第 37 卷 第 22 期
Vol.37 No.22

计 算 机 工 程
Computer Engineering

2011 年 11 月
November 2011

·开发研究与设计技术· 文章编号：1000—3428(2011)22—0261—03 文献标识码：A 中图分类号：TN912.34

基于特征分析的环境声音事件识别算法

刘波霞，陈建峰

(西北工业大学航海学院，西安 710072)

摘 要：对于环境声音事件，传统语音识别算法的识别效率低、稳定性差。为此，提出一种基于特征分析的环境声音事件识别算法。定义环境声音事件，分析常用的声音特征，不使用分类模型，仅利用特征对 4 种典型的环境声音事件进行分类。实验证明，该算法在识别率和稳定性上都优于传统识别算法，能够完成分类任务。

关键词：环境声音事件；特征分析；识别算法；Matlab 仿真

Environment Acoustic Event Recognition Algorithm Based on Feature Analysis

LIU Bo-xia, CHEN Jian-feng

(College of Marine, Northwestern Polytechnical University, Xi'an 710072, China)

【Abstract】The algorithm on Acoustic Event Classification(AEC) always uses traditional speech classification model at present, but to environment acoustic event, this algorithm has lower correct rate and bad stability. This paper puts forward an environment acoustic event classification algorithm based on feature analysis. It makes a definition to environment acoustic event. It analyzes familiar features of sound. It makes a classification to four typical kinds environment acoustic event just using features. Experiment proves that this method is better than traditional algorithm in correct rate and stability.

【Key words】environment acoustic event; feature analysis; recognition algorithm; Matlab simulation

DOI: 10.3969/j.issn.1000-3428.2011.22.087

1 概述

声音是人类感知环境的重要信息来源之一，也是反映人类行为的重要特征。这里所指的声音不仅指语音，还包括其他各种环境声音事件，如敲门声、爆炸声、脚步声等。因此，对声音事件进行探测和识别可以帮助人们从另外一个侧面探测和分析人类的行为。

环境声音事件识别是声音事件识别(Acoustic Event Classification, AEC)技术在特定环境中的一种应用。声音事件识别包括声音事件探测[1]与声音事件分类[1]两大任务。声音事件探测是指在连续的声音流中实时地辨别、定位事件，声音事件识别则是指识别已从声音流中分离出来的事件。它根据对声场中不同声音的理解，将其转换为有意义的信息符号。

近年来，人们对声音事件的识别技术逐步开展了深入细致的研究，国际上为此设立了每年一度的声音事件识别竞赛，目前已经成功开展了 CLEAR2006 和 CLEAR2008[1-2]。在实际生活中，对环境声音识别技术的应用尚不普及。不过，一些与声音识别技术相关的应用却已经发展了多年，例如说话人识别[3]、语种识别[4]、语音情感识别[5]、机械故障诊断[6]等。其中情感识别与故障诊断技术与环境声音识别技术比较接近。但是对于环境声音来说，其面临的困难更为突出：各种不同环境下出现的声音复杂多样，频带较宽，多种声音叠加、反射，声音长度差异大，动态范围广，发声声源种类多样等。因此，上述研究工作尚不能满足要求。针对环境声音事件，仍然需要对其特征提取与识别技术进行深入的研究。

目前各个领域所采用的识别系统还是传统的语音识别系统，即包括预处理、检测、特征提取、分类器和后处理 5 个步骤。应用这种系统对环境声音事件进行识别，不仅正确率低，而且实时性差。

针对上述不足，本文选取了以下 4 种典型环境声音事件，即爆炸声、破碎声、哭笑声、尖叫声，提出一种基于特征的环境声音事件识别算法，并对其性能进行识别概率的检验。

2 环境声音事件

为了更好地研究环境声音事件特征，根据相关的分类规则给环境声音事件一个明确的定义。

在特定的环境中，由某些物体不规律振动所引起的具有示警意义的声音，将其定义为环境声音事件。从人类感情色彩出发，这类声音在感觉特征上通常表现为混沌、粗糙、尖硬；在象征意义上，通常代表人类的财产、生命等受到某种侵犯。根据不同声音代表不同意义，又可将环境声音事件分为具体示警声音，比如破碎、爆炸等。

