾声进行分类,并进一步对鼾声和 OSA 分类识别.文献[98]采用类似方法,提取共振峰频率(Formant Frequency,FF)、MFCC 和新提出的基频能量比(f_0 energy ratio)特征,经 SVM 训练后可有效区分出 OSA 与单纯打鼾者.而且将呼吸、血氧信号与鼾声信号相结合,优势互补,提高了整个系统的筛查能力.文献[103]使用相机记录患者的视频和音频,并提取与 OSA 相关联的特征.进行视频时间域降噪后,跟踪患者的胸部和腹部运动.从视频和音频中分别提取特征,用于分类器训练和呼吸事件检测.文献[99]提取能够描述打鼾时声道特性的特征(即共振峰)后进行 K-means 聚类,将音频事件中的鼾声检测出来.

4.1.2 心脏系统疾病

心音信号(Heart Sounds, HS)是人体内一种能够反映心脏及心血管系统运行状况的重要生理信号.对心音信号进行检测分析,能够实现多种心脏疾病的预警和早期诊断.针对心音的分析研究已从传统的人工听诊定性分析,发展到对 T-F 特征的定量分析.

真实心脏声信号的录制可使用电子听诊器[104],或布置于人体心脏外胸腔表面的声音传感器[105].胎儿的心音可通过超声多普勒终端检测后经音频接口转换为声信号[106].利用心音信号的周期性和生理特征可对心音信号进行自动分段[107].

心音信号非常复杂且不稳定.在采集过程中,不可避免地会受到噪声和其他器官活动声音(如肺音等)的干扰,在 T-F 域上存在非线性混叠.文献[108]对原始心音信号通过 WT 进行降噪处理.文献[109]使用针对非平稳信号的 EMD 方法初步分离心音.为解决模态混叠问题,又对 EMD 获得的 IMFs 分量进行奇异值分解(Singular Value Decomposition, SVD).对各个特征分量进行筛选重构后,获得较为清晰的心音信号,优于传统的小波阈值消噪等方法.

心音信号检测使用的 T-F 表示包括 STFT、Wigner 分布(Wigner Distribution, WD)和 WT[110].使用的特征主要是第一心音(S1)和第二心音(S2)的共振峰频率 FF[104,108]、从功率谱分布中提取的特征[111]、心电图(Electrocardiograph,ECG)等辅助数据特征[112].S1 和 S2 具有重要的区分特性.实验表明,只依靠 S1 和 S2 这两个声音特征,无需参考 ECG,也不需要结合 S1 和 S2 的单个持续时间或 S1-S2 和 S2-S1 的时间间隔,即可得到好的识别结果[104].

心音信号检测使用的统计分类器有 SVM[108]、全贝叶斯神经网络模型(Full Bayesian Neural Network Model, FBNNM)[111]、DNN[104]、小波神经网络(Wavelet Neural Network, WNN)[113]等.文献[111]定义了 8 种不同类型的心音.由于临床采集困难,目前研究中心音数据量都不大.文献[111]中有 64 个样本,文献[107]有 48 例心音(异常 10 例),每例提取 2 个时长 5 s 的样本,共 96 个样本.

4.1.3 其他相关医疗

文献[114]使用自相关法提取嗓音的 f_0 特征,用 SVM 进行分类识别,区分病态嗓音和正常嗓音,完成对嗓音疾病的早期诊断.文献[115]采集胎音和胎动信号,获得胎音信号最强的位置,即胎儿心脏的位置,以此判断出胎儿头部位置和胎儿的体位姿态.文献[116]检测片剂、丸剂或胶囊暴露于肠胃系统时所产生的声波,以确定该人已经吞服了所述片剂、丸剂或胶囊.文献[117]使用 X 射线图像确定血液速度的空

间分布，根据速度分布人工合成可视谱所定义的声音.该方法允许心脏病学家和神经科学者以增强的方式分析血管，对脉管病变进行估计，并对血流质量进行更好的控制.肌音信号(Mechanomyographic，MMG)是人体发生动作时由于肌肉收缩所产生的声信号，蕴含了丰富的能够反映人体肢体运动状态的肌肉活动信息.文献[118]通过肌音传感器采集人体前臂特定肌肉的声信号，基于模式分类开发相应的假肢手控制系统.

4.2　安全保护

安全保护经常采用智能监控方式，按照地点可分为公共场所监控和私密场所监控两种.公共场所包括公园、车站、广场、商场、街道、学校、电影院、剧场等地点，经常人员密集，对其进行有效的安防智能监控来维护社会安全是最主要的应用.目前公共场所的监控系统主要都基于视频，但是视线被遮挡时存在盲区，而且容易受到光线、恶劣天气等因素的影响.异常事件通常会伴随异常声音的发生，异常声音本身即能有效地反应重大事故和危急情况的发生，且具有复杂度低、易获取、不受空间限制等优势[119-120].一个完整的公共场所智能监控系统应当充分利用场景中视听觉信息的相关性，将其有机地融合到一起[121].例如，文献[122]采集 ATM 机监控区域内的声信号，提取特征后判断是否为异常声音，与视频监控相结合可以解决 ATM 机暴力犯罪的问题.私密场所主要包括家庭、宿舍、医院病房、浴室、KTV 包房、军事基地等地点，由于或多或少的隐私性及保密性，不方便采用可能暴露被监护人隐私的视频监控，采用基于 AED 的音频监控更为合适[123-124].典型的应用包括老年人、残疾人、婴儿和儿童的家庭日常生活监控，病人的医疗监控及辅助护理，浴室、学生寝室等私密性公共场所的安全监控等[125-127].与已有的基于穿戴式设备的个体监护技术相比，音频监控受到的限制较小，成本也降低很多[128].

对公共场所及私密场所进行音频监控的技术框架相同，区别在于可能发生的异常声音种类不同.异常声音是指正常声音比如开门声、关门声、电话铃声、脚步声、谈话声、音乐声、车辆行驶声等之外的在特殊情况下才发出的声音.文献中研究较多的公共场合异常声音种类通常有枪声[129-132]、爆炸声[133-134]、玻璃破碎声[134]、乱扔垃圾声[135]等，私密场合研究较多的异常声音种类通常有摔门声[131]、跑步声[131,136]、玻璃破碎声[131,133]、人的尖叫声[131,133]、婴儿或小孩的哭声[133,137]、老人摔倒声[136,138-139]、呼救声[136]、漏水声[140]等.注意这种划分并不是绝对的，只是按照发生的可能性进行的粗略分类，有时也会交叉.比如人的尖叫声除了可能发生在家庭吵架场合，也会发生在广场恐怖事件这样比较少数的场合.音频监控系统主要基于软硬件的系统集成.文献[141]在智能家居领域发明了一种具有声音监听功能的智能电视，智能电视和声音监听模块通过无线通信连接.当声音监听模块监听到特定的声音或者音量超限时，智能电视会自动调成静音.

在已有的音频监控文献中，采集声音数据通常使用麦克风[136]或麦克风阵列(Microphone array)[138].文献[131]构建了一个大约 1 000 个声音片段的音频事件数据集和一个监视系统的真实情况数据集.文献[136]模拟了一个包含 105 个设计场景、21 个音频事件的音频事件数据库.

文献[133]使用 MFCC 的第 1 维系数改进声音活动检测算法，确定异常声音的端点.文献[142]针对公共场所异常声音的特点，提出一种综合短时优化 ZCR 和短时对数能量的自适应异常声音端点检测方法.文献[134]通过 WT 分析信号的高频特性，采用基于能量变化的算法检测异常声音片段.文献[119]基于 STE 时间阈值进行音频事件端点检测.文献[120]则另辟蹊径，首先用基于单类 SVM 的异常声音检测算法进行粗分类，根据 MFCC、STE、SC、短时平均 ZCR 等特征判断每一帧声音是否异常.当窗长 2 s 的滑动窗内有连续多个帧出现异常时，则判定这一段声音为异常声音.通过对各段声音进行中值滤波(Median filtering)平滑后得到音频事件的分割，从而直接省去端点检测的步骤.文献[143]使用了小波降噪方法进行信号提纯.

音频监控使用的音频特征包括 STE[129,143]、ZCR[144]、短时平均 ZCR[129]、SC[144]、滚降点(Roll-off point)[144]、MFCC[123,129,134,136-137,139,143-144]、ΔMFCC[134,136,143]、ΔΔMFCC[136]、Teager 能量算子[133]、感知特征(Perceptual features)[135]、MPEG-7 特征[145-146]等.考虑到异常声信号具有非平稳、突发性等特点，文献[120]将信号通过 EEMD 处理获得不同层的 IMF，对每一层的 IMF 提取 MFCC 等特征，并使用特征组合成最终称为 EEMD-MFCC 的特征矢量，识别效果比 MFCC 有明显提升.文献[41]在提取音频特征后不立

即进行分类,而是先送入概率潜在语义分析模型(Probabilistic Latent Semantic Analysis, PLSA),通过训练获取声音主题词袋模型,降低音频信号特征矩阵的维数[41].文献[128]认为特征融合很重要.文献[131]研究了不同的帧大小对音频特征提取的影响,结果表明不同的音频帧大小会引起分类精度变化.整合多帧特征生成一个新的特征集,可以实现更好的性能.

音频监控使用的音频事件匹配识别算法有模板匹配法[126]、DTW[129,137]、动态规划(Dynamic Programming, DP)[139].使用过的统计分类器包括 SVM[145]、KNN[41]、GMM[143-144]、HMM[123,133-134]、适合处理时间序列数据的脉冲神经网络(Pulsed Neural Networks, PulsedNN)[147]、层次结构神经网络(Hierarchical Structure Neural Network, HSNN)[148]、条件随机场(Conditional Random Field, CRF)[127]、基于模糊规则的单类分类器(Fuzzy rule-based one-class classifiers)[135]等.通常系统会根据音频事件的种类数量训练相同数量的模型,如文献[136]训练了与其音频事件数据库对应的 21 个 HMM.大多数异常声音监控系统采用直接识别法,只适用于少量异常声音种类的检测,当检测种类上升时效果变差[120].通过增加训练文件的数量和减少每个训练文件中样本的数量,可以获得更高的识别准确率[6].机器学习并不是识别音频事件的唯一办法,文献[140]研究了一种基于气泡声学物理模型的识别系统,不需要训练.

4.3 交通运输、仓储

CA 在交通运输、仓储业具有多个应用.例如,CA 可自动进行车辆检测、车型识别、车速判断、收费、交通事故认定、刹车片材质好坏识别、飞行数据分析等,对于水、陆、空智能交通都具有重要意义[149-151].

4.3.1 铁路运输业

文献[152]发明一种地铁故障检测装置,用麦克风检测列车发出的声信号并转换为电信号.若电信号的幅值变量与基准幅值变量相同,则继续检测;若不相同,则触发报警模块,记录当前时刻,并显示列车故障点的位置.

4.3.2 道路运输业

4.3.2.1 车型及车距识别

车型自动识别广泛应用于收费系统、交通数据统计等相关工作中.传统方法是在公路上埋设电缆线及感应线圈,通过摄像头抓拍进入视线的车辆照片进行车型识别.此外,还有超声波检测法、微波检测法、红外线检测法等.但对路段有破坏性,设备后期维护要求高,受雨雾等天气状况影响大,不适合沿道路大量铺设[149].基于音频信号的识别技术具有非接触性、维护简单、价格低等特点,在很大程度上弥补传统车辆检测设备易损坏、破坏路面、受环境影响明显、价格昂贵等不足,具有非常重要的现实意义[150].

早在 1998 年,文献[153]就提出一种根据物体发出的声音来对军用车辆进行分类的统计方法.文献[149]基于车辆声信号进行车型识别.文献[154]提出一种基于声音特征的运动车辆类型(Vehicle types)和距离的简单分类算法,对行驶车辆的接近程度进行识别,帮助不能听到车辆从背后接近的听障(Hearing impaired)人士降低户外行动的危险.记录车辆在不同环境条件和不同车速下的声音以及对应的车辆类型和距离作为训练数据.文献[155]的算法可以识别车辆类型.文献[156]将车辆与人的距离分为接近(Approaching)、通过(Passing)和远离(Receding)3 类,通过对道路行驶车辆在不同阶段感知到的噪声差异进行识别.为了防止碰撞,文献[157]研发了一种根据车辆轮胎发出的声音来识别接近车辆(Approaching vehicle)的方案.

车型识别的 CA 技术框架基本一致,只是对应的各种声音来源及种类有所不同.文献[158]选用了驻极体麦克风和 AD7606 数据采集模块,采集了东风农用三轮车和大众 Sagitar 1.4T 轿车的通过噪声.文献[159]使用 DARPA SensIT 实验中的真实数据,其中包含了履带车和重型卡车的大量声信号.文献[157]使用测量车上的一对麦克风来检测接近的车辆.文献[160]使用声音传感器,采集多条车道上行驶车辆的混叠声信号.

行驶车辆的声音可能会受到环境噪声(Ambient noises)和人所在车辆发出声音的影响.文献[157]利用多对麦克风的谱减技术(Spectral subtraction)来降低发动机、冷却风扇以及其他环境噪声的影响.盲信号分离或盲源分离(Blind Source Separation, BSS)在未知源信号与混合系统参数的情况下,仅由传感器

搜集的观测信号估计出源信号.文献[160]通过盲源分离模型估计信号分量个数及瞬时幅度，将单个车辆信号从混合信号中分离出来.文献[150]采用 MP 稀疏分解方法，用 Gabor 原子进行信号的分解及重构，重构后的信号能较好地反映原信号的特征.文献[150]认为发动机声信号相对平稳，信号分解后频域相对稳定，采用单帧进行识别可满足实时性要求.文献[161]采用 200 ms 的较长时间帧来计算频谱.

使用的音频特征有自回归(Autoregressive)[154]、STE[149]、ZCR[149]、基频周期[149]、MFCC[161]、基于听觉 Gammatone 滤波器的频谱特征[162]、使用 WPT 提取的 16 维信号特征[159]等.文献[160]利用 HHT 抽取信号分量的时域包络线，并提取特征向量.文献[155]使用零均值调整样本的协方差矩阵的均值向量和最重要主成分特征向量，来共同表征其声音特征.文献[162]首先在多个时间帧上对 Gammatone 过滤的特征向量进行组合，建立一个高维的时间谱表示(Spectro-Temporal Representation，STR).此外，由于运动车辆的确切声音特征是未知的，因此文献[162]采用非线性 Hebbian 学习(Nonlinear Hebbian Learning，NHL)规则从 T-F 特征提取出具代表性的独立特征并减少特征空间的维度.STR 和 NHL 均能准确提取原始输入数据的关键特征.该模型在噪声环境下的性能优于同类模型.对于加性高斯白噪声和一般有色噪声，该模型具有良好的鲁棒性.在 SNR 为 0 dB 时，它可以减少 3%的错误率，同时提高 21%～34%的性能；在 SNR 为－6 dB 时，其他模型已经不能正常工作，而它也才只有 7%～8%的错误率.

使用的统计分类器有 BPNN[150,154,160]、GMM[161]、HMM[161]、SVM[159]、基于 STFT 的贝叶斯子空间方法[161]等.在单节点识别结果上，文献[159]提出基于能量的全局决策融合算法，对多个节点做出的决策进行融合.文献[161]研究了在相似工作条件下产生的各种车辆声音的向量分布，使用一组典型的声音样本集合作为训练数据集.文献[156]将各种声音数据按层次分类，结果比没有层次结构的传统水平分类方案要好.文献[156]同时表明了当前 AI 系统的识别能力，通常低于人类专家，但高于未受训练的普通人.

4.3.2.2　交通事故识别

在重大交通事故发生时，车辆运行状态与正常行驶状态相比发生了很大变化，伴随有剧烈碰撞的声音，而且与周围的噪声存在较大的差别.因此，可以通过声音传感器实时采集并分析车辆周围的声音，判别车辆的运行情况，一旦有事故发生，可立即提取碰撞声并识别，并及时向后台救护系统发出报警信号[163].

声音采集装置成本低廉，体积小，安装方便，可靠性强，不易损坏，维护容易.声音检测系统的计算方法相对简单，信号处理量小，既可实时处理又可远程传输，快速准确，不易受雨雪天气和交通条件的影响，可以全天候工作.在事故发生后，报警信号应该将包括事故地理位置在内的信息尽快地传递到指挥中心，可用无线网络来传输数据[163].建立一个快速、高效的应急救援系统，能提高交通事故检测的实时性和准确度[164].

人耳对相同强度、不同频率的声音变化的敏感程度不同.文献[165]利用此特点，用基于人耳等响度曲线的 A 计权滤波器对声信号进行加权，使声信号映射到真实的人耳听觉频域，然后再进行音频事件检测.文献[164]采用单类 SVM 进行异常点检测.文献[165]采用互信息(Mutual information)分析噪声低频域与高频域的相关性，分别作为输入和输出向量，用 RBFNN 建模后估计高频域噪声，用谱减法降噪后获取较纯净的声信号.

在提取音频特征方面，文献[164]使用 Haar-WT 提取声信号的频域特征.文献[166]以小波分解后不同频带的重构信号能量作为特征向量.文献[165]首先二值化目标音频事件的频谱图，定位要保留的频带，提取其中最主要的频率成分.与全频域的 MFCC 特征相比，能降低计算量，提高检测速度，适用于行车环境下的实时音频事件检测.在类型识别方面，文献[166]采用多个 SVM 构成的交通事件分类器，对正常行驶、刹车、碰撞事件的声信号进行识别.

4.3.2.3　交通流量检测

现有交通流量数据采集设备造价高，采集精度不够，后期分析困难.文献[167]提取车辆噪声的时域特征 STE、ZCR，检测端点和特征跳变点，进行车型辨别和分类，统计出交通流量数据.为保证音频信息采集的有效性，数据采集设备安装在车辆加速行驶路段或凸形竖曲线顶部附近.文献[168]依据道路拥堵时机动车怠速声音在环境中所占比例较高的原理，发明一种道路拥堵检测方法.将一定时间内采集到的道路声音进行 FFT，在低频区域(20～40 Hz)内，拥堵与畅通两种状态下的频域能量谱有明显区别.拥堵时怠速

频率处将有明显尖峰,将尖峰陡峭程度转换成系数 k,基于 k 值进行道路状况评判.文献[169]基于声信号判断是否有汽车到来,尤其适用于车流量稀少、基础设施比较差的区域以及智能公路的前期建设阶段,同时对路灯进行智能控制,环保节能.

4.3.2.4 道路质量检测

汽车行驶产生的道路噪声与不同类型、不同磨损状况的路面直接相关.文献[170]基于正常车辆行驶下获得的轮胎声音,使用 ANN 分类器,能够正确预测 3 种路面类型及其磨损情况.该技术可用于创建数字地图,自动识别对车辆行驶道路噪声带来强烈影响的路段,估计道路宏观纹理.对于土木工程部门、道路基础设施运营商以及高级驾驶员辅助系统都有很大好处.文献[171]采集声信号,基于短时平均幅值对信号进行端点检测.以 MFCC 和基于 HHT 的希尔伯特边际谱作为特征,结合 BPNN 实现基于声振法的水泥混凝土路面脱空状况检测.