3 环境声音事件识别算法

人们普遍承认，各类模式识别技术的关键问题在于特征提取。环境声音事件识别算法就是在特征分析的基础上提出来的各种识别算法。

3.1 声音特征

特征提取一般可以在 3 个层面上进行，包括统计特征(statistic)、句法特征(syntactical)和语义特征(semantic)。句法特征刻画对象的结构，而语义特征需要了解对象的先验知识。

基金项目：国家创新基金资助项目(07C26226101997)；教育部博士点基金资助项目(20096102120013)

作者简介：刘波霞(1985－)，女，硕士研究生，主研方向：声音信号处理；陈建峰，教授

收稿日期：2011-06-10 **E-mail：**bocai211@163.com

文献[7]解释了语义特征即上下文知识对识别的重要性，揭示了这相当于人类对声音场景的理解过程：自上而下(语义特征)和自下而上(统计特征)。对语音识别而言，统计特征可用于辨认语音的出现，句法特征可将语音切分成单字，而语法特征则用词典中的词汇来解释每一个单字。这 3 个层次虽然主要是由语音识别技术中总结出来的，但对于声音事件的识别仍然适用。其区别在于，每个“单字”可能是一种声音事件。

语音识别技术目前已经广泛采用了这 3 层的特征，因此，在连续语音识别方面取得了长足的发展。对于环境声音，人们的理解程度远达不到语音情况，因此在句法特征、语义特征方面的工作相对滞后，目前主要依赖统计特征，如 MFCC、LPC、PLP、WVD、CWT、FWT 等。在此，总结出目前常用的各种声音信号的统计特征[8]，主要包括短时能量、短时过零率、短时自相关函数、线性预测编码(LPC)、线形预测倒谱系数(LPCC)、对数频率能量系数(LFPC)、子带能量、知觉线性预测(PLP)、Mel 倒谱系数(MFCC)、小波变换(WT)、频谱流量、语音持续时间、共振频率、基音频率。虽然相当一部分特征是在研究语音识别、情感识别、话者识别中建立的，但对环境声音的识别也具有不同程度的应用价值。

3.2 特征分析

3.2.1 数据库和仿真环境

选取了以下 4 种声音事件，即爆炸声、破碎声、哭笑声、尖叫声作为研究对象。声音的素材通过网络下载、影视截取等方式收集，每种声音收集 100 个样本，并统一声音的格式，一律转换为 WAV 格式、44 100 Hz 采样频率、16 bit，单通道。排除背景噪声大、波形畸变的声音。要求每个声音文件中不能有异类声音，即独立声音事件。

本文针对这 4 种典型的环境声音事件，对每种声音上述统计特征进行识别性能实验，具体实验条件：(1)软件平台：Matlab 6.0；(2)采样率：16 000 Hz；(3)声音长度：1.5 s，并对每段声音进行移动分帧处理，帧长 512，帧移 256。

3.2.2 特征分析结果

对 3.1 节中提到的所有特征进行了实现，并利用自行构建的数据库进行对比分析。本文的目的是希望从中寻求到特征突出，简单易行的特征识别方法，尽可能简化模型训练环节，提高识别算法在不同环境、不同信噪比条件下的适应性能。

下面针对 4 种声音，给出典型特征的分析结果，这些特征由于十分突出，特别适合于区别信号类型：

(1)子带能量[9]