4.3.3 水上运输业

CA 在江河海洋领域主要用于水声目标识别、船舶定位、安全监控等.利用被动声呐(Passive sonar),如安装在海床上的单水听器来检测船舶和自主水下航行器(Autonomous underwater vehicles)的活动,是对海洋保护区和受限水域进行远程监测的一种有效方法.传统方法利用水声数据的倒谱分析来测量直接路径到达和第一次多径到达之间的时间延迟,从而估计声源的实时范围[172].水下声道的环境不确定性常常是声场(Acoustic field)预测误差的主要来源[173].

近年来,基于 AI 测量船舶距离的方法开始发展起来.文献[172]基于数据增强进行模型训练.在不同 SNR 情况下,运用倒谱数据的 CNN 能够比传统的被动声呐测距方法更远距离地检测出船只,并估计出船只所在的范围.文献[174]在圣巴巴拉海峡进行深水(600 m)船只距离估计实验.将观测船的采集数据作为前馈神经网络(Feed-forward Neural Network, FNN)和 SVM 分类器的训练和测试数据.分类器表现良好,检测范围达到 10 km,远超传统匹配场处理的约 4 km 的检测范围.

CA 技术同样在水声目标识别领域得到应用.文献[175]在浅水环境中记录了 25 个包括干扰的声源信号.每个声源使用单独的类,基于子空间学习法(Subspace learning)和自组织特征映射(Self-Organizing Feature Maps, SOFM)进行分类.文献[176]采用基于核函数的 SVM 模型,在二类(Binary-class)和多类(Multi-class)分类的情况下,准确率均超过线性分类器(Linear classifiers).文献[177]使用水声传感器采集鱼群摄食时的声音,分析其与摄食量的关系,给出摄食时间、摄食量的估计,对于渔业养殖有重要意义.使用机器学习方法需要注意过拟合问题.如文献[175]中,测试时使用训练中出现的信号样本,准确率可以达到 80%~90%;若使用来自相同声源的全新记录样本,准确率则下降为 40%~50%.

4.3.4 航空运输业

4.3.4.1 航空飞行器识别

早在 1985 年,文献[178]就提出一种用声信号识别飞机类型和飞行模式(Flight patterns)的方法.特征参数选择在 A 加权声压峰值的峰值保持频谱(Peak-hold spectra)即 $\frac{1}{3}$ 倍频带(One-third octave band),使用判别分析.文献[179]通过检测直升机声信号来识别机型.利用 WPT 的 T-F 局部聚焦分析能力提取特征向量,用 BPNN 的自适应能力首先证实是直升飞机声信号后,再区分不同型号的直升机.

文献[45]使用嵌入式麦克风阵列采集一个四旋翼飞行器(Quadrotor)的声信号进行飞行事件识别.室外飞行环境很嘈杂,包括转子(Rotors)、风(Wind)和其他声源产生的噪声.对于单声道音频降噪使用鲁棒主成分分析(Robust Principal Component Analysis, RPCA)方法,对于多通道音频降噪使用几何高阶去相关的源分离方法(Geometric High-order Decorrelation based Source Separation, GHDSS).声源盲分离提高了输入声音的 SNR,然后对改善后的声音基于堆叠降噪自动编码机(Stacked Denoising Autoencoder, SDA)和 CNN 进行声源识别(Sound Source Identification, SSI).GHDSS 和 CNN 的结合效果更好.文献[180]同样通过声信号检测旋翼飞行器,基于 MFCC 特征和 DTW 匹配,实现对于直径范围为 40~60 cm 的旋翼飞行器的短距离检测和预警.

4.3.4.2 航空飞行数据分析

黑匣子于 1953 年由澳大利亚的载维·沃伦博士发明,是飞机上的记录仪器.一种是飞行数据记录仪

(Flight Data Recorder，FDR)，记录飞机的高度、速度、航向、爬升率、下降率、加速情况、耗油量、起落架放收、格林威治时间、系统工作状况、发动机工作参数等飞行参数.另一种是座舱话音记录仪(Cockpit Voice Recorder，CVR)，实际上就是一个无线电通话记录器，分 4 条音轨分别记录驾驶舱内所有的声音，包括飞行员与地面管制人员的通话，组员间的对话，机长、空中小姐对乘客的讲话，威胁、爆炸、发动机声音异常以及驾驶舱内各种声音如开关手柄的声音、机组座位的移动声、风挡玻璃刮水器的马达声等.FDR 可以向人们提供飞机失事瞬间和失事前一段时间里飞机的飞行状况、机上设备的工作情况等，CVR 能帮助人们根据机上人员的各种对话分析事故原因，以便对事故作出正确的结论[181-182].

我国在民航事故调查中仍然沿用传统的人耳辨听座舱声音，自动化程度很低.有些声音识别超出了人的生理功能极限，而且经常受到各种噪声掩盖，影响驾驶舱话音记录器作用的发挥.研发基于 CA 技术的驾驶舱话音记录器声音识别系统已迫在眉睫.文献[182]对舱音中的微弱信号——开关手柄声音特性进行分析，验证其符合暂态噪声脉冲模型.对信号进行 STFT 得到频谱，进行 WPT 得到信号在不同频带的能量.以归一化的频谱幅值、频谱幅值熵、归一化的小波 SE、小波 SE 熵作为开关手柄声音的特征，分析其各自的适用范围，使用 SVM 进行识别.

4.3.5　管道运输业

在各种管道传输中，可能会发生因人为损坏或自然因素造成的泄漏事故.如输水管道的漏水、油气输送管道的第三方破坏(Third Party Destroy，TPD)等.此外，在传输管道中频繁使用的阀门也会出现泄漏现象.管道和阀门的泄露现象不易检测.传统的方式是人工监听，需要有丰富的经验，容易造成误判.基于泄漏声音的自动检测是一类很有希望的方法.

早在 1991 年，文献[183]就报导了日本电力中央研究所和东亚阀门公司根据声音检测阀门漏泄.文献[184]研究基于 FFT 自相关算法并嵌入到 DSP 芯片的便携式智能听漏仪，能够在复杂背景噪声中检测出漏水点.文献[185]采用小波降噪，快速有效地提取 TPD 信号，对其奇异点进行定位，以小波分解 SE 和相关统计量作为特征输入 SVM 进行分类，能正确区分切割、挖掘、敲击等典型的 TPD 信号，监控的有效检测距离达到 1 400 m.文献[186]基于 LPCC 特征，利用 HMM 识别损伤或泄漏信号.文献[187]用声音传感器采集声信号，提取 MFCC 特征输入 HMM 识别异常声音，及时发现阀门泄漏并报警.文献[188]研究软管隔膜活塞泵进出口阀门声音实时检测系统，该系统使用 MFCC 作为特征，利用 HMM 分类器识别故障.

管道内检测器用来检测管道腐蚀、局部形变以及焊缝裂纹等缺陷.检测器进行检测工作时，容易在管壁的形变处、三通处和阀门处等位置发生卡堵事件.轻则影响管道正常运输，重则引发凝管事故、导致整条管道报废.因此，研究地面管道内检测器追踪定位技术具有重要意义.文献[189]通过建立声音在土壤中的传播模型实现对卡堵位置的准确定位，后续可用机器学习模型加以研究.

4.3.6　仓储业

制炼厂中产生的声音可以用来检测在容器内发生反应的进展，或检测生产线内的流体流动.声音通过安装在容器外部的传感器来接收.该技术是非侵入性(Non-invasive)的，不需要对过程流体进行采样，避免了污染等潜在风险[190].

在农业上，由于粮食储藏后期技术不过关，虫害导致的玉米损失总量非常庞大.基于声音的害虫检测技术逐渐成为研究热点[191]，已开始实仓多点应用[192].文献[193]研究玉米象、米象、杂拟谷盗等 3 种害虫在玉米中活动的声信号.首先进行加汉宁(Hanning)窗，50 阶带通滤波，小波降噪等预处理，计算 STE、ZCR，在时域进行声信号端点检测，然后提取能量峰值频率，MFCC、ΔMFCC 作为音频特征.当信号能量达到 11 dB 左右时判断可能有害虫存在.采用两种识别办法：一是将声信号的第 1，4，5，6 能量峰值频率输入 Probalistic NN 进行分类识别；二是将声信号的 MFCC、ΔMFCC，振动信号的 LPC、ΔLPC 输入 HMM 进行分类识别.前者比后者识别效果要好.文献[194]在隔音环境下，采集谷蠹、米象和赤拟谷盗等 3 种储粮害虫的爬行声信号，然后进行频域分析获取其功率谱，提取特征向量，输入 BPNN 进行分类识别.

4.4　制造业

近些年，CA 技术在制造业的数十个细分领域中开始逐步产生应用.例如，基于声信号的故障诊断技术被大量应用在机械工程的各个领域，逐渐成为故障诊断领域的一个研究热点.对于很多设备如发动机、

螺旋桨、扬声器等,故障发生在内部,在视觉、触觉、嗅觉等方面经常没有明显变化.而产生的声音作为特例却通常具有明显变化,可用于机械损伤检测[195],成为独特的优势.此外,传统上采用的基于摄像机和传感器的方法,也不能进行早期的故障异常检测[18,196].

4.4.1 铁路、船舶、航空航天和其他运输设备制造业

转辙机用于铁路道岔的转换和锁闭,其结构损伤会直接影响行车安全.在生产过程中,需要对高铁转辙机的重要零件全部进行无损检测.基于声信号进行结构损伤检测具有非接触、高效等优点.文献[197]基于核主分量分析提取声信号特征,用 SVM 进行结构损伤分类识别.

水泥厂输送带托辊运行工况恶劣,数量众多,又要求连续运转,并且在线检修不便.要保证输送机长期连续稳定的运行,对有故障托辊的快速发现和及时处理非常重要.为快速安全可靠地发现有故障隐患的托辊,需适时安排检修,避免托辊带病运转可能造成的更高的停机维修成本及产量损失,减少工人的工作强度[198].瑞典的 SKF 轴承公司发明了一种托辊声音检测仪,原理是对运行中的托辊发出的声音进行辨别,从而判断托辊是否正常,并对异常声音发出报警信号.该装置设有声音遮盖技术,可以区分托辊良好运行和带故障运行所发声音的区别.即使在高噪声环境下,亦能过滤出周边部件的信号,准确捕捉故障托辊信号.

4.4.2 通用设备制造业

4.4.2.1 发动机

发动机是飞机、船舶、各种行走机械的核心部件[199],有柴油机(Diesel engine)、汽油机(Gasoline engine)、内燃机(Internal combustion engine)、燃气涡轮发动机(Gas turbine engines)等几种.发动机故障是发动机内部发生的严重事故,传统的发动机故障诊断高度依赖于工程师的技术能力,如文献[200]根据发动机的高、中、低 3 个频带的频谱特性对其进行分析,通过分析汽车噪声的强度可大致判断出汽车发动机部件的故障.人工判断具有很大的局限性,一些经验丰富的技术人员也会有一些失败率,造成时间和金钱的严重浪费.因此,急需一种自动化的故障诊断(Fault diagnosis)方法[201].系统既可直接用于自动诊断,提高系统可靠性,节约维护成本,也可作为经验不足的技术人员的训练模块.而且避免了拆分机器安装振动传感器的传统诊断方式的麻烦[202].

发动机在正常工作时,其振动的声音及振动频谱是有规律的.在发生各种故障时,会发出各种异常响声[203],频谱会出现变异和失真.每一个发动机故障都有一个特定的可以区分的声音相对应[201,204],可用于进行基于声信号的故障诊断,此类研究早在 1989 年即已开始[205].常见的发动机故障有失速[204],正时链张紧器损坏[206],定时链条故障(Timing chain faults)[207],阀门调整(Valve-setting)[207-208],消声器泄漏(Muffler leakage)[207],发动机启动问题(Engine start problem)[208],驱动带分析(Drive-belt analysis)[208],发动机轴瓦故障[209],漏气[210],齿轮异常啮合[210],连杆大瓦异响[210],断缸故障[211],油底壳处异响[212]、前部异响[212]、气门挺柱异响[212],发动机喘振[213],滑动主轴承磨损故障[214],箱体异响[215],右盖异响[215],左盖异响[215]等.

发动机声信号的采集通常使用麦克风/声音传感器[211,216-219],也有的系统使用智能手机[208].声音采集具有非接触式的特点,如文献[218]利用发动机缸盖上方的声压信号对发动机进行故障诊断.文献[208]采用基于频谱功率求和(Spectral power sum)与频谱功率跳跃(Spectral power hop)两种不同的聚类技术将音频流分割.使用的 T-F 表示有 CWT[220]、STFT[196,208,213,221]、WT[222]、HHT[209]、稀疏表示[223]等.

使用的声信号降噪采用各种滤波,如 SVD 滤波、WT 滤波、EMD 滤波[224].理论描述表明,发动机噪声产生机理与独立成分分析(Independent Component Analysis, ICA)模型的原理相同.文献[220]用 ICA 将发动机噪声信号分解成多个独立成分(Independent Components, IC).文献[215]研究表明,小波阈值降噪效果较好,但是具有突变、不连续特性的发动机声信号会产生伪 Gibbs 现象,进一步改进为基于平移不变小波的阈值降噪法.文献[209]基于一种改进的 HHT 进行 EMD 分解,利用端点优化对称延拓和镜像延拓联合法抑制端点效应,同时采用相关性分析法去除 EMD 分解的虚假分量,用快速独立成分分析(Fast ICA)去除噪声.文献[213]对低频区域的声信号使用 db8 小波的 7 层分解进行降噪.文献[225]利用 Fast ICA 盲源分离法对船舶柴油机的噪声信号进行分离.

提取对应于发动机各种故障状态的音频特征是研究的难点.由于声源的数量和环境的影响,这些特征非常复杂,可能被严重破坏,使得故障检测和诊断变得困难[226].使用的特征有基于 WPT 的能量[224],经验模态分解(Empirical Mode Decomposition, EMD)的能量[224],MFCC[216,224,227-228],ZCR[228],f_0[228],归一化均方误差(Normalized Mean Square Error, NMSE)[201],FF[201],声信号伪谱(Pseudospectrum)的链式编码(Chaincode)[207],小波频带(Wavelet subbands)导出的统计特征[207],自回归系数(Autoregressive coefficients)[205],自回归倒谱系数(Autoregressive cepstral coefficients)[205],小波熵[222],码激励线性预测编码(Code Excited Linear Prediction, CELP)[216],$\frac{1}{3}$倍频程值[215],基于混沌(Chaos)技术提取的时间序列动力学轨道非平稳运行特征[229],共振峰的位置[210],基于 WT 的自功率谱密度[202],基于频带局部能量的区间小波包特征[202],信号的波形指标、峰值指标、脉冲指标、裕度指标、峭度系数、偏度系数[218]等.文献[230]使用 SVD 方法确定观察矩阵中哪个特征能够最好地识别内燃机的技术状态.提取音频特征集后经常采用 PCA 法,在不损失有效信息的情况下,将原始特征向量中的冗余信息约简[203,231].

初级的故障检测可以只区分正常和异常[232],更高级的方法可识别具体的故障种类.故障识别可采用模板匹配的方法[216].文献[201]收集和分析了不同类型汽车的声音样本,代表不同类型的故障,并建立了一个频谱图数据库.将测试中的故障与数据库中的故障进行比较,匹配度最高的数据库中的故障被认为是检测到的故障.使用的距离有灰色系统(Grey system)的关联度量(Relational measure)[205]、马氏距离(Mahalanobis distance)[205]、Kullback-Leiber 距离[205].文献[203]采用线性预测方法模拟发动机声音时域特征与转速(表征发动机状态)之间的关系.更多的方法是基于机器学习统计分类器,如 SVM[224,231],HMM[228],高斯混合模型-通用背景模型(Gaussian Mixture Model-Universal Background Model, GMM-UBM)[227],模糊逻辑推理(Fuzzy logic inference)系统[208],BPNN[196,208,213,217],概率神经网络(Probabilistic Neural Network, Probabilistic NN)[215],小波包与 BPNN 相结合的 WNN[202].文献[207]采用 DTW 进行两级故障检测.第一阶段将样本粗分为健康和故障两类,第二阶段细分故障种类.若有其他相关证据,可利用信息融合理论对发动机故障进行综合诊断[218].

4.4.2.2　金属加工机械制造

刀具状态是保证切削加工过程顺利进行的关键,迫切需要研制准确、可靠、成本低廉的刀具磨损状态监控系统.切削声信号采集装置成本低廉,结构简单,安放位置可调整.基于它的检测技术,信号直接来源于切削区,灵敏度高,响应快,非常适用于刀具磨损监控.需要注意的是,切削声信号频率低,容易受到环境噪声、机床噪声等的干扰,获取高 SNR 的刀具状态声音是监控系统的关键[233].

早在 1991 年,文献[234]已利用金属切削过程中的声音辐射检测工具的状态,即锋利、磨损、破损.以 5 kHz 为边界,低频和高频带的频谱成分作为特征,可以很容易地区分锋利和磨损工具.对于破损的情况,鉴别需要更多的特征.

文献[233]首先采集刀具在不同磨损状态下的切削声信号.通过时域统计分析和频域功率谱分析,发现时域统计特征均方值与刀具磨损状态具有明显的对应关系,与刀具磨损相关的特征频率段为 2～3 kHz.还实验研究了不同主轴转速、进给速率对刀具磨损状态的影响.基于小波分析,将声信号分为 8 个不同的频带,以不同 SE 占信号总能量的百分比作为识别刀具磨损状态的特征向量,用 BPNN 进行状态识别.

加工的主要目标是产生高质量的表面光洁度,但是只能在加工周期结束时才能进行测量.文献[235]在加工过程中对加工质量进行检测,形成一种实时、低成本、准确的检测方法,能够动态调整加工参数,保持目标表面的光洁度,并且调查了车削过程中发出的声信号与表面光洁度的关系.AISI 52100 淬火钢的实验表明,这种相关性确实存在,从声音中提取 MFCC 可以检测出不同的表面粗糙度水平.

文献[236]利用采煤机切割的声信号进行切割模式的识别.将工业麦克风安装在采煤机上,采集声信号.利用多分辨率 WPT 分解原始声音,提取每个节点的归一化能量(Normalized energy)作为特征向量.结合果蝇和遗传优化算法(Fruitfly and Genetic Optimization Algorithm, FGOA),利用模糊 C 均值(Fuzzy C-Means, FCM)和混合优化算法对信号进行聚类.通过在基本果蝇优化算法(Fruitfly Optimization Algorithm, FOA)中引入遗传比例系数,克服传统 FCM 算法耗时且对初始质心敏感的缺点.

冲压工具磨损会显著降低其冲压的产品的质量,其状态检测为许多制造行业迫切需求.文献[237]研究了发出的声信号与钣金冲压件磨损状态的关系.原始信号和提取信号的频谱分析表明,磨损进程与发出的声音特征之间存在重要的定性关系.文献[238]介绍了一种金刚石压机顶锤检测与防护装置.运用声纹识别技术,提取顶锤断裂声特征参数,建立顶锤断裂声模板库.再将金刚石压机工作现场声音特征参数与顶锤断裂声模板库进行比对,相符则切断金刚石压机工作电源,实现了对其余完好顶锤的保护.

有经验的焊接工人仅凭焊接电弧声音的响度和音调特征就可以判断焊缝质量.文献[239]基于焊接自动化系统采集焊接声信号,可忽略噪声的影响.根据铝合金脉冲焊接声信号的特点,提取 3 164～4 335 Hz 内声信号的短时幅值平均值、幅值标准差、能量和、对数能量平均值作为特征,通过 SVM 识别铝合金脉冲熔透状态,用粒子群优化算法对 SVM 模型的参数进行优化.

4.4.2.3 轴承、齿轮和传动部件制造

旋转机械(轴承、齿轮等)在整个机械领域中有着举足轻重的地位,发生故障的概率又远远高于其他机械结构,因此对该类部件进行状态检测与故障诊断就尤为重要[240].针对传统的振动传感器需要拆分机器、不易安装的缺点,可通过在整机状态下检测特定部位的噪声来判定轴承与齿轮等是否异常[241].

滚动轴承是列车中极易损坏的部件,其故障会导致列车故障甚至脱轨.非接触式的轨旁声学检测系统(Trackside Acoustic Detector System, TADS)采集并分析包含圆锥或球面轴承运动信息的振动、声音等信号[240,242-243].由美国 Seryo 公司设计的轴承检测探伤器[244]除了用轨道旁的声音传感器收集滚动轴承发出的声音,还包括红外线探伤器.文献[245]提出一种铁路车轮自动化探伤装置,研究所需探测的缺陷类型.通过传声器检测发射到空气中的声音可用于发现轮辋或辐板的裂纹,而擦伤或轮辋破损则最好由安装在钢轨上的加速度计来探测.

文献[240]提出两种针对列车轴承信号的分离技术.第一种通过多普勒畸变信号的伪 T-F 分布,来获取不同声源的时间中心和原始频率等参数,利用多普勒滤波器实现对不同声源信号的逐一滤波分离;第二种基于 T-F 信号融合和多普勒匹配追踪获取相关参数,再通过 T-F 滤波器组的设计运用,得到各个声源的单一信号.

使用的音频特征有 MFCC[242]、小波熵比值即峭熵比(Kurtosis Entropy Ratio, KER)[243]和 EEMD[243].分类器有 BPNN[242]、SVM[242].文献[244]则采用类似单类识别的方法,识别从某一轴承中产生的任何所接收到的标准信号.一旦检测出非标准频率信号,将报警.能在因表面发热导致红外线探测器触发前检测出损坏的轴承.

4.4.2.4 包装专用设备制造

文献[246]公开了一种基于声信号的瓶盖密封性检测方法.声信号的产生由电磁激振装置对瓶子封盖激振产生,由麦克风采集.文献[247]基于声信号实现啤酒瓶密封性快速检测.瓶盖受激发后产生受迫振动,其振动幅度和振动频率与瓶盖的密封性存在一定的关系.瓶内压力增高时,若瓶盖密封性好,其振动频率就高,振幅就小;反之,若密封性差,振动频率就比较低,振幅也比较大.

4.4.3 电气机械和器材制造业

电机是用于驱动各种机械和工业设备、家用电器的最通用装置.电机有很多种,如同步电机(Synchronous motors)[248]、直流电机(DC machine)[249]、感应电机(Induction motor)[250].为保证其安全稳定运行,常常需要工作人员定期检修、维护.电机在发生故障时,维护人员听电机发出的声音,以人工方式判断故障的类型,耗费大量人力,而且无法保证及时检测到故障,急需自动化检测系统[251].基于声信号的声纹识别系统将提取的音频特征与某一类型的故障联系起来[250],可以识别出电机异响[252]及各种类型的故障,如线圈破碎和定子线圈短路[253].

文献[251]利用声音传感器在电机轴向位置采集电机的声信号.文献[254]结合 EMD 与 ICA,通过 EMD 的自适应分解能力,解决 ICA 中信号源数目的限制问题;同时利用 ICA 方法的盲源分离能力,避免 EMD 分解的模态混叠现象.通常需要对音频信号进行预加重、分帧、加窗等预处理[255].文献[255]使用自适应门限的音频流端点检测进行分割.

使用的 T-F 表示有 FFT[253]、WT 及 WPT[252,256-260].小波分析对信号的高频部分分辨率差,小波包分

解方法能够对信号高频部分进行更加细化地分解并能更有效地检测出发电机故障.因为人耳对相位不敏感,只需要对幅度谱进行分析[252].使用的音频特征有 LPC[249],LPCC[255],根据 SVD 得到的特征向量[252],MFCC[255,261],基于加权、差分的 MFCC 动态特征[255],故障信号与正常信号小波能量包的相对熵、各频带的综合小波包能量相对熵[259].PCA 被用来进行特征维度压缩[252].

使用的统计分类器有线性 SVM[253]、KNN[248]、HMM[255,261]、BPNN[146,256,257,260].针对 BPNN 收敛速度慢的问题,文献[260]提出了两点改进：利用区域映射代替点映射和动态改变学习速率.考虑到电机的故障率很低,很难收集到足够多的各类故障样本,且电机异音形成过程复杂,文献[251]和[252]基于 SVM 进行单类学习(Single class learning)实现异音电机检测.以足够数量的正常、无异音电机样本为基础建立一个判别电机声音是否异常的判别函数,不需要异音样本,凡是检测有不符合正常电机声音特征的样本一律判为有故障样本.文献[259]根据小波包能量相对熵首先确定电机是否有故障,之后通过比较大小判断故障所处的频带位置,从而确定电机为何种故障.

电力系统中的许多设备在运行或操作时会产生声音,对应于各种状态.高压断路器是电力系统不间断供电的关键性保护装置,断路器合闸的声信号可用于识别其运行时的机械状态[262].变压器是变电站中的重要设备.变压器在正常运行时,有较轻微、均匀的嗡嗡声.如果突然出现异常的声音,则表明发生故障.不同的声音对应于不同的故障[263].电力电缆发生故障时,故障电弧会发出声音[264],可用于故障定位.电力开关柜的内部故障电弧在剧烈放电前的局部放电会产生电弧声音,可用于故障电孤检测与预警[265].航天继电器中多余物的存在会导致其可靠性下降,不同的声音对应于不同的材质.

各种电力设备主要依靠人工进行故障检测,耗时耗力.电力设备在运行时经常是高电压和强电磁场等复杂环境,不利于接触式设备故障检测方法.有经验的技术人员可以直接凭借电气设备工作时所发出的声音来判断设备是否发生异常,基于声信号的故障诊断近年来逐渐发展起来.采集声音数据的方法各不相同.文献[264]在低压电气输电线路导线绝缘层上设置声音传感器,文献[266]采用麦克风阵列,有效抑制周围噪声干扰并将波束对准目标信号.

声音采集过程中经常会混合干扰信号如人的说话声,与电气设备发出的声音是统计独立的[266].文献[266]采用 ICA 来分离有用的电气设备声信号.文献[262]利用改进的势函数法进行声源数估计,通过 EEMD 得到多个 IMF 分量,重构形成符合聚类声源数的多维信号,利用拟牛顿法优化快速 ICA 算法提取断路器操作产生的声信号.文献[267]总结了常见的线性模型盲信号分离算法：基于负熵的固定点算法,信息极大化的自然梯度算法,联合近似对角化算法,并将这 3 种算法分别用于对电力设备作业现场多种混合声源信号进行分离.文献[268]提出一种基于 WPT 分解信号、自适应滤波估计噪声与遗传算法寻优重构相结合的声信号增强算法.

文献[262]根据包络特征比对识别断路器的状态.文献[269]使用 SVM 实现对断路器当前状态的识别.文献[270]对航天继电器中多余颗粒物碰撞噪声的声音脉冲包络进行分析,使用 RBFNN 将颗粒自动分为金属、非金属两类.文献[271]提取 0～1 000 Hz 内的 21 个谐波作为特征,建立样本库,利用 VQ 的 LBG 算法训练得到变压器和高抗设备的码本,与未知声音特征匹配后实现运行状态的识别.文献[266]用 MFCC 作为声信号特征,与专家故障诊断库中各种各样的故障信号进行匹配,根据 DTW 判断是否发生电气设备故障.

4.4.4　纺织业

细纱断头的低成本自动检测一直是纺纱企业急需解决的一个问题.文献[272]利用定向麦克风采集 5 个周期的钢丝圈转动产生的声信号.正常纺纱时的声信号都具有分布均匀的 5 个较高波峰,而发生纺纱断头时采集到的声信号不具有该特点.按照此标准即可判断纱线是否发生断头.

4.4.5　黑色及有色金属冶炼和压延加工业

文献[273]对金属和非金属粘接结构施加微力,在频域提取与粘接有关的声信号的特征用于后续模式识别.文献[274]撞击非晶合金产品使其产生振动,并采集发出的声信号.以声信号衰减时间的长短作为特征,判断产品的合格性,可以准确地检测出非晶合金产品内部是否存在收孔或裂纹等缺陷.

文献[275]采集氧化铝熟料与滚筒窑撞击所产生的声音,通过分析频谱、幅度等数据区别出熟料的 3 种

状态：正常、过烧、欠烧，进行自动质量检测.文献[276]采集成品熟料与滚筒窑撞击所产生的声音，经滤波、谱分析等处理后，对烧结工序中的异常状态进行判断并报警.

在铝电解生产过程中，电解槽内电解质和锪液循环流动、界面波动、槽内阳极气体的排出、阳极效应的出现都伴随着相应的特征声音.检测这些特征声信号并分析，能够判断出铝电解槽的运行状况[277].针对铝锭铸造是否脱模的故障检测难题，文献[278]尝试利用铸模敲击声信号进行诊断分析.首先基于改进的小波包算法对敲击声音进行降噪.进行频域分析后发现，某次敲击后如果铝锭脱模，那么将与下一次敲击声音存在明显的峰值频率差.此现象可作为故障特征，进行基于阈值的检测.

角钢是铁塔加工的必备原料.若不同材质的钢材混用，将对铁塔的强度、韧性、硬度产生很大影响.在铁塔加工过程中，角钢进行冲孔时会发出一定的声音，不同材质的角钢加工时会发出不同的声音.Q235 和 Q345 是两种标准角钢材质.文献[279]利用传感器采集并提取单个冲孔周期的声信号，基于 MFCC 和 DTW 计算待测模板与 Q235 和 Q345 两种标准模板之间的距离，距离小者判定为该种角钢材质.文献[280]分析 Q235 和 Q345 两种材质角钢声信号的频谱特征，计算在特定高频频带与低频频带的能量比值，找到能区别两种材质的能量比取值范围作为特征.

4.4.6 非金属矿物制品业

热障涂层(Thermal Barrier Coatings，TBC)是一层陶瓷涂层，沉积在耐高温金属或超合金的表面，对基底材料起到隔热作用，使得用其制成的器件(如发动机涡轮叶片)能在高温下运行.TBC 有 4 种典型的失效模式：表面裂纹、滑动界面裂纹、开口界面裂纹、底层变形.文献[281]以 WPT 特征频带的小波系数为特征，BPNN 为分类器，基于声信号进行 TBC 失效检测.文献[282]提取冲击声的 T-F 域特征及听觉感知特征，通过模式识别研究基于冲击声的声源材料自动识别.

4.4.7 汽车制造业

汽车的 NVH(Noise，Vibmtion，Harshness)表示噪声、振动与舒适性.汽车噪声主要来自发动机，是影响汽车乘坐舒适性的重要因素.对发动机、车辆传动系等进行声品质分析及控制的研究具有重要意义.声品质的改善目标是获得容易被人接受的、不令人厌烦的声音[283-284].

文献[285]针对 C 级车，在一汽技术中心的半消声室内采集 4 个车型、5 个匀速工况下由发动机引起的车内噪声，用等级评分法对声音样本的烦躁度打分，计算出声音样本的 7 个客观心理声学参数，对主观评价值和客观参数进行相关分析.与主观评价值相关性较大的心理声学参数是响度、尖锐度、粗糙度.文献[284]使用 EEMD 获得的 IMF 的熵作为特征，比心理声学参量效果更佳.

以心理声学参数作为声品质预测模型的输入，主观评价值作为声品质预测模型的输出，建立声品质烦躁度的预测模型[283].文献[285]训练确定 BPNN 的结构，包括输入、输出层神经元个数、隐含层数、隐含层神经元个数和传递函数.用遗传算法(GA)对 BPNN 的权值和阈值进行编码，采用选择、交叉和变异等操作寻求全局最优解，将遗传输出结果作为 BPNN 的初始权值和阈值，得到声品质烦躁度的 GA-BPNN 预测模型.文献[284]以 Morlet 小波基函数作为隐含层节点的传递函数构建 WNN，同时运用 GA 优化 WNN 的层间权值和层内阈值，构造 GA-WNN 模型用于传动系声品质预测.

文献[283]研究结果表明，响度是影响人们对车辆排气噪声主观感受的最主要因素，和满意度呈负相关.使用多元线性回归(Multiple Linear Regression，MLR)与 BPNN 理论分别建立了柴油发动机噪声声品质预测模型，实验表明 BPNN 模型预测值与实测值更接近，能够更好地反映客观参数和主观满意度间的非线性关系.文献[285]表明，在网络训练误差目标相同的情况下，GA-BPNN 预测模型比 BPNN 预测模型的收敛速度提高了 5 倍.由于 BPNN 预测模型初始权值和阈值的随机性，导致相同样本每次的预测结果都存在较大差异.而 GA-BPNN 预测模型采用遗传算法对 BPNN 的初始权值和阈值进行优化，保证了网络的稳定性，对声音样本声品质预测结果有较高的一致性.文献[284]研究表明 GA-WNN 网络较 GA-BPNN 网络能更准确、有效地对传动系声品质进行预测.

汽车内部安静并不是好汽车的唯一目标，不同的汽车要有对其合适的声音.文献[286]研究发动机声音和客户偏好之间的关系，对汽车声音进行主观评价.研究发现，加速度和恒定速度下的声音感知明显不同，不同的车主群体有不同的感知.

4.4.8　农副食品加工业

在鸡蛋、鸭蛋等的加工过程中，从生产线上分选出破损蛋是一道重要工序.国内主要依靠工人在灯光下观察是否有裂纹，或转动互碰时听蛋壳发出的声音等方法来识别和剔除破损鸡蛋.这种方法效率低下，精度差，劳动强度大，成本高.研究自动化的禽蛋破损检测方法意义重大[287].经验表明，好蛋的蛋壳发出的声音清脆，而破损蛋的蛋壳发出的声音沙哑、沉闷[287]，这使得基于声音音色进行蛋类质量判别成为可能.

文献[288]以鸡蛋赤道部位的 4 个点(1,2,3,4)作为敲击位置，采集鸡蛋的声信号.文献[287]对鸭蛋自动连续敲击，采集鸭蛋的声信号.在实际环境中，还需要音频分离或降噪技术.文献[289]根据海兰褐蛋鸡声音与风机噪声的 PSD 在 1 000～1 500 Hz 频率范围内存在的差异，从风机噪声环境中分离提取蛋鸡声音.文献[290]用自制的橡胶棒分别敲击鸡蛋中间、中间偏大头一点、中间偏小头一点等 3 个位置，低通滤波消除噪声干扰，每次采样 128 点数据.

已用的音频特征各不相同，文献[288]使用鸡蛋最大、最小 2 个特征频率(f_{max}，f_{min})的差值 $\Delta f(=f_{max}-f_{min})$，文献[291]使用敲击声信号的衰竭时间、最小 FF、4 点最大频率差，文献[292]使用共振峰对应的模拟量频率值、功率谱面积、高频带额外峰功率谱幅值和第 32 点前后频带功率谱面积的比值.除了常规的好、坏两种分类，文献[291]进一步将鸡蛋分类为正常蛋、破损蛋、钢壳蛋、尖嘴蛋等 4 种.已用的识别方法有的基于规则，如文献[288]以 1 000 Hz 作为裂纹鸡蛋的识别阈值.有的基于机器学习模式识别，如 Bayes 判别[287,292]、基于最大隶属度原则的模糊识别[290-291]、ANN[293]等.

4.4.9　机器人制造

机器人需要对周围环境的声音具有听觉感知能力.AED 在技术角度也属于 CA，但专用于机器人的各种应用场景[294].如文献[295]面向消费者的服务消费机器人，在室内环境中识别日常音频事件.文献[296]面向灾难响应的特殊作业机器人，识别噪声环境中的某些音频事件，并执行给定的操作.文献[297]面向阀厅智能巡检的工业机器人，对设备进行智能检测和状态识别.

文献[295]将机器人听觉的整体技术框架分为分割连续音频流、用稳定的听觉图像(Stabilized Auditory Image，SAI)对声音进行 T-F 表示、提取特征、分类识别等步骤.使用的音频特征有 PSD[294]，MFCC[294]，对数尺度频谱图的视觉显著性[294]，小波分解的第五层细节信号的质心、方差、能量和熵[297]，从 Gammatone 对数频谱图中提取的多频带 LBP 特征，提高对噪声的鲁棒性，更好地捕捉频谱图的纹理信息[298].使用的机器学习模型有 SVM[294]、BPNN[297]、深度学习中的受限玻尔兹曼机(Restricted Boltzmann Machine，RBM)[296].基于人与机器人的交互，建立了一个新的音频事件分类数据库，即 NTUSEC 数据库[298].

4.5　农、林、牧、渔业

4.5.1　农业

在现代绿色农业中，喷洒农药需首先判断农作物上的昆虫是否是害虫.害虫活动的声音经常具有明显特点，例如文献[299]使用麦克风在隔音箱内录制黄粉虫成虫的爬行和咬食活动的声音，发现咬食活动声音脉冲信号的时间带有明显规律性，时间间隔约为 0.68 s.咬食活动声音频率的主峰值在 70～93 Hz，低于爬行活动的 140～180 Hz.文献[300]结合声信号分离和声音活动端点检测，基于频谱图模板进行害虫的匹配识别.在确定存在害虫后，为避免喷洒农药量过多或不足，需根据病虫害的实际情况和分布种类混药进行变量式喷雾.文献[301]首先识别混杂在复杂背景音下的不同病虫害的声音，用 DNN 自动学习特征并分类，并根据识别的病虫害种类及分布情况进行自动在线混药.

文献[302]将听诊器改装成一种装置，用以在检疫检验中探测在水果和谷粒中昆虫嚼食的声音.先是在实验室进行实验，从柚子、枇杷、木瓜中迅速而准确地将实蝇检测出来.仅一条刚刚孵化出一天的幼虫也能从柚子中检测出来.后来发现谷蠹和麦蛾也能从玉米、水稻和小麦的谷粒中检测出来.