子带能量表征各频段上的短时能量。其计算过程为，将通频带分为若干个频段，求出每一子带的能量，即：

$$E_{i,j}=\sum_{k=i}^{j}|X_k|^2\Delta f \tag{1}$$

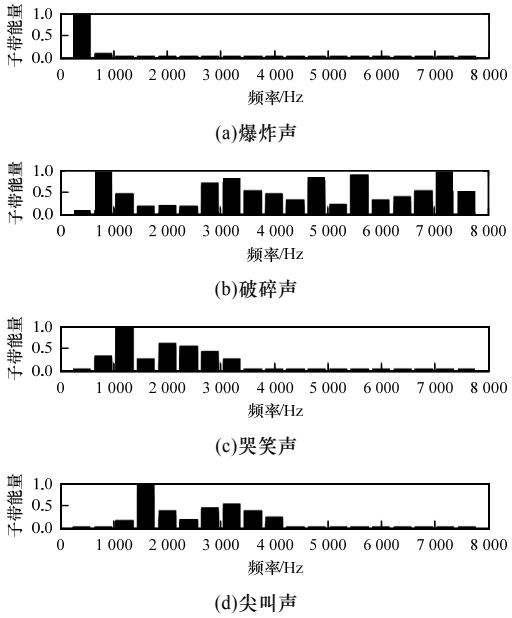
其中，i 为子带的下限频率；j 为子带的上限频率；X_k 为第 k 条谱线对应的振幅值；Δf 为子带频率分辨率。

对数据库中的 100 组样本进行分析，选取一组声音为例，其他样本与此十分接近，如图 1 所示。可以看到：

1)破碎声的子带能量存在的跨度很大，4 kHz 以上能量仍然明显存在；

2)爆炸声的子带能量只要集中在 2 kHz 以下，起伏很大，只有一个明显地峰值；

3)尖叫、哭笑声的子带能量在 2 kHz~4.5 kHz 区域分布，其他频段极少。



(a)爆炸声　(b)破碎声　(c)哭笑声　(d)尖叫声

图 1　4 种声音子带能量对比

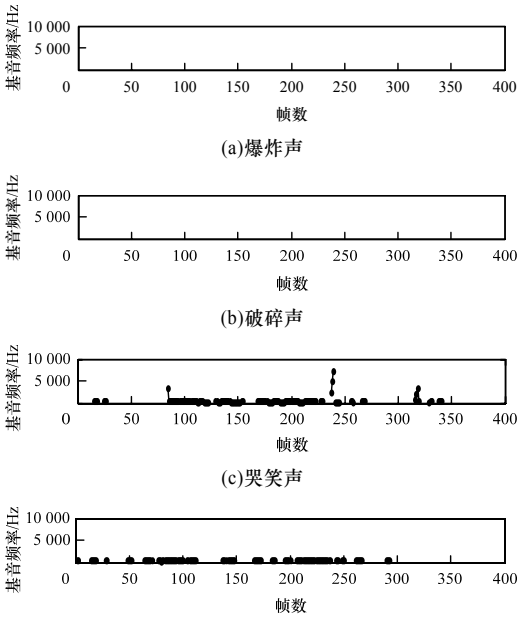
(2)基音频率

基频是指声带的振动频率，其倒数为基音周期，是指发浊音时声带的振动周期。

对声音数据库中的声音进行了分析，选取一组声音为例，如图 2 所示。可以看到：

1)爆炸声和破碎声不存在基音：所有的 100 组爆炸和玻璃破碎样本的声音都不包含基音成分。这是因为这 2 组信号从信号本质上讲就是一种噪声信号，没有基音产生的条件。

2)尖叫声、哭笑声的基音总是存在：从图 2 中可以看出，哭笑声和尖叫声的基音成分明显，并且时断时续，这是由于这些声音都是人发出的，在发出“有声”信号时，绝大都有较强的基音成分存在。



(a)爆炸声　(b)破碎声　(c)哭笑声　(d)尖叫声

图 2　4 种声音基音频率对比

(3)过零率

短时过零率表示一帧中语音信号穿过零点平的次数，其计算如下：

$$ZCR=\sum_{n=-\infty}^{\infty}\left|sign[x(n)]-sign[x(n-1)]\right|\times w(n-m)=\left|sign[x(n)]-sign[x(n-1)]\right|*w(n) \quad (2)$$