小麦是最重要的农作物之一，其硬度是评价小麦品质的重要指标，需建立自动、客观、准确的检测技术.文献[303]采集单粒小麦籽粒下落碰撞产生的声信号，进行谱估计和 WT，提取时域和频域的 16 个特征，采用回归分析(Regression analysis)和 ANN 建立小麦声音特性与千粒重和硬度之间的数学模型，以达到预测小麦品质的目的.文献[304]自制小麦自动进料器，使小麦逐粒、自然地下落击靶，采用声音传感

器接收小麦击靶发出的声信号.经调理、放大、A/D 转换及预处理后,在时域提取 ZCR、波形指标、脉冲因子等特征,在频域提取基于 FFT 和 DCT 的特征,利用线性回归(Linear Regression, LR)、BPNN 建立特征参数和对应的小麦硬度指数之间的预测模型.文献[305]进一步在不同采样频率、不同下落高度情况下,在时域和 FFT、DCT、WT 等频域分别提取特征.研究表明,无论是时域还是频域,在采样频率为 200 kHz、下落高度为 40 cm 时,声音特征与小麦硬度指数相关性较好,最后运用 LR 分析和 BPNN 建立了小麦硬度基于声音的预测模型.

榴莲是东南亚的一种绿色尖刺水果.因为价格昂贵,又很难从外观上判断榴莲的成熟度,迫切需要开发一种在不进行切割或破坏条件下的自动识别榴莲成熟度的方法,这对果农、消费者和零售商都很重要.文献[306]提取信号的频谱特征,用 HMM 模型识别榴莲是否已成熟,并确定成熟的程度.当敲击次数从 1 次增加到 5 次时(每次不超过 80 ms),识别准确率会随之增加.文献[307]提取声音特征后使用 N-gram 模型识别榴莲是否成熟,利用多数投票从 N-best 列表中找到成熟度.

同样的道理,为满足采收前后对西瓜成熟度的无损检测的需求,文献[308]实现了在田间环境下通过声音自动检测西瓜成熟度的方法.使用 STE 和 ZCR 判断击打信号的起止点,完整提取每次敲击西瓜的声音片段,滤波消除干扰噪声.不同成熟度的西瓜敲击声音对应不同的功率谱峰值频率范围,作为西瓜成熟度检测的规则.

4.5.2 林业

我国的森林盗伐现象猖獗.文献[309]专门设计实现了一种基于声音识别的森林盗伐检测传感器.文献[310]通过对声信号的频谱特征分析、相似度值及 SNR 计算,检测是否存在链锯伐木行为.

蛀干害虫是一类危害严重的森林害虫.因其生活隐蔽,林木受害表现滞后,使得检测和防治极其困难.基于声音识别的害虫检测技术具有无损、快速、准确等优势,潜力巨大.文献[311]研究红棕象甲虫、亚洲长角草甲虫、天牛甲虫幼虫等 3 种木蛀虫的生物声学(Bioacoustics)规律.发现通过咬音和摩擦音可以有效地进行物种识别.

文献[312]用高灵敏度录音机采集双条杉天牛害虫的活动声信号.采用 ANN 和滤波器消噪,提取较为纯净的双条杉天牛幼虫活动声音.发现其幼虫活动声音脉冲数量随害虫密度增加而增加,呈线性关系,且取食声信号能量大于爬行声信号能量.

文献[313]在野外环境下,距离 50 cm 内,采集云杉大墨天牛、光肩星天牛和臭椿沟眶象 3 种蛀干害虫的幼虫在活动、取食时产生的声信号.受风声和汽车噪声影响较大,但是与鸟鸣和虫鸣噪声在 T-F 域有显著差别,可相对容易地分离.研究发现不同种类幼虫产生的声信号在 T-F 域特征上均有明显差异,但与数量无明显关系.幼虫声音脉冲个数与幼虫数量正相关,可利用脉冲个数估计幼虫数量.

4.5.3 畜牧业

在养殖业中,准确高效地检测畜禽信息,有助于提高养殖及加工效率,及时发现生病或异常个体,减少经济损失.人工观察方式主观性强且精度低,嵌入式检测手段又会造成动物应激反应,发展智能自动检测手段是目前的研究热点[314].禽畜的声音直接反应了它们的各种状况,可用于状态监测.例如,针对猪的大规模养殖中频发的呼吸道疾病问题,可通过检测咳嗽状况对猪的健康状况进行预警[315].

对采集的猪的声音,首先进行加窗分帧[316]等预处理.音频流分割需要端点检测[315].文献[317]通过 ZCR 和 STE 进行端点检测,文献[318]基于双门限进行端点检测.之后进行降噪处理,如谱减法[315]、小波阈值法[318].已用的音频特征有 MFCC[315,317-318]、ΔMFCC[318].文献[316]和[318]分别定义了猪在 8 种行为状态下的声音.常用的识别匹配及分类算法有 VQ[319]、HMM[315-316,318]、SVM[316-317]、Adaboost[316]等.

4.6 水利、环境和公共设施管理业

4.6.1 水利管理业

钱塘江潮涌高且迅猛,伤人事故频发.为提高潮涌实时检测与预报水平,文献[320]提出一种基于音频能量幅值技术的潮涌识别方法.通过采集沿江各危险点潮涌来临前后的声音,经滤波后进行 FFT 幅频特性分析,提取潮涌音频能量幅值特征值,自动识别并进行潮涌实时检测与预报.

为最大限度开发利用空中水资源,减轻干旱、冰雹等造成的损失,利用高炮、火箭实施人工影响天气

作业是解决水资源紧缺的有效途径.文献[321]实现了一种基于炮弹声音采集、识别、处理的高炮作业用弹量统计系统.

4.6.2　生态保护和环境治理业

动物发出的各种声音具有不同的声学特点,作为交流的手段.例如,沙虾虎鱼发出的声音由一系列脉冲组成,以每秒 23～29 次的速度重复.单脉冲的频谱为 20～500 Hz,峰值在 100 Hz 左右.绝对声压水平在 1～3 cm 范围内为 118～138 dB[322].雄性石首鱼集体的声音甚至可以掩盖捕鱼船的引擎噪声[323].大熊猫"唔"的叫声是警告性行为,"唔"音的长短和强弱反映大熊猫的情绪及警告程度.若警告无效,"唔"音加强和变急,进一步转变成发怒的叫声"汪"、"呢"和"哞",下一步即可能发生打斗行为[324].

生态环境中的声音在自动物种识别(Species recognition)与保护,野生动物及濒危鸟类监控,森林声学和健康检测,以及对相关环境、进化、生物多样性、气候变化、个体交流等的理解分析上都有重要应用[325-334].文献中根据声音研究分析过的动物已有很多种,如海豹[335],海豚[336],大象[337],鱼类[322-323,338-339],蛙类[340],鸟类[341-348],昆虫[349-353]等.

文献[342]在鸟类背上绑定麦克风采集声音.除了真实录制的数据,还可以采用合成声音数据[354].在真实场景中,存在风或其他动物的叫声等背景噪声干扰[341],需要来抑制噪声[327].文献[355]采用 ICA 进行野外动物声音的声源分离.文献[353]和[333]分别使用 Adobe Adition 和 Gold Wave 软件对录制的声音文件进行人工降噪.文献[325]将早期的短时谱估计算法与一种基于双向路径搜索的噪声功率谱动态估计算法相结合,提出一种适用于高度非平稳噪声环境下的音频增强算法.文献[356]使用改进的多频带谱减法进行降噪.文献[332]研究了基于 DWT 的声音降噪方法.传统的噪声估计需要假设背景噪声是平稳的,不能适应实际的非平稳环境噪声.文献[347]将一种基于双向路径搜索的动态噪声功率谱估计算法与经典的短时谱声音增强技术相结合,进行非平稳环境噪声下的声音增强.此外,传感器节点的能量消耗也是实际系统的一个问题[345].

进行动物识别需要将连续音频流分割为有意义的单元.文献[356]和[325]采用基于 STE 的门限进行端点检测.文献[329]通过聚类在声音记录中检测 4 种音频事件,即哨声(Whistles)、点击(Clicks)、含糊音(Slurs)和块(Blocks).文献[329]对通过 WT 后的中、低频声信号进行端点检测,不但可以去除高斯噪声,而且可以去除高频脉冲噪声对系统的影响.文献[347]通过比较每个 2 维 T-F 矩阵点的幅度谱来定位每个鸟叫音节(Syllable)在整个 T-F 图中的起始位置,实现连续鸟叫声音的音节分割.文献[348]将遥感领域使用的图像分割技术引入频谱图进行鸟叫声分割.

频谱图是最常用的 T-F 表示,有时需要形态学滤波(Morphological filtering)等预处理[343].文献[339]为克服特征提取时间长、数量多等问题,采用稀疏表示.文献[357]从神经机制方面研究了听觉的特征.使用的音频特征有 LPC[328,358]、MFCC[328,351,353,358-359]、频谱图特征(Spectrogram feature)[340]、音色特征[360]、基于特征学习自动提取的特征[342,359]、基于频带的倒谱(Sub-Band based Cepstral, SBC)[361].此外,文献[341]从频谱图提取特征.文献[335]采用海豹叫声的持续时间作为特征反映海豹之间的个体差异.文献[334]使用 MP 算法提取有效信号的 T-F 特征.动物叫声经常在 T-F 图上表现出不同的纹理特征.文献[325]用和差统计法进行 T-F 纹理特征提取,在 4 种不同位置关系下计算 5 个二次统计特征,得到一个 20 维的 T-F 纹理特征向量.文献[347]使用图像处理中的灰度共生矩阵纹理分析法,提取 T-F 图 4 个方向上的 5 种纹理特征.文献[362]使用 A-DCTNet(Adaptive DCTNet)提取鸟叫的声音特征作为分类器的输入.A-DCTNet 与 CQT 类似,其滤波器组的中心频率以几何间距排列,能比 MFCC 等特征更好地捕获对人类听觉敏感的低频声音信息.文献[344]在研究鸣禽的过程中,发现除了传统的绝对音高(Absolute Pitch, AP)信息,频谱形状等音色类特征也可以用于鸣禽的叫声.文献[345]首先基于 Sigmoid 函数进行音调区域探测(Tonal Region Detection, TRD),然后采用基于分位数的倒谱归一化(Quantile-based cepstral normalization)方法提取 Gammatone-Teager 能量倒谱系数(Gammatone-Teager Energy Cepstral Coefficients, GTECC),形成最终的 TRD-GTECC 特征.文献[356]对频谱图进行 Radon 变换和 WT 提取特征.文献[332]针对不同频带的重要程度,提出了基于 WT 和 MFCC 的小波 Mel 倒谱系数 WT-MFCC.文献[346]为克服 MFCC 对噪声的敏感性,提取更符合人耳听觉特性的 Gammatone 滤波器倒谱系数

(GFCC)及小波系数,组合后作为特征向量.文献[339]基于稀疏表示利用正交匹配追踪法(Orthogonal Matching Pursuit, OMP)提取与水声信号最为匹配的少数原子作为特征.

对于待识别的声音种类,文献[329]首先为这些目标构建模板,之后用 DTW 等进行匹配[341],这适用于数据有限的情况.文献[363]基于鸟声在 T-F 平面高度结构化的特点,利用阈值方法对鸟类声音进行帧级的二元决策,并融合得到最终结果.文献[360]基于频谱-时间激发模式(Spectro-Temporal Excitation Patterns, STEP)进行听觉距离匹配.更多的方法采用机器学习分类器,如 HMM[328,341,343,345],GMM[333,343],RF[325,347],KNN[358],RNN[362],ANN[352],DNN[345],SVM[328,334,339,356],Probalistic NN[351,353],PLCA[342],迁移学习[359],CNN[359,364-365],基于内核的极限学习机(Kernel-based Extreme Learning Machine, KELM)[326]等.分类模型的设计及调试需考虑实际应用场景.例如,文献[366]对每种鸟类的鸣叫声和鸣唱声建立双重 GMM 模型,并讨论不同阶数对 GMM 模型的影响.使用多个模型时,可使用后期融合(Late-fusion)方法将模型融合起来[364].文献[349]采用 Probalistic NN 和 GMM 的分数级融合(Score-level fusion),提出一种针对昆虫层次结构(如亚目、科、亚科、属和种)的高效的分层(Hierarchic)分类方案.

机器学习的方法需要较多的标注数据.例如文献[340]的数据集包括来自美国的 48 个无尾目类动物物种的 736 个叫声数据,文献[367]使用数千个未处理的鸟类现场录音.数据量不足时可使用数据增强方法增加训练数据[364].为充分利用大量无标签的动物声音(如鸟叫),文献[324]使用基于稀疏实例的主动学习(Sparse-Instance based Active Learning, SI-AL)和基于最小置信度的主动学习(Least-Confidence-Score-based Active Learning, LCS-AL)方法,有效地减少专家标注.

以色列科学家发现一种检测水污染的新方法——听水生植物发出的声音.用一束激光照射浮在水面的藻类植物,根据藻类反射的声波,分析出水中的污染物类型以及水受污染的程度.激光能刺激藻类吸收热量完成光合作用,在这一过程中,一部分热量会被反射到水中,形成声波.健康状况不同的藻类的光合作用能力不同,反射出的热量形成的声波强度也不一样.

4.7 建筑业

4.7.1 土木工程建筑业

地下电缆经常遭到手持电镐、电锤、切割机、机械破碎锤、液压冲击锤、挖掘机等工程机械的破坏[368-369],影响供电系统稳定性.电缆防破坏成为电力部门所面临的一个重大技术难题,急需研发基于声音的地下电缆防外力破坏方法,识别挖掘设备的声音,进行预警判断,对事发地定位.

文献[368]对声信号采集、预加重、分帧、加窗预处理后,使用 LPCC 及提出的单边自相关线性预测系数倒谱系数(One-Sided Autocorrelation LPCC, OSA-LPCC)作为特征,用 SVM 进行分类,OSA-LPCC 的抗噪声性能优于 LPCC.文献[369]采用 8 通道的麦克风十字阵列,在夜晚环境下对 4 种挖掘设备在不同距离作业下采集声信号,建立声音特征库.使用 MFCC、ΔMFCC、ΔΔMFCC、频谱动态特征,输入 BPNN、KNN 和极限学习机(Extreme Learning Machine, ELM)进行设备识别.文献[370]使用 STE 比值 SFER(Short-term Frames Energy Ratio)、短时 T-F 谱幅值比(Short-term Spectrum Amplitude Ratio, SSAR)、短时 T-F 谱幅值比占比(Short-term Spectrum Amplitude Ratio Rate, SSARR)、冲击脉冲宽度(Width of Pulse, WoP)、冲击脉冲间隔(Interval of Pulse, IoP)等统计特征识别,受距离变化影响较小,性能稳定,比 LPCC、MFCC 等经典特征泛化能力更好.

4.7.2 房屋建筑业

文献[371]通过单点单次敲击抹灰墙采集声信号,通过 MFCC 特征和 DTW 对抹灰墙黏结缺陷进行识别.文献[372]通过烧砖的敲击声音判断烧砖内部是否存在缺陷,并进一步区分缺陷类别.采用无限冲击响应(Infinite Impulse Response, IIR)滤波器进行降噪,采用近似熵方法判断敲击声音端点.以频谱峰值点之间的关系作为特征,用 PCA 方法进行故障检测.老房子的木质结构和家具中可能存有木蛀虫,是物体腐朽的主要原因.文献[373]基于木蛀虫的活动声音检测其是否存在.因为幼虫发出的声音相对较低,背景噪声会大大降低检测的准确性.文献[374]采集建筑物内部金属断裂的声音进行分析,识别可能出现在建筑物内部的裂缝,避免倒塌等灾难性后果的发生.

4.8　采矿业、日常生活、身份识别、军事等

4.8.1　采矿业

为监测钻井过程中的井壁坍塌、井底岩爆等井下工况信息，文献[375]采集返出岩屑在排砂管中运输所产生的声信号.根据 STE 确定声音段的起止点，利用 NN 算法去噪，DTW 识别岩屑的大小，计算岩屑流量，进而判断井下工况.

4.8.2　日常生活

CA 技术在日常生活中也有许多应用.烹饪过程中会产生特定的声音，可用于进行烹饪过程的检测和控制.文献[376]基于声信号识别水沸腾的状态.文献[377]发明另一种基于声信号的装置，检测电磁炉水沸腾状态，而且还能自动关机.文献[378]发明一种风扇异音检测系统.文献[379]发明的一种智能吸油烟机能对厨房的各种环境声音进行分析检测，判断该声音是否是烹饪过程发出的声音.进而判断该烹饪声音所对应的油烟量级别，设置对应的吸油烟机的启动或关闭或调节风机转速，实现对吸油烟机的智能控制.文献[380]发明一种带有保健检测的手表，通过翻身声响检测人的睡眠质量.文献[381]使用耳垫声音传感器采集咀嚼食物的声信号，基于模式识别技术实时获取咀嚼周期和食物类型，预测固体食物的食量，进行饮食指导.文献[382]分别使用动圈式麦克风(Dynamic microphone)和电容式麦克风(Condenser microphone)采集在有偿自动回收机(Reverse Vending Machines，RVM)中进入废物的声音，基于 SVM 和 HMM 对废物的种类和大小进行分类，如自由落体、气动撞击、液体冲击.文献[383]基于 PCA 处理后的声音的帧能量，根据方差最小原则判断同型号待测打印纸的柔软度，分为 5 级.文献[384]发明一种日用陶瓷裂纹检测装置.通过敲击碗坯发出声音，声音传感器捕获信号后判断是否有裂纹.文献[385]中的地震声响测定仪基于 FFT 模型快速识别不同声音的地震脉冲，预测将要发生危险的地带.

4.8.3　身份识别

脚步声是人最主要的行为特征之一.正常情况下每个人走路的脚步声是不一样的，蕴含着性格、年龄、性别等多方面信息，具有可靠性和唯一性.脚步声识别在家庭监控、安全防盗、军事侦察等领域具有重要意义.常规算法采用 MFCC 特征，GMM 分类器识别.由于同一人穿不同的鞋，在不同的地板上走路时脚步声会有差异，这类对不同发声机制较为敏感的方法具有很大的约束性和限制性，鲁棒性不足.

文献[386]采用双门限比较端点检测法分割脚步声，维纳滤波降噪.提出一种新的特征，即脚步声的持续时间与脚步声的间隔时间，使用 KNN 分类识别.对于同一个人在不同发声机制下的脚步声识别具有良好的鲁棒性和适用性.文献[387]用谱减法对频谱图降噪.在训练过程中，计算在安静环境下采集的每个训练样本的对数能量，形成 2 维频谱图.应用数字图像中的关键点检测与表征技术在 2 维频谱图中检测关键点，形成每个关键点的局部频谱特征.在识别过程中，利用基于最小错误率的贝叶斯决策(Bayesian decision)理论对待识别样本进行分类.

手写声音(Hand writing sound)是真实环境中存在的一种噪声，其信息不仅可以用来识别文字如数字字符，还可以进行书写者身份识别(Writer recognition).文献[388]记录受试者用圆珠笔在纸上写字时的声音.采用 MFCC、ΔMFCC、ΔΔMFCC 作为特征，HMM 作为分类器模型，进行书写者身份识别.

4.8.4　军事

CA 在军事上也有许多重要应用.下边仅举几例.

4.8.4.1　目标识别

现代化的智能侦察与作战方式需要准确感知到自身周围是否出现机动目标，并判别它们的类别和数量，以配合目标定位、跟踪和攻击等功能.文献[389]设计实现一个车辆声音识别系统.提取 STE、ZCR、谐波集、SC、LPC、MFCC 和小波能量等音频特征，用遗传算法对备选特征库进行优化产生最终的特征子集，对两类目标车辆进行分类.文献[390]基于声信号对战场上的车辆进行分类识别，集成谐波集、MFCC、小波能量等 3 种特征，并用 PCA 进行降维融合处理.

被动声音目标识别也称为被动式声雷达(Passive acoustic radar).与传统雷达探测技术相比，有抗干扰、低功耗、不易被发现等优点，可以弥补雷达低空探测存在盲区的不足.声音传感器实时接收目标的声音信息，与典型的声信号(如坦克、轮式车辆、直升机等)通过模式匹配进行自动识别.文献[391]基于 MFCC

和 DTW 对低空四旋翼飞行器的声信号进行声纹识别.文献[392]提出在战场上对同时多低空目标进行分类的方法.采用 ICA 将混合信号分为若干个声源并去除噪声.提取 MFCC 作为特征,使用 K-means 聚类后产生训练和识别的特征向量(Eigenvector),输入模拟声信号时域变化的 HMM 进行分类.

文献[393]基于无线声音传感器网络(Wireless Sound Sensor Networks, WSSN)搜集数据,结合 MFCC 和 DTW 实现一个海上无人值守侦察系统,对进入侦察区域的目标进行外形轮廓和声音的识别.由于海上船只、海面飞行物、海鸟以及海洋背景声音的复杂性,只能对进入侦察海域的声音进行初步感知.

在复杂的电磁环境中,对雷达辐射源音频信号进行人工识别耗时长、易于误判和错判.文献[394]结合 MFCC 和 DTW 实现基于声纹技术的雷达辐射源音频自动识别.文献[395]利用战术无人机上的声音传感器探测和定位地面间接火力源(如迫击炮和火炮),需先对发动机噪声和空气流动噪声进行降噪处理.

4.8.4.2 其他应用

枪声分析在现实中有着很多应用.枪声信号的声音特征显示出强烈的空间依赖性,文献[396]使用空间信息和一种基于它的决策融合规则来处理多声道声音武器分类.文献[397]在自行火炮实车测试中,利用瞬态过程中的声信号对齿轮箱进行故障诊断,避免了常规振动测试方法无法实现非接触、不解体、无损在线检测的弊端,采用倒谱分析克服 FFT 不能分析非稳态信号的不足.文献[398]基于振动信号和声信号用于火炮发射现场对发射次数的计数,解决了火炮发射人工计数准确性差的问题.文献[399]采用 Probabilistic NN 在火炮音频特征和火炮零部件(凸轮轴)硬度之间进行非线性映射,实现零部件的硬度分类.

5 总结与展望

本文全面总结了基于一般音频/环境声的计算机听觉技术涉及的相关声学基础、概念与原理、典型技术框架、已有的应用领域.与语音信息处理、音乐信息检索(MIR)、自然语言处理(Natural Language Processing, NLP)、计算机视觉(Computer Vision, CV)等相关领域相比,该学科在国内外发展都比较缓慢.

影响 CA 发展的几个原因包括:(1) 环境声音具有非平稳、强噪声、弱信号、多声源混合等特点.一个实际系统必须经过音频分割、声源分离或增强/去噪后,才能进行后续的内容分析理解.音频特征经常需要根据具体应用场景下声音的特点进行专门设计,直接套用语音信息处理或 MIR 中的特征则效果较差.(2) 各种音频数据都源自特定场合和物体,难以全面搜集和标注.文献中使用最多的两个公共数据库是 DCASE 和 RWCP,但是这两个数据库主要面向日常生活场景中的一些典型声音种类.对于其他绝大多数 CA 应用领域,不仅数据不公开,而且数据规模小,种类不全甚至完全不同,严重影响了算法的研究及比较.(3) 基于一般音频/环境声的 CA 几乎都是交叉学科,除了日常生活场景,绝大多数应用需要了解相关各领域的专业知识和经验.(4) 作为新兴学科,还存在社会发展水平、科研环境、科技评价、人员储备等各种非技术类原因阻碍着 CA 技术的发展.

声音信号具有丰富的信息量,在很多视觉、触觉、嗅觉不合适的场合下,具有独特的优势.声音信号通常被认为与振动信号具有较大的相关性,但声音信号具有非接触性,避免了振动信号采集数据的困难.基于一般音频/环境声的 CA 技术属于 AI 在音频领域的分支,直接面向社会经济生活的各个方面,在医疗卫生,安全保护,交通运输、仓储,制造业,农、林、牧、渔业,水利、环境和公共设施管理业,建筑业,采矿业,日常生活,身份识别,军事等数十个领域具有众多应用,是一门非常实用的技术.目前该领域在国内外已开始起步发展,但在许多研究和应用领域仍接近于空白,具有无限广阔的发展前景.

参考文献:

[1] ZHUANG X, ZHOU X, HASEGAWA-JOHNSON M A, et al. Real-world acoustic event detection[J]. *Pattern Recognition Letters*, 2010, **31**(12): 1543-1551.

[2] LU T, WENG Y, WANG G. Audiotory movie summarization by detecting scene changes and sound events[C]// 2014 22nd International Conference on Pattern Recognition. Stockholm, Sweden: IEEE, 2014: 756-760.

[3] PHAN H, KOCH P, KATZBERG F, et al. What makes audio event detection harder than classification?[C]// 2017 25th European Signal Processing Conference (EUSIPCO). Kos, Greece: IEEE, 2017: 2739-2743.

[4] JAEGER C P, LASZLO C A. Machine recognition of sound sources[J]. *Canadian Acoustics*, 1999, **27**(3): 124-125.

[5] CHACHADA S, KUO C C J. Environmental sound recognition: A survey[J]. *APSIPA Transactions on Signal and Information Processing*, 2014, **3**: 1-15.

[6] DIBAZAR A A, BANGALORE A, PARK H, et al. Hardware implementation of dynamic synapse neural networks for acoustic sound recognition[C]// The 2006 IEEE International Joint Conference on Neural Network Proceedings. Vancouver, BC, Canada: IEEE, 2006: 2015-2022.

[7] TERENCE N W Z, DAT T H, DENNIS J, et al. Robust sound event recognition under TV playing conditions[C]// 2013 IEEE China Summit and International Conference on Signal and Information Processing. Beijing, China: IEEE, 2013: 332-336.

[8] KHUNARSAL P, LURSINSAP C, RAICHAROEN T. Very short time environmental sound classification based on spectrogram pattern matching[J]. *Information Sciences*, 2013, **243**: 57-74.

[9] WON M, ALSAADAN H, EUN Y. Adaptive audio classification for smartphone in noisy car environment[C]// Proceedings of the 25th ACM international conference on Multimedia. California, USA: ACM, 2017: 1672-1679.

[10] MASUM S M A, HIROSE K, PRENDINGER H. Context awareness using environmental sound cues and commonsense knowledge [C]// International Conference on Signal Processing and Multimedia Applications. Milan, Italy: ICSPMA, 2009: 193-196.

[11] MA L, MILNER B, SMITH D. Acoustic environment classification[J]. *ACM Transactions on Speech and Language Processing*, 2006, **3**(2): 1-22.

[12] MARTÍN-MORATÓ I, COBOS M, FERRI F J. Analysis of data fusion techniques for multi-microphone audio event detection in adverse environments [C]// 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP). Luton, UK: IEEE, 2017: 1-6.

[13] DENNIS J, TRAN H D, CHNG E S. Overlapping sound event recognition using local spectrogram features and the generalised hough transform[J]. *Pattern Recognition Letters*, 2013, **34**(9): 1085-1093.

[14] LIN W, LI Y. Lower SNR sound event recognition using noisy training sample [C]// 2015 8th International Congress on Image and Signal Processing (CISP). Shenyang, China: IEEE, 2015: 1448-1453.

[15] HATTORI Y, ISHIHARA K, KOMATANI K, et al. Repeat recognition for environmental sounds[C]// RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication. Kurashiki, Okayama, Japan: IEEE, 2004: 83-88.

[16] TERENCE N W Z. Sound event recognition in home environments [D]. Singapore: Nanyang Technological University, 2014.

[17] YE J, KOBAYASHI T, MURAKAWA M, et al. Robust acoustic feature extraction for sound classification based on noise reduction[C]// 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Florence, Italy: IEEE, 2014: 5944-5948.

[18] DAVY M. Application of time-frequency techniques to sound signals: Recognition and diagnosis[M]// Time-Frequency Analysis: Concepts and Methods. London, UK; Hoboken, NJ, USA: ISTE Ltd; John Wiley & Sons, Inc., 2008: 383-408.

[19] YE J, KOBAYASHI T, MURAKAWA M, et al. Kernel discriminant analysis for environmental sound recognition based on acoustic subspace[C]// 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC, Canada: IEEE, 2013: 808-812.

[20] VAN NORT D, BRAASCH J, OLIVEROS P. Sound texture recognition through dynamical systems modeling of empirical mode decomposition[J]. *The Journal of the Acoustical Society of America*, 2012, **132**(4): 2734-2744.

[21] ROMA G, NOGUEIRA W, HERRERA P. Recurrence quantification analysis features for environmental

sound recognition [C] // 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, NY, USA: IEEE, 2013: 1-4.

[22] LI Y, ZHANG X, JIN H, et al. Using multi-stream hierarchical deep neural network to extract deep audio feature for acoustic event detection[J]. *Multimedia Tools and Applications*, 2018,**77**(1): 897-916.

[23] LI J, DAI W, METZE F, et al. A comparison of deep learning methods for environmental sound detection[C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 126-130.

[24] COWLING M, SITTE R. Comparison of techniques for environmental sound recognition[J]. *Pattern Recognition Letters*, 2003,**24**(15): 2895-2907.

[25] DAI W, LI J C, PHAM P, et al. Acoustic scene recognition with deep neural networks (DCASE challenge 2016) [C] // Detection and Classification of Acoustic Scenes and Events 2016. Budapest, Hungary: DCASE, 2016,**3**: 1-6.

[26] **张小梅,杨鼎才.基于支持向量机模型的环境音分类研究[J].电子测量技术**,2008,**31**(9): 121-123.

[27] VALERO X, ALIAS F. Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification[J]. *IEEE Transactions on Multimedia*, 2012,**14**(6): 1684-1689.

[28] CASEY M A. Reduced-rank spectra and minimumentropy priors as consistent and reliable cues for generalized sound recognition[C]// 7th European Conference on Speech Communication and Technology (EUROSPEECH 2001). Aalborg, Denmark: IEEE, 2001: 167-170.

[29] ADAVANNE S, PERTILÄ P, VIRTANEN T. Sound event detection using spatial features and convolutional recurrent neural network [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 771-775.

[30] DENG S, HAN J, ZHANG C, et al. Robust minimum statistics project coefficients feature for acoustic environment recognition [C] // 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Florence, Italy: IEEE, 2014: 8232-8236.

[31] PARK S, LEE Y, HAN D K, et al. Subspace projection cepstral coefficients for noise robust acoustic event recognition[C]// 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 761-765.

[32] ABIDIN S, TOGNERI R, SOHEL F. Enhanced LBP texture features from time frequency representations for acoustic scene classification [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 626-630.

[33] KOBAYASHI T, YE J. Acoustic feature extraction by statistics based local binary pattern for environmental sound classification [C] // 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Florence, Italy: IEEE, 2014: 3052-3056.

[34] HERSHEY S, CHAUDHURI S, ELLIS D P W, et al. CNN architectures for large-scale audio classification [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 131-135.

[35] MITROVIC D, ZEPPELZAUER M, EIDENBERGER H. Towards an optimal feature set for environmental sound recognition[R/OL]. Technical Report TR-188-2-2006-03, 2006. http: // www. ims. tuwien. ac. at/publication_master. php.

[36] CHMULIK M, JARINA R. Bio-inspired optimization of acoustic features for generic sound recognition [C]// 2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP). Vienna, Austria: IEEE, 2012: 629-632.

[37] GRZESZICK R, PLINGE A, 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.

[38] TAKAHASHI N, GYGLI M, PFISTER B, et al. Deep convolutional neural networks and data augmentation for acoustic event recognition [C] // Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH.San Francisco, USA: INTERSPEECH, 2016,**8**: 2982-2986.

[39] 李荣杰，蒋兴浩，孙锬锋.一种基于音频词袋的暴力视频分类方法[J].上海交通大学学报，2011，**45**(2)：214-218.

[40] 卢斌.基于非监督特征的多模态色情视频检测算法研究与系统实现[D].上海：上海交通大学，2016.

[41] 辛欣.基于潜在概率语义模型的异常声音检测的研究与应用[D].北京：中国科学院大学，2016.

[42] 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.

[43] TOKOZUME Y, HARADA T. Learning environmental sounds with end-to-end convolutional neural network[C]// 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 2721-2725.

[44] ZHANG J, YIN J, ZHANG Q, et al. Robust sound event classification with bilinear multi-column ELM-AE and two-stage ensemble learning[J]. *Eurasip Journal on Audio, Speech, and Music Processing*, 2017(1): 2-11.

[45] SUGIYAMA O, UEMURA S, NAGAMINE A, et al. Outdoor acoustic event identification with DNN using a quadrotor-embedded microphone array[J]. *Journal of Robotics and Mechatronics*, 2017, **29**(1): 188-197.

[46] SIVAPRAKASAM T, DHANALAKSHMI P. A robust environmental sound recognition system using frequency domain features[J]. *International Journal of Computer Applications*, 2013, **80**(9): 5-10.

[47] SHAUKAT A, AHSAN M, HASSAN A, et al. Daily sound recognition for elderly people using ensemble methods[C]// 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). Xiamen, China: IEEE, 2014: 418-423.

[48] SCHRODER J, MORITZ N, ANEMULLER J, et al. Classifier architectures for acoustic scenes and events: implications for DNNs, TDNNs, and perceptual features from DCASE 2016[J]. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 2017, **25**(6): 1304-1314.

[49] KUN Z, TAKUYA M. Gamification based Participatory Environmental Sound Collection Framework for Human Activity Recognition[J]. **研究報告ユビキタスコンピューティングシステム** (UBI), 2013(2): 1-8.

[50] TERENCE N W Z, DAT T H, HOA H T, et al. Adaptive semi-supervised tree SVM for sound event recognition in home environments[C]// 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Kaohsiung, Taiwan, China: IEEE, 2013: 1-4.

[51] SHUYANG Z, HEITTOLA T, VIRTANEN T. Active learning for sound event classification by clustering unlabeled data[C]// 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 751-755.

[52] SU T W, LIU J Y, YANG Y H. Weakly-supervised audio event detection using event-specific Gaussian filters and fully convolutional networks[C]// 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 791-795.

[53] ELIZALDE B, LEI H, FRIEDLAND G, et al. An i-vector based approach for audio scene detection[C]// IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events.[s.l.]: IEEE, 2013: 1-3.

[54] ELIZALDE B, SHAH A, DALMIA S, et al. An approach for self-training audio event detectors using web data[C]// 2017 25th European Signal Processing Conference (EUSIPCO). Kos, Greece: IEEE, 2017: 1863-1867.

[55] MUN S, SHON S, KIM W, et al. Deep neural network based learning and transferring mid-level audio features for acoustic scene classification[C]// 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 796-800.

[56] ARORA P, HAEB-UMBACH R. A study on transfer learning for acoustic event detection in a real life scenario[C]// 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP). Luton, UK: IEEE, 2017: 1-6.

[57] TROWITZSCH I, MOHR J, KASHEF Y, et al. Robust detection of environmental sounds in binaural

auditory scenes[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017, **25**(6): 1344-1356.

[58] JANSEN A, GEMMEKE J F, ELLIS D P W, et al. Large-scale audio event discovery in one million youtube videos[C]// 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 786-790.

[59] SHI Y Y, WEN X, SHE B. ESRD03 Database and the Labeling for Environmental Sound Recognition [C].The 18th International Congress on Acoustics(ICA). Kyoto, Japan: ICA, 2004: 1687-1690.

[60] HUANG Q, XU Y, JACKSON P J B, et al. Fast tagging of natural sounds using marginal co-regularization[C]// 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 2991-2995.

[61] **成丽君,张丽萍,张宇波.基于改进特征** HMM **的尖叫音频检测算法**[J].**山西农业大学学报(自然科学版)**,2009,**29**(4): 365-369.

[62] ZHANG T, KUO C C J. Video content parsing based on combined audio and visual information[C]// Multimedia Storage and Archiving Systems IV. Boston, MA,USA: International Society for Optics and Photonics, 1999, **3846**: 78-90.

[63] **薛涛,杜军朝,刘惠,刘悦韡,陈文靖.一种基于环境声音的场景识别方法及装置及移动终端**.CN: 103456301A[P].2013.

[64] **倪宁,卢刚,卜佳俊.基于音频分析的视频场景检测**[J].**计算机仿真**,2006,**23**(8): 184-187.

[65] SUNDARAM H, CHANG S F. Audio scene segmentation using multiple features, models and time scales[J]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, **4**: 2441-2444.

[66] BREGMAN A S. Auditory scene analysis: The perceptual organization of sound[M]. Cambridge, Massachusetts, USA: MIT Press, 1994.

[67] NIU F, GOELA N, DIVAKARAN A, et al. Audio scene segmentation for video with generic content[C]// Multimedia Content Access: Algorithms and Systems II. San Jose, California, USA: International Society for Optics and Photonics, 2008: 1-5.

[68] NITANDA N, HASEYAMA M, KITAJIMA H. An audio-scene cut detection method using fuzzy c-means algorithm for audio-visual indexing[C]// 2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No. 04CH37512). Vancouver, BC, Canada: IEEE, 2004,**2**: 89-92.

[69] LENG Y, ZHOU N, SUN C, et al. Audio scene recognition based on audio events and topic model[J]. *Knowledge-Based Systems*, 2017,**125**: 1-12.

[70] YANG M, KANG J. Pitch features of environmental sounds[J]. *Journal of Sound and Vibration*, 2016, **374**: 312-328.

[71] CHU S, NARAYANAN S, KUO C C J. Environmental sound recognition using MP-based features[C]// 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. Las Vegas, NV, USA: IEEE, 2008: 1-4.