其中，$sign[x(n)]=\begin{cases}1 & x(n)\geqslant 0\\ -1 & x(n)<0\end{cases}$ 是符号函数；$x(n)$是每个语音帧的时域信号；N 为帧长；$w(n)$为窗口序列，一般采用矩形窗，为了平均，窗的幅度取为 $1/N$，为了使过零率作为“频率”的概念理解，窗的幅度再除以 2。

对声音数据库中的声音进行分析，选取一组声音为例，如图 3 所示。

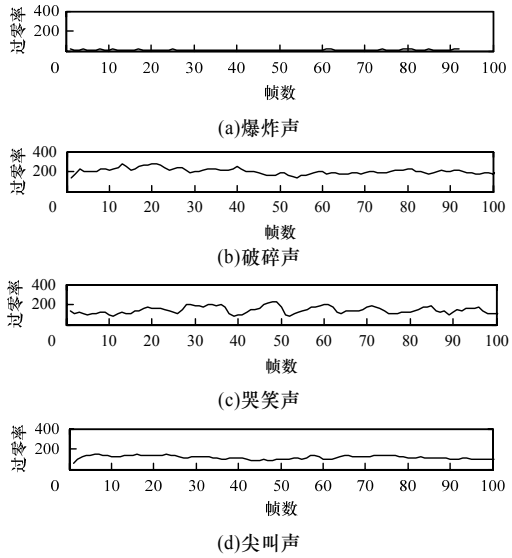


图 3　4 种声音过零率对比

从图 3 可以看到，爆炸声的过零率非常小，是 4 种声音里面最小的，主要是由于爆炸声的频率主要集中在较低频率。

(4)短时能量

语音信号常常可以假定为短时平稳的，信号 $x(n)$的短时能量定义为：

$$E(n)=\sum_{n=-\infty}^{\infty}[x(m)\cdot w(n-m)]^2=\sum_{m=n-N+1}^{n}[x(m)\cdot w(n-m)]^2 \quad (3)$$

其中，$w(n)$为一个长度有限的窗函数；N 为窗长。

对声音数据库中的声音进行分析，选取一组声音为例，如图 4 所示。

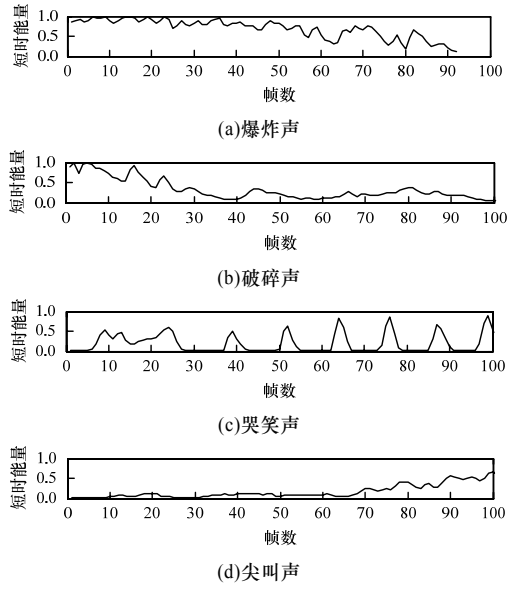


图 4　4 种声音短时能量对比

从图 4 可以看到，哭声和笑声的短时能量存在多个明显地峰值，即峰值和相邻谷值之间差异较大。

3.3　算法流程

基于上述的统计分析结果，选用 3 个特征，即基音、子带能量、短时能量将 4 种声音事件可靠地区分开来，过零率由于通过子带能量就可以与破碎声区别开，就没有再选用。具体流程如图 5 所示。

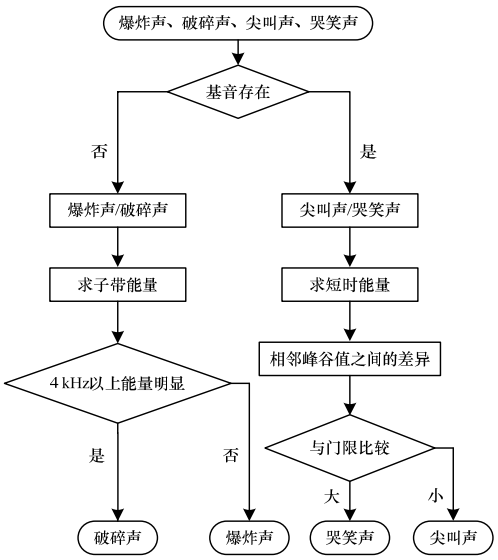


图 5　本文算法流程

3.4　实验结果

本文采用 2 种方法对 4 种声音事件进行分类测试：

(1)使用基音、子带能量、短时能量 3 种特征进行分类，不使用任何模型。

(2)采用 MFCC 特征，MFCC 是一种常用的语音识别算法，提取 13 维 MFCC。每类声音单独建立 GMM 模型，128 个高斯分量，对角型协方差阵，测试结果如表 1 所示。

表 1　2 种方法测试结果对比　(%)

名称	方法(1)正确率	方法(2)正确率
哭笑声	62.4	64.3
爆炸声	98.3	97.1
破碎声	97.6	100.0
尖叫声	78.6	56.4

为了进一步考察该方法在实际环境下的适应性，又做了一项实验。将数据库中的样本在实验室环境中用一般的扬声器播放，由分析软件采集后再处理，这时，相当于原来的数据又经过了扬声器和麦克风 2 个“传递函数”的处理，以及周围环境背景噪声的干扰，在相同的条件下，使用以上 2 种方法进行分类，结果如表 2 所示。

表 2　处理后声音 2 种方法测试结果对比　(%)

名称	方法(1)正确率	方法(2)正确率
哭笑声	61.9	52.6
爆炸声	97.2	89.3
破碎声	96.8	85.2
尖叫声	78.1	41.7

通过以上实验可以看到，从特征分析的角度出发，可以实现 4 种声音事件有效的分类，其分类结果与传统分类结果相当。同时不使用分类模型，计算方法均有快速算法，可以有效地提高识别系统的实时性。通过实际环境播放后，由于声音特征细节发生了一定的变化，方法(2)在某些类别上的性能下降了一些，而方法(1)则由于利用了典型的特征，基本不受影响。

(下转第 267 页)

是声誉较好的风格，TFT 型在收益和声誉方面都比较优秀。

表 2　基于 IREEG 模型的企业决策风格存活率比较

学习对象	企业决策风格	存活率/(%)			
		迭代 20 次	迭代 40 次	迭代 60 次	迭代 80 次
收益	TFT	28.40	37.35	45.50	51.10
	TOL	18.40	17.90	15.80	15.05
	REV	17.65	15.40	13.25	11.20
	EVI	17.80	15.00	13.50	12.55
	RAN	17.75	14.35	11.95	10.10
声誉	TFT	20.75	19.55	21.15	22.60
	TOL	23.50	30.10	30.95	37.15
	REV	17.35	14.35	12.65	9.80
	EVI	19.20	17.15	17.95	15.40
	RAN	19.20	18.85	17.30	15.05
吸引值 (IREEG)	TFT	26.15	30.40	32.90	36.65
	TOL	22.50	27.35	29.30	32.40
	REV	15.25	10.70	10.05	8.75
	EVI	17.55	15.05	13.05	10.10
	RAN	18.55	16.50	14.70	12.10

综上所述，IREEG 模型可以很好地区分各决策风格的声誉，并使具有良好声誉的决策风格获得较好的收益，在进化中获得较强的生存优势，有较高的存活率。

6　结束语

本文总结 TFT 型、TOL 型、REV 型、EVI 型、RAN 型 5 种决策风格，并结合现实经济生活，提出一种引入声誉影响的 IREEG 模型，可以很好地区分各决策风格，使具有良好声誉的决策风格获得较好的收益。今后将针对决策风格类型及竞争合作博弈研究收益矩阵值的选择问题。

参考文献

[1] 杨　城, 孙世新. 非完备策略的演化少数者博弈模型研究[J]. 计算机工程, 2007, 33(11): 26-28.