[72] YANG W, KRISHNAN S, YANG W, et al. Combining temporal features by local binary pattern for acoustic scene classification[J]. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 2017,**25**(6): 1315-1321.

[73] NTALAMPIRAS S, POTAMITIS I, FAKOTAKIS N. Automatic recognition of urban environmental sounds events[C]// International Association for Pattern Recognition Workshop on Cognitive Information Processing (IAPR).Santorini,Greece: IEEE,2008: 110-113.

[74] IMOTO K, ONO N, IMOTO K, et al. Spatial cepstrum as a spatial feature using a distributed microphone array for acoustic scene analysis[J]. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 2017,**25**(6): 1335-1343.

[75] BISOT V, SERIZEL R, ESSID S, et al. Feature learning with matrix factorization applied to acoustic scene classification[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017, **25**(6): 1216-1229.

[76] RAKOTOMAMONJY A. Supervised representation learning for audio scene classification[J]. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 2017,**25**(6): 1253-1265.

[77] PHAN H, HERTEL L, MAASS M, et al. Improved audio scene classification based on label-tree embeddings and convolutional neural networks[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017,**25**(6): 1278-1290.

[78] 蔡群,陆松年,杨树堂.基于音视特征的视频内容检测方法[J].计算机工程,2007,**33**(22): 240-242.

[79] 庄越挺,傅正钢,叶朝阳,等.基于视听分层模型的实时爆炸场景识别[J].计算机辅助设计与图形学学报,2004,**16**(1): 90-97.

[80] DOV D, TALMON R, COHEN I. Multimodal kernel method for activity detection of sound sources[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017,**25**(6): 1322-1334.

[81] DIVAKARAN A, RADHAKRISHNAN R, XIONG Z, et al. A procedure for audio-assisted browsing of news video using generalized sound recognition[C]// SPIE proceedings series. Santa Clara, CA, United States: Society of Photo-Optical Instrumentation Engineers, 2003: 160-166.

[82] 许荣.融合音视频特征的足球视频检索研究[D].长春: 吉林大学,2008.

[83] 李伟,李子晋,高永伟.理解数字音乐-音乐信息检索技术综述[J].复旦学报(自然科学版),2018,**57**(3): 271-313.

[84] 中华人民共和国国家质量监督检验检疫总局、中国国家标准化管理委员会.国民经济行业分类,GB/T4754—2017[S].北京: 中国标准出版社,2017.

[85] R·M·阿尔特斯.利用声音分类器和麦克风的声学患者监测.CN: 102088911B[P].2014-10-29.

[86] 魏山城,张家平,王国华,等.光电型医用听诊器的研究[C]// 鲁豫赣黑苏五省光学(激光)学会 2011 年会.山东日照: 山东省科学技术协会,2011: 59.

[87] YOU M, LIU Z, CHEN C, et al. Cough detection by ensembling multiple frequency subband features[J]. *Biomedical Signal Processing and Control*, 2017,**33**: 132-140.

[88] 王博.基于高斯混合模型的咳嗽音检测研究[D].重庆: 重庆大学,2011.

[89] AMRULLOH Y A, ABEYRATNE U R, SWARNKAR V, et al. Automatic cough segmentation from non-contact sound recordings in pediatric wards[J]. *Biomedical Signal Processing and Control*, 2015, **21**: 126-136.

[90] TURNER R D, BOTHAMLEY G H. How to count coughs?Counting by ear, the effect of visual data and the evaluation of an automated cough monitor[J]. *Respiratory Medicine*, 2014, **108**(12): 1808-1815.

[91] DRUGMAN T. Using mutual information in supervised temporal event detection: Application to cough detection[J]. *Biomedical Signal Processing and Control*, 2014,**10**: 50-57.

[92] 谢忠好,曾碧新. 24 小时便携式咳嗽音信号监测[J].数理医药学杂志,2015(6): 893-895.

[93] 魏栋,陈纪赞,田联房,等.基于隐马尔可夫模型的咳嗽声识别研究[C]// 中国语音学学术会议暨庆贺吴宗济先生百岁华诞语音科学前沿问题国际研讨会.北京: 中国语言学会语音学分会,2008: 1-6.

[94] 郭锐,杜平,陈先友,等.听声诊断—一种尘肺辅助检测方法的临床研究[J].中国职业医学,2008,**35**(3): 191-193.

[95] 张晓燕.基于 BP 神经网络的肺音识别与诊断研究[J].电子测试,2016,**13**: 111-113.

[96] 傅星瑜,李凡.基于小波变换和 BP 神经网络的肺炎呼吸音识别算法研究[J].无线互联科技,2017(5): 105-108.

[97] 杨纯,朱昌平,韩苏,等.一种基于声音检测技术的睡眠质量辅助系统.CN: 103126650A[P].2013-6-5.

[98] 张诗玓.基于安卓手机的睡眠呼吸暂停综合症筛查系统[D].大连: 大连理工大学,2016.

[99] 马干军.鼾声信号检测与分析算法研究[D].南京: 南京理工大学,2016.

[100] 张健,曾辉.基于 SOPC 的呼吸音检测系统分析[J].无线互联科技,2017(21): 56-58.

[101] 王笑咪.鼾声声压级参数在单纯鼾症和阻塞性睡眠呼吸暂停低通气综合征患者间的鉴别意义[D].温州: 温州医科大学,2016.

[102] LIAO W H, LIN Y K. Classification of non-speech human sounds: Feature selection and snoring sound analysis[C]// 2009 IEEE International Conference on Systems, Man and Cybernetics. San Antonio, TX,

USA: IEEE, 2009: 2695-2700.

[103] YANG C, CHEUNG G, STANKOVIC V, et al. Sleep apnea detection via depth video and audio feature learning[J]. *IEEE Transactions on Multimedia*, 2017, **19**(4): 822-835.

[104] CHEN T E, YANG S I, HO L T, et al. S1 and S2 heart sound recognition using deep neural networks [J]. *IEEE Transactions on Biomedical Engineering*, 2017, **64**(2): 372-380.

[105] **刘胜,付兴铭,徐涌.一种基于声音传感器识别诊断风湿性心脏病的系统及诊断方法**.CN: 103479386A[P].2014-01-01.

[106] **王欧阳,甄辉,姚剑.基于智能手机的胎心率检测系统**[J].**计算机工程**,2016,**42**(4): 288-294.

[107] **许莉莉,郭学谦.基于心音周期性的自动分段研究**[J].**中国医疗设备**,2018(1): 86-88.

[108] **成谢锋,陈亚敏.S1 和 S2 共振峰频率在心音分类识别中的应用**[J].**南京邮电大学学报(自然科学版)**, 2017, **37**(5): 7-12.

[109] **雍希.基于 EMD 及 SVD 的心音信号提取方法研究**[D].**重庆: 重庆大学**,2016.

[110] KIM I Y, LEE S M, YEO H S, et al. Feature extraction for heart sound recognition based on time-frequency analysis[C]// Proceedings of the First Joint BMES/EMBS Conference. 1999 IEEE Engineering in Medicine and Biology 21st Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Society. Atlanta, GA, USA: IEEE,1999: 960.

[111] ZHOU J, HE W, DAN C, et al. Feature extraction and recognition of heart sound[C]// 2008 World Automation Congress. Hawaii, HI, USA: IEEE, 2008: 1-4.

[112] NELSON A T . Wavelet transform and pattern recognition method for heart sound analysis. US: 8152731B2[P]. 2012.

[113] **成谢锋,傅女婷,陈胤,等.一种心音小波神经网络识别系统**[J].**振动与冲击**,2017,**36**(3): 1-6.

[114] **闫峰.基于移动计算的病态声音识别**[D].**秦皇岛: 燕山大学**,2010.

[115] **崔建国,张雅雅.具有胎音监护、胎动检测及胎动轨迹诱导功能的可穿戴式胎教腹带系统**.CN: 104306018A[P].2015-01-28.

[116] **D·P·怀特,H·拉克劳特,K·达瑙斯基,等.利用声音检测进行口服药物依从性监测**.CN: 101252918A[P].2008.

[117] **O·博纳富,R·弗洛朗,V·M·A·奥夫雷.X 射线检查中的流声音**.CN: 102245105A[P].2011-11-16.

[118] **占小杰.肌音信号采集及其在假肢手控制中的研究应用**[D].**上海: 上海师范大学**,2016.

[119] **钱桂萍.监控系统中异常点声源的检测与定位**[D].**南京: 南京理工大学**,2016.

[120] **陈志全.异常声音监控系统关键技术研究**[D].**北京: 中国科学院大学**,2016.

[121] **李超,熊璋,孟岩,等.基于视听信息融合的智能监控系统**[J].**计算机工程与应用**,2004,**40**(31): 218-221.

[122] **聂蓉.应用于 ATM 机的异常声音检测方法及系统**.CN: 102148032A[P].2011-08-10.

[123] CHEN J, KAM A H, ZHANG J, et al. Bathroom activity monitoring based on sound [C]// International Conference on Pervasive Computing. Munich, Germany: Springer-Verlag, 2005: 47-61.

[124] KOMOGUCHI N, YAMANE K, TANAKA S. Recognition of Sound Environment by a Bathroom Monitoring System [J]. *IEEJ Transactions on Electronics, Information and Systems*, 2007, **127**: 1902-1908.

[125] ISTRATE D, BOUDY J, MEDJAHED H, et al. Medical remote monitoring using sound environment analysis and wearable sensors[M]. Rijeka, Croatia: IntechOpen, 2009: 1-16.

[126] **曹军,李传江,牟映峰.学生寝室纪律监管系统**.CN: 203012904U[P].2013-06-19.

[127] YUSUF S A, BROWN D J, MACKINNON A. Application of acoustic directional data for audio event recognition via HMM/CRF in perimeter surveillance systems[J]. *Robotics and Autonomous Systems*, 2015, **72**: 15-28.

[128] **张贤华.基于音频的老年人居家监护系统研究**[D].**杭州: 浙江大学**,2016.

[129] **张涛,苏春玲.一种用于枪声的多级检测识别技术**[J].**电子设计工程**,2013,**21**(18): 56-58.

[130] ARSLAN Y, GÜLDOĞAN B. Impulsive sound detection and gunshot recognition[C]// 2015 23nd Signal Processing and Communications Applications Conference (SIU). Malatya, Turkey: IEEE, 2015: 511-514.

[131] PENG L, YANG D, CHEN X. Multi frame size feature extraction for acoustic event detection[C]// Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific. Siem Reap, Cambodia: IEEE, 2014: 1-4.

[132] CLAVEL C, EHRETTE T, RICHARD G. Events detection for an audio-based surveillance system[C]// 2005 IEEE International Conference on Multimedia and Expo. Amsterdam, Netherlands: IEEE, 2005: 1306-1309.

[133] **叶剑杰.监督式分级异常声音检测系统的设计与实现**[D].**广州：华南理工大学**,2015.

[134] **栾少文.智能监控系统中公共场所异常声音检测的研究**[D].**重庆：重庆大学**,2009.

[135] TRIPATHI A M, BARUAH D, BARUAH R D. Acoustic sensor based activity recognition using ensemble of one-class classifiers[C]// 2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS). Douai, France: IEEE, 2015: 1-7.

[136] HUANG K Y, HSIA C C, TSAI M, et al. Activity recognition by detecting acoustic events for eldercare [C]// 6th World Congress of Biomechanics (WCB 2010). Singapore: Springer-Verlag Berlin Heidelberg, 2010: 1522-1525.

[137] **杜仲平,李一博,叶霆.基于音频监控的婴儿智能监护系统设计**[J].**计算机测量与控制**,2016,**24**(7): 105-108.

[138] LI Y, POPESCU M, HO K C, et al. Improving acoustic fall recognition by adaptive signal windowing [C]// 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Boston, USA: IEEE, 2011: 7589-7592.

[139] NISHI H, KIN K, KIMURA Y, et al. A watching system for aged people using acoustic informations [J]. *IEICE Technical Report Life Intelligence & Office Information Systems*, 2013,**113**(381): 81-84.

[140] GUYOT P, PINQUIER J, ANDRÉ-OBRECHT R. Water sound recognition based on physical models [C]// 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada: IEEE, 2013: 793-797.

[141] **谢圻.一种具有声音监听功能的智能电视**.CN: 104284224A[P].2015-01-14.

[142] **徐保民,李文婧.一种自适应的异常声音端点检测方法**[J].**软件导刊**,2017,**16**(8): 1-4.

[143] **郭延萍.特殊点声源声信号的检测和识别**[D].**南京：南京理工大学**,2014.

[144] LOZANO H, HERNÁEZ I, CAMARENA J, et al. A real-time sound recognition system in an assisted environment[M]. Berlin-Heidelberg, Germany: Springer-Verlag, 2012: 385-391.

[145] ŁOPATKA K, ZWAN P, CZY ŹEWSKI A. Dangerous sound event recognition using support vector machine classifiers[M]// Advances in Multimedia and Network Information System Technologies. Berlin Heidelberg, Germany: Springer-Verlag, 2010: 49-57.

[146] ALQAHTANI M O, MUHAMMAD G, ALOTAIBI Y A. Environment sound recognition using zero crossing features and MPEG-7 [C] // 2010 Fifth International Conference on Digital Information Management (ICDIM). Thunder Bay, ON, Canada: IEEE, 2010: 502-506.

[147] IWASA K, KUGLER M, KUROYANAGI S, et al. A sound localization and recognition system using pulsed neural networks on FPGA [C] // 2007 International Joint Conference on Neural Networks. Orlando, FL, USA: IEEE, 2007: 902-907.

[148] ZHU Z, HU X J, WANG J. Hierarchical structure neural network and it's application in unusual sound recognition[C]// Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693). Xi'an, China: IEEE, 2003,**5**: 3229-3231.

[149] **刘波.车辆音频特征分析及车型识别研究**[D].**武汉：武汉理工大学**,2007.

[150] **姜愉,赵荣阳,董振华.基于** MP **稀疏分解的发动机声音识别研究**[J].**软件导刊**,2015(3): 64-66.

[151] **刘明华.基于** MATLAB **声学检测刹车片材质的研究**[J].**科技传播**,2014(3): 96-97.

[152] **陈倩倩,周海平,陆永耕,等.一种地铁故障检测装置**.CN: 202853911U[P].2013-04-03.

[153] JARNICKI J, MAZURKIEWICZ J, MACIEJEWSKI H. Mobile object recognition based on acoustic information[C] // IECON'98. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (Cat. No. 98CH36200). Aachen, Germany: IEEE, 1998,**3**: 1564-1569.

[154] PAULRAJ M P,ABDUL H A,HEMA C R,et al . Acoustic Signature Recognition of Moving Vehicles using Elman Neural Network[J]. *Karpagam Jcs*, 2012,**6**(4): 183-189.

[155] WU H, SIEGEL M, KHOSLA P. Vehicle sound signature recognition by frequency vector principal component analysis[C]// IMTC/98 Conference Proceedings. IEEE Instrumentation and Measurement Technology Conference. Where Instrumentation is Going (Cat. No. 98CH36222). Paul, MN, USA: IEEE, 1998, 1: 429-434.

[156] VALERO X, ALÍAS F. Hierarchical classification of environmental noise sources considering the acoustic signature of vehicle pass-bys[J]. *Archives of Acoustics*, 2012,**37**(4): 423-434.

[157] MIYOSHI F, ASAHI K, OGAWA A. A Study on the Sound from the Own Car in the Recognition of Approaching Cars[J]. *IEICE Technical Report*, 2007,**107**(161): 27-32.

[158] **李云焕.基于声音识别的交通信息检测技术研究**[D].**西安: 长安大学**,2014.

[159] **刘桂林,孔祥维,刘航.基于无线传感器网络的车辆检测识别算法研究**[J].**传感器与微系统**,2010,**29**(2): 9-12.

[160] **张伟,谭国真,史慧敏,等.基于声敏传感器和盲信号处理的多车辆检测算法**[J].**计算机科学**,2009,**36**(10): 14-17.

[161] MUNICH M E. Bayesian subspace methods for acoustic signature recognition of vehicles[C]// 2004 12th European signal processing conference. Vienna, Austria: IEEE, 2004: 2107-2110.

[162] LU B, DIBAZAR A, BERGER T W. Nonlinear Hebbian Learning for noise-independent vehicle sound recognition[C]// 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). Hong Kong, China: IEEE, 2008: 1336-1343.

[163] **王慧.基于** GPS/GPRS **交通事故声音检测自动报警系统**[J].**赤峰学院学报(自然科学版)**,2012(18): 62-63.

[164] **戴硕.基于声信号处理的交通事故自动检测方法研究**[D].**合肥: 中国科学技术大学**,2010.

[165] **朱强华.行车噪声环境下的快速声学事件检测方法研究**[D].**哈尔滨: 哈尔滨工业大学**,2014.

[166] **罗向龙,高静怀,牛国宏,等.交通事件小波分解与支持向量机声频识别方法**[J].**交通运输工程学报**,2010(2): 116-121.

[167] **高晗,裴玉龙.基于车辆噪声时域特征的交通量统计方法**[J].**公路交通科技**,2008,**25**(4): 113-116121.

[168] **杨建国,王碧阳,石元杰,等.基于汽车怠速声音频谱的道路拥堵监测方法**.CN: 103247175A[P].2013-08-14.

[169] **陈宇峥.基于声信号检测的来车识别装置**.CN: 103714827A[P].2014-04-09.

[170] MASINO J, FOITZIK M J, FREY M, et al. Pavement type and wear condition classification from tire cavity acoustic measurements with artificial neural networks[J]. *The Journal of the Acoustical Society of America*, 2017,**141**(6): 4220-4229.

[171] **王楠.声振法检测混凝土路面脱空的信号处理方法研究**[D].**西安: 长安大学**,2012.

[172] FERGUSON E L, RAMAKRISHNAN R, WILLIAMS S B, et al. Convolutional neural networks for passive monitoring of a shallow water environment using a single sensor[C]// 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, USA: IEEE, 2017: 2657-2661.

[173] JAMES K R, DOWLING D R. A robust method for approximating acoustic field uncertainty in underwater sound channels[J]. *The Journal of the Acoustical Society of America*, 2007,**121**(5): 3076-3076.

[174] NIU H, OZANICH E, GERSTOFT P. Ship localization in Santa Barbara Channel using machine learning classifiers[J]. *The Journal of the Acoustical Society of America*, 2017,**142**(5): 455-460.

[175] MADEKIVI S. Experiments on automatic classification of shallow water acoustic signal sources using two pattern recognition methods[C]// ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing. New York, USA: IEEE, 1988: 2693-2696.

[176] TUMA M, IGEL C, PRIOR M. Hydroacoustic signal classification using kernel functions for variable feature sets[C]// 2010 20th International Conference on Pattern Recognition. Istanbul, Turkey: IEEE, 2010: 1011-1014.

[177] 崔秀华，胡国兵，崔金魁.鱼声信号检测系统的设计与实现[J].科学技术与工程，2012，**12**(25)：6416-6419.
[178] YAMADA I，YOKOTA A，YAMAMOTO K，et al. Acoustic recognition of aircraft types in flight[J]. *Journal of the Acoustical Society of Japan(E)*，1985，**6**(3)：203-213.
[179] 马宁，高勇.直升机声信号的检测和识别[J].太赫兹科学与电子信息学报，2006，**4**(3)：165-169.
[180] 雷丁，田燃，王柳彬，等.基于声纹识别的旋翼飞行器检测系统[J].电子世界，2017(12)：167.
[181] 杨琳.驾驶舱话音记录器背景声识别方法及实现[C]//大型飞机关键技术高层论坛暨中国航空学会2007年年会.深圳：中国航空学会，2007：253-257.
[182] 郭超.飞机舱音中微弱信号的检测与辨识[D].南京：南京航空航天大学，2009.
[183] 石磊.根据声音检测阀门漏泄的方法[J].发电设备，1991(3)：31-33.
[184] 方宇昌.关于输水管线智能听漏仪的研制及漏水声信号处理的研究[D].沈阳：沈阳工业大学，2005.
[185] 王强.基于声信号检测的管道TPD预警系统研究[D].杭州：浙江大学，2005.
[186] AI C，SUN X，ZHAO H，et al. Pipeline damage and leak sound recognition based on HMM[C]//2008 7th World Congress on Intelligent Control and Automation. Chongqing，China：IEEE，2008：1940-1944.
[187] 李明霞，路莹.基于声音识别的多阀门泄漏检测系统[J].大连工业大学学报，2008，**27**(1)：87-90.
[188] 李明霞.基于声音识别的新型输油泵阀门泄露检测原型系统[D].大连：大连工业大学，2008.
[189] 崔尧尧.基于声传感器阵列的管道内检测器追踪定位系统关键技术研究[D].天津：天津大学，2012.
[190] BOUCHARD J G，BEESLEY M J，FARRINGTON D C F. Acoustic diagnostics in process plant：application of pattern recognition techniques[C]// International Conference on Acoustic Sensing and Imaging，1993. London，UK：IET，1993：189-194.
[191] 甄彤，董志杰，郭嘉，等.基于声音的储粮害虫检测系统设计[J].河南工业大学学报(自然科学版)，2012，**33**(5)：79-82.
[192] 郭敏，尚志远.储粮害虫声信号的检测和应用[J].物理，2001，**30**(1)：39-42.
[193] 方旭君.基于声音和振动的玉米储存中害虫检测方法研究[D].长春：吉林大学，2012.
[194] 董志杰.基于声信号的储粮害虫检测技术研究[D].郑州：河南工业大学，2013.
[195] ORTIZ J. Identifying Mechanical Damage using Sound Samples. US：5884264[P]. 1999-03-16.
[196] SHIBATA A，KONISHI M，ABE Y，et al. Recognition of facility sound with background noises[C]// 2009 7th Asian Control Conference. Hong Kong，China：IEEE，2009：887-892.
[197] 瞿金秀，杨飞宇，张周锁，等.基于声音信号的结构损伤识别方法[J].振动、测试与诊断，2014，**34**(4)：638-643.
[198] 王艳丽.SKF托辊声音监测装置快速安全可靠地检测输送带托辊故障[J].水泥，2011(4)：56-56.
[199] WU H，YAN Q，LING X. Survey on fault diagnosis of diesel engine based on vibration signal[C]// 2017 3rd International Conference on Information Management (ICIM). Chengdu，China：IEEE，2017：494-495.
[200] 张永肃，高宝成.基于Linux系统的汽车噪声故障诊断系统[J].电子测量技术，2011，**34**(12)：80-83.
[201] MADAIN M，AL-MOSAIDEN A，AL-KHASSAWENEH M. Fault diagnosis in vehicle engines using sound recognition techniques[C]// 2010 IEEE International Conference on Electro/Information Technology. Normal，IL，USA：IEEE，2010：1-4.
[202] 任志英.基于声信号技术的发动机故障诊断系统研究[D].福州：福州大学，2006.
[203] 刘帅师.基于发动机声响信号分析的转速估计及异响特征提取研究[D].长春：吉林大学，2006.
[204] 吴道虎，王绪军.船舶发动机噪声与振动监测系统的设计[J].电子设计应用，2004(2)：139-140.
[205] 徐朴.内燃机故障声响信号时间序列模型诊断方法[J].中南林学院学报，1989，**9**(1)：77-84.
[206] FIGLUS T. The application of a continuous wavelet transform for diagnosing damage to the timing chain tensioner in a motorcycle engine[J]. *Journal of Vibroengineering*，2015，**17**(3)：1286-1294.
[207] ANAMI B S，PAGI V B. Acoustic signal based detection and localisation of faults in motorcycles[J]. *IET Intelligent Transport Systems*，2013，**8**(4)：345-351.
[208] NAVEA R F，SYBINGCO E. Design and implementation of an acoustic-based car engine fault diagnostic system in the android platform[C]// International Research Conference in Higher Education(IRCHE). Manila，Philippines，USA：Polytechnic University of the Philippines，2013.

[209] 周小龙,马风雷.改进希尔伯特-黄变换的发动机轴瓦故障诊断[J].机械设计与制造,2016(11):71-75.
[210] 侯文魁,吕川,刘红梅,等.基于连续小波消噪的发动机异响提取及诊断[J].华中科技大学学报,2009,**37**(1):160-163.
[211] 梁春艳,李志雄.基于频谱分析的发动机断缸异响检测[J].轻工科技,2016(9):107-108.
[212] 齐伟.基于声信号的汽车发动机机械异响故障诊断方法研究[D].苏州:苏州大学,2016.
[213] 曹昳劼.基于压力、振动、声信号的压气机喘振故障诊断和监测[D].上海:上海交通大学,2010.
[214] 吕琛,王桂增,叶昊.基于噪声测量的主轴承间隙状态监测[J].振动与冲击,2003,**22**(3):33-36.
[215] 阚磊.小波去噪和概率神经网络在发动机声信号识别中的应用[D].重庆:重庆大学,2016.
[216] GEIB A F, KUO C C, GAWECKI M, et al. MFCC and CELP to detect turbine engine faults. US: 8655571[P]. 2014-2-18.
[217] 陈宇鹏.基于发动机工作噪声的故障诊断研究[D].太原:中北大学,2010.
[218] 刘福.基于声音的发动机诊断方法研究与实践[D].沈阳:东北大学,2005.
[219] 史忠科.基于振动和音频信息的汽车发动机故障诊断系统.CN:102103036A[P].2011-6-22.
[220] LI W, GU F, BALL A D, et al. A study of the noise from diesel engines using the independent component analysis[J]. *Mechanical Systems and Signal Processing*, 2001,**15**(6):1165-1184.
[221] 高志彬,刘尊民.基于傅立叶变换的发动机缺缸故障检测[J].农业装备与车辆工程,2009(1):28-30.
[222] PAULRAJ M P, SAZALI Y, ZUBIR M, et al. Motorbike engine faults diagnosing system using entropy and functional link neural network in wavelet domain[J]. *Proceedings of the International Conference on Man-Machine Systems*, 2009,**2**(4):1-5.
[223] 俊吕,杨祖元,谢胜利.基于稀疏表示的汽车发动机故障诊断系统和方法.CN:101382468A[P].2009-3-11.
[224] 穆峰.基于声信号的故障诊断研究及应用[D].济南:山东大学,2016.
[225] 梁中远.基于盲源分离的船舶柴油机噪声分离研究[D].大连:大连海事大学,2016.
[226] ALHOULI Y, ALKHALEDI A, ALZAYEDI A, et al. Acoustic measurements in diesel engine condition monitoring[J]. *International Journal of Engineering Science and Innovative Technology*, 2016,**5**(6):1-8.
[227] 杨毫鸽,孙成立.基于 GMM-UBM 的飞机发动机声音识别方法研究[J].计算机科学与应用,2017,**7**(8):781-787.
[228] LEE J S. Engine fault diagnosis using sound source analysis based on hidden Markov mode[J]. 한국통신학회논문지, 2014,**39**(5):244-250.
[229] 徐光华,贾维银,侯成刚,等.基于轨迹平行测量的发动机异响诊断方法[J].西安交通大学学报,2002,**36**(5):519-522.
[230] MARTINOD R M, BETANCUR G R, CASTAÑEDA L F, et al. Estimation of combustion engine technical state by multidimensional analysis using SVD method[J]. *International Journal of Vehicle Systems Modelling and Testing*, 2013,**8**(2):105-118.
[231] 唐浩,屈梁生.基于支持向量机的发动机故障诊断[J].西安交通大学学报,2007,**41**(9):1124-1126.
[232] ROMA G, HERRERA P, NOGUEIRA W. Environmental sound recognition using short-time feature aggregation[J]. *Journal of Intelligent Information Systems*, 2018,**51**(3):457-475.
[233] 谢政.基于切削声信号的刀具状态识别研究[D].上海:上海交通大学,2008.
[234] TRABELSI H, KANNATEY-ASIBU E. Pattern-recognition analysis of sound radiation in metal cutting[J]. *The International Journal of Advanced Manufacturing Technology*, 1991,**6**(3):220-231.
[235] FRIGIERI E P, BRITO T G, YNOGUTI C A, et al. Pattern recognition in audible sound energy emissions of AISI 52100 hardened steel turning: A MFCC-based approach[J]. *The International Journal of Advanced Manufacturing Technology*, 2017,**88**(5-8):1383-1392.
[236] XU J, WANG Z, WANG J, et al. Acoustic-based cutting pattern recognition for shearer through fuzzy c-means and a hybrid optimization algorithm[J]. *Applied Sciences*, 2016,**6**(10):294.
[237] UBHAYARATNE I, PEREIRA M P, XIANG Y, et al. Audio signal analysis for tool wear monitoring in sheet metal stamping[J]. *Mechanical Systems and Signal Processing*, 2017,**85**:809-826.

[238] 李炜恒，王林生.声纹识别技术在金刚石压机顶锤防护中的应用[J].金刚石与磨料磨具工程，2013，**33**(3)：71-74.
[239] 张欢欢.基于声信号频率特征的铝合金脉冲 GTAW 熔透识别方法研究[D].上海：上海交通大学，2016.
[240] 张海滨.列车轴承轨边声学故障信号的声源分离及其去噪研究[D].合肥：中国科学技术大学，2016.
[241] 刘洋，张栓，张桐.航空发动机轴承噪声检测技术分析研究[C]//第十五届中国科协年会.贵阳：中国科学技术学会，2013：1-6.
[242] 王前，于嘉成，宁永杰，等.基于 MFCC 与 PCA 的滚动轴承故障诊断[J].组合机床与自动化加工技术，2017(12)：103-105.
[243] 蒋杰.基于轨旁声学信号的城轨列车滚动轴承故障诊断研究[D].南京：南京理工大学，2017.
[244] 梁莹.声频传感器检测轴承[J].中国铁路，1990(6)：32.
[245] NAGY K，DOUSIS D A，FINCH R D.铁路车轮的特征声探伤[J].国外机车车辆工艺，1980(3)：42-49.
[246] 马思乐，李现明，王海相，等.基于声信号处理的瓶盖密封性检测装置及方法.CN：101929913A[P].2010-12-29.
[247] 巩小东.啤酒瓶在线检测相关技术的研究[D].青岛：山东科技大学，2011.
[248] GLOWACZ A. Recognition of acoustic signals of synchronous motors with the use of MoFS and selected classifiers[J]. *Measurement Science Review*，2015，**15**(4)：167-175.
[249] GLOWACZ A. Diagnostics of dc machine based on sound recognition with application of LPC and GSDM [J]. *Przeglad Elektrotechniczny*，2010，**86**(6)：243-246.
[250] GŁOWACZ A. Sound recognition of induction motor with the use of discrete Meyer wavelet transform and classifier based on words[J]. *Przeglad Elektrotechniczny*，2013，**89**(6)：152-154.
[251] 李明超.基于异音检测的电机故障诊断方法[D].江门：五邑大学，2014.
[252] 刘力源.基于机器学习方法的电机异音检测研究[D].江门：五邑大学，2014.
[253] GLOWACZ A. Recognition of acoustic signals of loaded synchronous motor using FFT，MSAF-5 and LSVM[J]. *Archives of Acoustics*，2015，**40**(2)：197-203.
[254] 宗银雪，张靓，李铁军，等.机电设备故障音频特征提取方法研究[J].仪表技术与传感器，2016(4)：105-107.
[255] 李伟鑫.基于音频识别的防爆电机故障监测研究[D].大庆：东北石油大学，2015.
[256] 李春雷，董志学，王新杰.发电机声音检测与故障诊断研究[J].内蒙古工业大学学报，2015，**34**(3)：201-208.
[257] 王新杰，董志学，潘颖辉.机车牵引电机声音检测与故障诊断系统应用研究[J].计算机应用与软件，2016，**33**(10)：103-107.
[258] 李敏，董志学.基于 Android 的嵌入式机器声音故障检测系统的设计与实现[J].计算机应用与软件，2013，**30**(7)：301-304.
[259] 莫慧芳，谷爱昱，饶明辉，等.基于小波包能量相对熵的电机振声信号故障检测[J].煤矿机械，2014，**35**(3)：231-233.
[260] 张宇波.声信号识别研究及在机械运行状态预测中的应用[D].长沙：湖南大学，2005.
[261] 李军，李伟鑫，李建平.基于声学特征的化工防爆电机在线监测[J].计算机与现代化，2014(12)：103-106.
[262] 赵书涛，李沐峰，王亚潇，等.断路器操动状态声音辨识的优化算法的研究[J].电测与仪表，2017，**54**(10)：26-31.
[263] 邱卫东.配电变压器运行中的声音检查[J].农村电工，2003，11(1)：31-31.
[264] 范平安，施秀琴，丁玮，等.一种基于声波检测的故障电弧检测装置.CN：203502548U[P].2014-3-26.
[265] 孟繁庆，王红，杨士元，黄立明.基于电弧声音检测的电气故障监测方法[J].计算机科学，2013，**40**(7)：78-82.
[266] 杜世斌.基于音频特征的电气设备故障监测算法研究[D].济南：山东大学，2014.
[267] 张康荣.基于盲源分离的设备故障音检测算法与应用[D].济南：山东大学，2016.
[268] 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.
[269] 陈梦晓.基于声音信号的断路器故障诊断算法及系统设计[J].仪器仪表与分析监测，2017(3)：21-24.
[270] 高宏亮，王世成，翟国富.基于 PIND 声音脉冲分类的航天继电器多余物特征识别方法的研究[J].机电元件，2007，**27**(3)：7-12.

[271] 李晶,孙农亮,滕升华.基于声音识别的设备状态检测算法[J].信息技术,2015(6):94-98.
[272] 吕鑫.环锭纺细纱断头检测技术研究[D].天津:天津工业大学,2014.
[273] 金冉,杨风暴,韩焱.声激励检测信号的特征分析[J].测试技术学报,2002,**16**(4):245-248.
[274] 卢昆,文丹华,李运春,等.一种通过声音检测非晶合金产品合格性的方法.CN:102435676B[P].2013-12-18.
[275] 莫正鹏,张常年,王泽来.基于碰撞声音的氧化铝熟料质量检测系统设计[J].陕西科技大学学报,2011,**29**(6):64-66.
[276] 王泽来,张常年,莫正鹏.基于声音的氧化铝烧结工艺中异常状态的检测[J].数据采集与处理,2012(s1):143-146.
[277] 胡春松.基于 DSP 的铝电解槽声音检测和处理系统的设计与实现[D].北京:北方工业大学,2005.
[278] 喻国丽.基于声信号的铝锭脱模故障诊断研究[D].兰州:兰州理工大学,2016.
[279] 秦志英,齐康花,董桂西,等.基于声音信号的钢材材质检测及试验研究[J].河北科技大学学报,2016,**37**(3):275-282.
[280] 秦志英,刘尧,董桂西,等.利用声信号能量比在线识别钢材材质[J].机械科学与技术,2016,**35**(5):800-804.
[281] YANG L, KANG H S, ZHOU Y C, et al. Intelligent discrimination of failure modes in thermal barrier coatings: wavelet transform and neural network analysis of acoustic emission signals[J]. *Experimental Mechanics*, 2015,**55**(2): 321-330.
[282] 张冰瑞,陈克安,马玺越.冲击声的特征提取及其在声源材料识别中的应用[J].噪音与振动控制,2012(S1):152-156.
[283] 王芳.车用柴油机声品质主观评价的分析研究[D].天津:天津大学,2011.
[284] 黄海波,黄晓蓉,苏瑞强,等.基于 EEMD 与 GA-小波神经网络的传动系声品质预测[J].振动与冲击,2017,**36**(9):130-137.
[285] 孙慧慧.基于 GA-BP 神经网络的车内声品质评价研究[D].长春:吉林大学,2012.
[286] KUBO N, MELLERT V, WEBER R, et al. Categorisation of engine sound[C]// INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Prague, Czech: Institute of Noise Control Engineering, 2004,**2004**(5): 2284-2291.
[287] 陈家焱.鸭蛋破损声检与分级技术的研究[D].武汉:华中农业大学,2004.
[288] 郭小军,张海东,王孟,等.基于声学特性和 Labview 的鸡蛋裂纹检测研究[J].湖北农业科学,2017,**56**(21):4131-4136.
[289] 曹晏飞,陈红茜,滕光辉,等.基于功率谱密度的蛋鸡声音检测方法[J].农业机械学报,2015,**46**(2):276-280.
[290] 姜勇,郭文川.基于 DSP 的鸡蛋蛋壳破损检测系统硬件设计[J].农机化研究,2008(8):103-105.
[291] 王树才,任奕林,陈红,等.利用敲击声信号进行禽蛋破损检测和模糊识别[J].农业工程学报,2004,**20**(4):130-133.
[292] 刘俭英,田茂胜,文友先.基于 DSP 的鸡蛋破损检测方法[J].华中农业大学学报,2008,**27**(6):807-810.
[293] 黄耀志,余劼.基于神经网络分析的鲜蛋破损检测[J].振动、测试与诊断,2004,**23**(3):205-209.
[294] WANG J, ZHANG K, MADANI K, et al. Multi-scale feature based salient environmental sound recognition for machine awareness[C]// 2014 IEEE 6th International Conference on Awareness Science and Technology (iCAST). Paris, France: IEEE, 2014: 1-6.
[295] JANVIER M, ALAMEDA-PINEDA X, GIRINZ L, et al. Sound-event recognition with a companion humanoid[C]// 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012). Osaka, Japan: IEEE, 2012: 104-111.
[296] NIWA T, KAWAKAMI T, OOE R, et al. An Acoustic Events Recognition for Robotic Systems based on a Deep Learning Method[J]. *Journal of Computer & Communications*, 2015(3): 46-51.
[297] 赵殿全,李金龙,谢蓓敏.基于阀厅智能巡检机器人的设备声音故障识别算法[J].电子设计工程,2016,**24**(21):63-65.
[298] REN J, JIANG X, YUAN J, et al. Sound-event classification using robust texture features for robot hearing[J]. *IEEE Transactions on Multimedia*, 2017,**19**(3): 447-458.