[2] 陈金波. 企业竞争的进化博弈论与种群生态学模型[J]. 数学的实践与认识, 2009, 39(1): 111-119.

[3] 晏国祥. 企业声誉测评指标体系[M]. 北京: 经济科学出版社, 2009.

[4] 程德华. 有限理性下企业合作竞争的进化博弈分析[J]. 计算机与数字工程, 2005, 33(5): 44-47.

[5] Ohdaria T, Terano T. Cooperation in the Prisoner's Dilemma Game Based on the Second-best Decision[J]. Journal of Artificial Societies and Social Simulation, 2009, 12(4): 71-90.

[6] Lam Ka-Man, Leung Ho-Fung. Incorporating Risk Attitude and Reputation into Infinitely Repeated Games and an Analysis on the Iterated Prisoner's Dilemma[C]//Proc. of the 19th IEEE International Conference on Tool with Artificial Intelligence. Patras, Greece: [s. n.], 2007.

编辑　陆燕菲

(上接第 263 页)

4　结束语

本文通过对大量实测样本的统计分析，得出一种简易可行的 4 种典型声音的识别算法，无须再进行训练，环境适应性和实时性均优于传统语音识别采用的 MFCC 特征向量方法。

但是同时也发现，有些声音的识别效率较低，比如哭笑声和尖叫声。这个问题主要是因为数据库分类不仔细，比如哭笑声里面，就包含男子哭笑声、儿童哭笑声等。这种不细致的分类对特征分析带来了一定的困难。

环境声音识别是一个涉及面十分广泛的领域，今后将在以下方面重点开展研究：(1)数据库建立。代表性强、一致性好的数据库是研究声音事件识别技术的基础。(2)针对特定环境声音事件的特征提取技术研究。为了更好地反映环境声音特征，需要在信号定义、描述、特征参量的选取、构建等方面加大基础理论研究力度。(3)加强时频域和空域先验信息和声音逻辑关系的理解和应用。在利用“语义”和“句法”来理解环境声音类别这一环节上目前尚属空白，有待进一步加强理论研究。

参考文献

[1] Temko A, Malkin R, Zieger C, et al. CLEAR Evaluation of Acoustic Event Detection and Classification Systems[C]//Proc. of the 1st International Evaluation Conference on Classification of Events, Activities and Relationships. Heidelberg, Germany: Springer-Verlag, 2007: 311-322.

[2] Heittola T, Klapuri A. TUT Acoustic Event Detection System[C]//Proc. of the 2nd International Evaluation Conference on Classification of Events, Activities and Relationships. Heidelberg, Germany: Springer-Verlag, 2008: 364-370.

[3] 王书诏, 邱天爽. 说话人识别研究综述[J]. 电声技术, 2007, 31(1): 51-65.

[4] 姜洪臣, 郑　榕, 张树武, 等. 基于SDC特征和GMM-UBM 模型的自动语种识别[J]. 中文信息学报, 2007, 21(1): 49-53.

[5] 朱永崇. 语音情感识别的特征分析与多子模式投票方法的研究[D]. 哈尔滨: 哈尔滨工业大学, 2005.

[6] 常西畅. 机械设备噪声故障诊断的新进展[C]//2002 年全国振动工程及应用学术会议论文集. 上海: 上海高教电子音像出版社, 2002: 347-349.

[7] 曹　华, 李　伟, 谭艳梅. 线性预测及其 Matlab 实现[J]. 现代电子技术, 2009, 294(7): 133-135.

[8] 赵　力. 语音信号处理[M]. 北京: 机械工业出版社, 2008.

[9] 高明明, 常太华, 杨国田, 等. 基于子带主频率信息的语音特征提取算法[J]. 计算机工程, 2009, 35(18): 161-163.

编辑　任吉慧