[299] 孙开源，曹阳，李燕羽，等.黄粉虫成虫爬行和咬食活动声音采集分析研究[J].河南工业大学学报(自然科学版)，2008，**29**(6)：39-44.
[300] 贾琪.基于声信号处理的农业虫害识别系统[D].济南：山东大学，2016.
[301] 陈爱武，郭丙琴，李荣.音频分析在自动喷雾技术方面的应用[J].湖南科技学院学报，2016，**37**(10)：23-25.
[302] PYRAH D. 检测昆虫嚼食声音的装置[J].*Agricultural Research*，1985，33(4)：13-15.
[303] 杨丽.基于信号处理的小麦品质声学检测方法研究[D].郑州：河南工业大学，2007.
[304] 李齐超.小麦硬度声学测定方法的研究[D].郑州：河南工业大学，2011.
[305] 王薇.小麦硬度声学测定方法的优化研究[D].郑州：河南工业大学，2014.
[306] PHOOPHUANGPAIROJ R. Computerized unripe and ripe durian striking sound recognition using Syllable-based HMMs[J]. *Applied Mechanics & Materials*，2013，**446/447**：927-935.
[307] PHOOPHUANGPAIROJ R. Durian ripeness striking sound recognition using N-gram models with N-best lists and majority voting[M]// Recent Advances in Information and Communication Technology. Cham，Switzerland：Springer，2014：167-176.
[308] 肖珂，高冠东，滕桂法，等.西瓜成熟度音频无损检测技术[J].农机化研究，2009，**31**(8)：150-152.
[309] 王再欢，韩鹏，莫斌，等.基于声音识别的森林盗伐传感器设计与实现[J].传感器与微系统，2013，**32**(8)：102-104.
[310] 王再欢，唐云建，韩鹏.一种利用声音识别的森林盗伐检测方法[J].计算机工程与应用，2012，**48**(30)：216-219.
[311] BRANDSTETTER M，HÜBNER S. Bioacoustic—A method to detect woodboring insects[J]. *Waldwissen Author Translation of Article in Fortschutz Aktuell*，2015，**60/61**：31-36.
[312] 赵源吉.双条杉天牛幼虫声信号监测技术研究[D].北京：北京林业大学，2009.
[313] 祁骁杰.蛀干害虫幼虫声信号特征及其影响因素研究[D].北京：北京林业大学，2016.
[314] 汪开英，赵晓洋，何勇.畜禽行为及生理信息的无损监测技术研究进展[J].农业工程学报，2017，**33**(20)：197-209.
[315] 袁瑞临，张栖铭，王峰，等.基于 HTK 的嵌入式猪只声音识别系统设计[J].电脑知识与技术，2017，**13**(4)：186-188.
[316] 张苏楠.基于视频跟踪与多模型声音识别的猪行为检测与分析[D].太原：太原理工大学，2016.
[317] 张栖铭，袁瑞临，范凡，等.基于 SVM 算法的猪声音识别的研究[J].电脑知识与技术，2017，**13**(10)：162-164.
[318] 张苏楠，王芳，阎高伟，等.基于隐马尔科夫模型的猪只状态自动识别技术[J].黑龙江畜牧兽医，2016(11)：97-99.
[319] 龚永杰，黎煊，高云，等.基于矢量量化的猪咳嗽声识别[J].华中农业大学学报，2017，**36**(3)：119-124.
[320] 吴佩贤，王瑞荣，王建中，等.基于音频识别的钱塘江潮涌实时监测技术[J].机电工程，2009，**26**(4)：74-76.
[321] 韩英奎，赵金涛.基于声音高炮作业用弹量检测电路设计[J].吉林气象，2014(2)：34-35.
[322] LINDSTRÖM K，LUGLI M. A quantitative analysis of the courtship acoustic behaviour and sound patterning in male sand goby，Pomatoschistus minutus[J]. *Environmental Biology of Fishes*，2000，**58**(4)：411-424.
[323] ERISMAN B E，ROWELL T J. A sound worth saving：Acoustic characteristics of a massive fish spawning aggregation[J]. *Biology letters*，2017，**13**(12)：1-5.
[324] 韩春琏，陕方，曹家林，等.大熊猫"唔"叫声的时域和频域特征[J].声学学报，2000，**25**(4)：364-370.
[325] 魏静明.用时频纹理特征识别特定环境声音[D].福州：福州大学，2014.
[326] QIAN K，ZHANG Z，BAIRD A，et al. Active learning for bird sound classification via a kernel-based extreme learning machine[J]. *The Journal of the Acoustical Society of America*，2017，**142**(4)：1796-1804.
[327] 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.
[328] CHENG J，XIE B，LIN C，et al. A comparative study in birds：Call-type-independent species and individual recognition using four machine-learning methods and two acoustic features[J]. *Bioacoustics*，2012，**21**(2)：157-171.

[329] DUAN S, TOWSEY M, ZHANG J, et al. Acoustic component detection for automatic species recognition in environmental monitoring [C] // 2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing. Adelaide, Australia: IEEE, 2011: 514-519.

[330] RAZALI M H M, AZMI N, ZAKARIA A, et al. Sound analyser for bioacoustic monitoring system using LabVIEW [C] // 2014 International Conference on Computer, Communications, and Control Technology (I4CT). Langkawi, Malaysia: IEEE, 2014: 434-437.

[331] BEECHER M D, MEDVIN M B, STODDARD P K, et al. Acoustic adaptations for parent-offspring recognition in swallows [J]. *Experimental Biology*, 1986, **45**(3): 179-193.

[332] **王岩.基于动物叫声的物种识别技术的研究** [D].**哈尔滨: 东北林业大学**, 2008.

[333] **皮阳.基于声音的生物种群识别** [D].**成都: 电子科技大学**, 2015.

[334] **余清清.噪声环境下基于时-频特征的生态环境声音的分类** [J].**计算机与数字工程**, 2017, **45**(1): 8-14.

[335] NIKOLICH K, FROUIN-MOUY H, ACEVEDO-GUTIÉRREZ A. Quantitative classification of harbor seal breeding calls in Georgia Strait, Canada [J]. *The Journal of the Acoustical Society of America*, 2016, **140**(2): 1300-1308.

[336] ERBS F, ELWEN S H, GRIDLEY T. Automatic classification of whistles from coastal dolphins of the southern African subregion [J]. *The Journal of the Acoustical Society of America*, 2017, **141**(4): 2489-2500.

[337] KEEN S C, SHIU Y, WREGE P H, et al. Automated detection of low-frequency rumbles of forest elephants: A critical tool for their conservation [J]. *The Journal of the Acoustical Society of America*, 2017, **141**(4): 2715-2726.

[338] URAZGHILDIIEV I, CLARK C W, KREIN T. Acoustic detection and recognition of fin whale and North Atlantic right whale sounds [C] // 2008 New Trends for Environmental Monitoring Using Passive Systems. Hyeres, French Riviera, France: IEEE, 2008: 1-6.

[339] **叶蓁,孙海信,颜佳泉,等.鲸类声信号的分类系统设计** [J].**厦门大学学报(自然科学版)**, 2017, **56**(1): 144-148.

[340] TOMASINI M, SMART K, MENEZES R, et al. Automated robust Anuran classification by extracting elliptical feature pairs from audio spectrograms [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 2517-2521.

[341] KAEWTIP K, ALWAN A, O'REILLY C, et al. A robust automatic birdsong phrase classification: A template-based approach [J]. *The Journal of the Acoustical Society of America*, 2016, **140**(5): 3691-3701.

[342] STOWELL D, BENETOS E, GILL L F. On-bird sound recordings: automatic acoustic recognition of activities and contexts [J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017, **25**(6): 1193-1206.

[343] de OLIVEIRA A G, VENTURA T M, Ganchev T D, et al. Bird acoustic activity detection based on morphological filtering of the spectrogram [J]. *Applied Acoustics*, 2015, **98**: 34-42.

[344] BREGMAN M R, PATEL A D, GENTNER T Q. Songbirds use spectral shape, not pitch, for sound pattern recognition [J]. *Proceedings of the National Academy of Sciences*, 2016, **113**(6): 1666-1671.

[345] BOULMAIZ A, MESSADEG D, DOGHMANE N, et al. Robust acoustic bird recognition for habitat monitoring with wireless sensor networks [J]. *International Journal of Speech Technology*, 2016, **19**(3): 631-645.

[346] **王浩安.基于多层能量检测的动物声音检测与识别** [D].**福州: 福州大学**, 2013.

[347] **陈莎莎.基于时频纹理和随机森林的鸟声识别** [D].**福州: 福州大学**, 2013.

[348] **蒋锦刚,邵小云,万海波,等.基于语谱图特征信息分割提取的声景观中鸟类生物多样性分析** [J].**生态学报**, 2016, **36**(23): 7713-7723.

[349] GANCHEV T, POTAMITIS I. Automatic acoustic identification of singing insects [J]. *Bioacoustics*, 2007, **16**(3): 281-328.

[350] POTAMITIS I, GANCHEV T, FAKOTAKIS N. Automatic acoustic identification of insects inspired

by the speaker recognition paradigm [C] // InterSpeech-2006. Pittsburgh, PA, USA: ICSLP, 2006: 2126-2129.

[351] LE-QING Z. Insect sound recognition based on mfcc and pnn [C] // 2011 International Conference on Multimedia and Signal Processing. Guilin, Guangxi, China: IEEE, 2011, **2**: 42-46.

[352] CHESMORE E D, NELLENBACH C. Acoustic Methods for the Automated Detection and Identification of Insects [J]. *ACTA Horticulturae*, 2001, **562**: 223-231.

[353] **罗茜，王鸿斌，张真，等.基于** MFCC **和** BP **神经网络的小蠹声音种类自动鉴别** [J].**北京林业大学学报**, 2011, **33**(5): 81-85.

[354] MORFI V, STOWELL D. Deductive refinement of species labelling in weakly labelled birdsong recordings [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 656-660.

[355] **李泽鑫.基于** ICA **算法的野外动物声音探测仪设计** [J].**求知导刊**, 2017(25): 57-57.

[356] **周晓敏，李应.基于** Radon **和平移不变性小波变换的鸟类声音识别** [J].**计算机应用**, 2014, **34**(5): 1391-1396.

[357] SCHÖNEICH S, KOSTARAKOS K, HEDWIG B. An auditory feature detection circuit for sound pattern recognition [J]. *Science Advances*, 2015, **1**(8): e1500325.

[358] YUAN C L T, RAMLI D A. Frog sound identification system for frog species recognition [C] // 2012 International Conference on Context-Aware Systems and Applications (ICCASA). Ho Chi Minh City, Vietnam: Springer-Verlag, 2012: 41-50.

[359] STROUT J, ROGAN B, SEYEDNEZHAD S M M, et al. Anuran call classification with deep learning [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 2662-2665.

[360] ISNARD V, TAFFOU M, VIAUD-DELMON I, et al. Acoustic and auditory sketches: Recognition of severely simplified natural sounds by human listeners [J]. *The Journal of the Acoustical Society of America*, 2015, **138**(3): 1918-1918.

[361] ZHU L Q, ZHANG Z. Insect sound recognition based on SBC and HMM [C] // 2010 International Conference on Intelligent Computation Technology and Automation. Changsha, China: IEEE, 2010, **2**: 544-548.

[362] XIAN Y, PU Y, GAN Z, et al. Adaptive DCTNet for audio signal classification [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 3999-4003.

[363] KOLUGURI N R, MEENAKSHI G N, GHOSH P K, et al. Spectrogram Enhancement Using Multiple Window Savitzky-Golay MWSG Filter for Robust Bird Sound Detection [J]. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 2017, **25**(6): 1183-1192.

[364] SALAMON J, BELLO J P, FARNSWORTH A, et al. Fusing shallow and deep learning for bioacoustic bird species classification [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 141-145.

[365] NARASIMHAN R, FERN X Z, RAICH R. Simultaneous segmentation and classification of bird song using CNN [C] // 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 146-150.

[366] **王恩泽.基于鸣声的鸟类智能识别方法研究** [D].**杨凌：西北农林科技大学**, 2014.

[367] POTAMITIS I, NTALAMPIRAS S, JAHN O, et al. Automatic bird sound detection in long real-field recordings: Applications and tools [J]. *Applied Acoustics*, 2014, **80**: 1-9.

[368] **杨三伟.基于** OSALPCC **与** SVM **的工程机械声音识别算法研究** [D].**杭州：杭州电子科技大学**, 2016.

[369] **赵拓.基于频谱动态特征和** ELM **的挖掘设备识别方法研究** [D].**杭州：杭州电子科技大学**, 2016.

[370] **王玮.基于作业声信号统计分析的工程机械识别研究** [D].**杭州：杭州电子科技大学**, 2016.

[371] **张旭清，刘文白，孔戈，等.基于声信号的抹灰墙质量检测及试验研究** [J].**住宅科技**, 2017, **37**(1): 45-49.

[372] **王子娟.基于声音的烧砖内部缺陷检测** [D].**沈阳：东北大学**, 2008.

[373] BILSKI P, BOBIŃSKI P, KRAJEWSKI A, et al. Detection of wood boring insects' larvae based on the acoustic signal analysis and the artificial intelligence algorithm[J]. *Archives of Acoustics*, 2017, **42**(1): 61-70.

[374] **赵义湘,赵鲁军,杨先碧.听声识别问题建筑**[J].**建筑工人**,2008(1): 56-56.

[375] **夏文鹤,潘硕,孟英峰,等.基于音频信号的气体钻井返出岩屑量监测方法研究**[J].**石油钻探技术**,2017, **45**(3): 121-126.

[376] TABACCHI M, ASENSIO C, PAVÓN I, et al. Water boiling stages classification using acoustic features. Toward a cooking appliance monitoring and control[C]// INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Innsbruck, Austria: Institute of Noise Control Engineering, 2013, **247**(6): 2323-2329.

[377] **刘军.一种电磁炉的水沸腾检测装置**.CN: 102128677A[P].2011-07-20.

[378] **刘力源,孙博,吴川辉,等.风扇异音检测系统**.CN: 105588636A[P].2016-05-18.

[379] **方献良,茅忠群,诸永定.一种智能吸油烟机以及该吸油烟机的控制方法**.CN: 103673008A[P].2014-03-26.

[380] **黎祥英.一种带有保健监测的手表**.CN: 102608906A[P].2012-07-25.

[381] AMFT O, KUSSEROW M, TRÖSTER G. Bite weight prediction from acoustic recognition of chewing [J]. *IEEE Trans Biomed Engineering*, 2009, **56**(6): 1663-1672.

[382] KORUCU M K, KAPLAN Ö, BÜYÜK O, et al. An investigation of the usability of sound recognition for source separation of packaging wastes in reverse vending machines[J]. *Waste Management*, 2016, **56**: 46-52.

[383] **邱斌,李占锋,王顺利,等.基于声音的纸张柔软度检测方法研究**[J].**计算机测量与控制**,2014, **22**(6): 1979-1983.

[384] **梁康宁,何迁,杨苑森.一种日用陶瓷裂纹检测装置**.CN: 106153726A[P].2016-11-23.

[385] **佚名.根据不同声音类型识别煤岩体地震冲量**[J].**煤矿安全**,2008(10): 10.

[386] **张瑞兴.基于脚步声身份识别的算法研究**[D].**西安: 陕西师范大学**,2014.

[387] **余瑶.基于局部频谱特征的脚步声识别算法研究**[D].**西安: 陕西师范大学**,2014.

[388] ASAKURA D, NISHINO T, NARUSE H. Writer recognition with a sound in hand-writing[J]. *The Journal of the Acoustical Society of America*, 2016, **140**(4): 3374-3375.

[389] **董伟.特征提取及特征优选在车辆声识别中的应用研究**[D].**太原: 中北大学**,2010.

[390] **董伟,王红亮,黄洋文.基于多重特征提取的战场车辆声目标识别**[J].**传感器与微系统**,2010, **29**(7): 30-32.

[391] **雷鸣,乔柯.低空目标被动声识别关键技术研究**[J].**计算机与数字工程**,2017, **45**(4): 645-649.

[392] LIU H, YANG J, CHEN H. A novel approach research on low altitude passive acoustic target recognition based on ICA and HMM[C]// 2008 Fourth International Conference on Natural Computation. Jǐnan, China: IEEE, 2008, **5**: 371-375.

[393] **曹慧敏.基于海上侦察系统的声音识别技术研究**[D].**南京: 南京理工大学**,2010.

[394] **聂晓华,周庆军,王毅等.雷达辐射源音频信号声纹识别方法**[J].**探测与控制学报**,2009, **31**(3): 9-13.

[395] FERGUSON B G, WYBER R J. Detection and localisation of a ground based impulsive sound source using acoustic sensors onboard a tactical unmanned aerial vehicle[R]// Defence Science and Technology Organisation Edinburgh (Australia) Maritime Operations Div. Neuilly-sur-Seine, Franc: Science and Technology Organization, 2006.

[396] SÁNCHEZ-HEVIA H A, AYLLÓN D, GIL-PITA R, et al. Maximum likelihood decision fusion for weapon classification in wireless acoustic sensor networks[J]. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 2017, **25**(6): 1172-1182.

[397] **田昊,唐力伟,陈红,等.基于瞬态声与阶次倒谱的齿轮箱故障诊断**[J].**振动、测试与诊断**,2009, **29**(2): 137-140.

[398] **丁明理,吕飞,杨冬梅,等.火炮发射次数检测装置及检测方法**.CN: 102410781A[P].2012-04-11.

[399] **陆军仁,柏逢明,王俊平.基于** PNN **的音频检测硬度分类器研究**[J].**长春理工大学学报(自然科学版)**, 2009, **32**(1): 82-84.

Understanding Digital Audio—A Review of General Audio/Ambient Sound based Computer Audition

LI Wei[1,2], LI Shuo[1]

(1. *School of Computer Science and Technology, Fudan University, Shanghai* 201203, *China*;
2. *Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai* 200433, *China*)

Abstract: Sound is one of the main sources for human to obtain information. It is of great importance to automatically analyze and understand the content of sound. This paper introduces the basic knowledge of sound, classifies various sounds from three digital perspectives, i.e., signal, auditory perception and sound characteristics, and clarifies the relationship between each class, as well as makes clear the research object and discipline position of general audio based Computer Audition(CA). Next, we describe the basic concept, principles, research topics, and technical framework of computer audition. Typical applications of computer audition are comprehensively summarized in various fields, including health care, safety protection, transportation, storage, manufacturing, agriculture, forestry, animal husbandry, fishery, water conservancy, environment protection, construction, mining, daily life, identification, military etc. With each kind of CA application, the basic principles and technical route in typical papers are outlined. Finally, we analyze the problems that hinder the development of CA, and prospect the bright future.

Keywords: digital audio; general audio/ambient sound; computer audition; audio signal processing; artificial intelligence