一种用于枪声的多级检测识别技术

张 涛，苏春玲
（天津大学 电子信息工程学院，天津 300072）

摘要：声音无处不在，声音的检测与识别一直是声音研究领域的重要内容。其中公共场所环境中的突发声音的检测一直是一个难题，本文提出了一种利用短时能量和短时平均过量率以及 MFCC 和 DTW 对枪声等特定声音的多级检测算法。实验结果表明，本算法具有较低的计算复杂度，易于实现，而且检测的漏检率和误检率都很理想。

关键词：枪声；短时能量；短时过零率；MFCC；DTW
中图分类号：TP391 **文献标识码**：A **文章编号**：1674-6236(2013)18-0056-03

A multi-level detection and identification technique for gunshots

ZHANG Tao, SU Chun-ling
(*School of Electronic Information Engineering, Tianjin University, Tianjin* 300072, *China*)

Abstract: As we known, the sound is everywhere. The detection and recognition of the sound has been an important content of the sound's research. It has been a problem that the detection of a burst of sound in the environment of the public places. This paper proposes a multi-stage way to detect gun shots. This way takes advantage of short-term energy and short-time average zero crossing rates and MFCC and DTW. The experiments show that this method has low computation complexity, and can be implemented easily. From the results, we can also find that ideal undetected rates and false detection rates can be obtained.

Key words: gun shots; short-term energy; short-time average zero crossing rates; MFCC; DTW

声音无处不在，人们的生活中伴随着各种各样的声音，有熟悉的也有不熟悉的，甚至有从未听过的。声音的检测与识别一直是声音研究领域的重要内容。声音识别系统就是将待测声音与已经训练好的已知的声音模板进行匹配，从而得知待测声音是什么声音的过程。在声音识别领域中，可以将其分为语音识别和非语音识别。对于语音识别的研究比较深入[1]，非语音识别的研究相对滞后一些。对于非语音识别的用处主要有以下几个方面，如军事与国防、工业生产与控制、医疗护理和安全控制等。在非语音声音信号的检测研究中，公共场所环境中的突发异常声音检测一直是一个被公认的难题。一些学者对特定场景下的某些声音做了相关研究，例如：Clavel 等人利用 LPCC 表征声音信号，通过 GMM 检测识别噪声环境下发生的枪声，在他们的实验中主要研究不同信噪比的训练序列对检测结果中漏检率的影响，得到的结果是当测试序列与训练序列的信噪比相同时，检测结果的漏检率最低；当用没有噪声的序列作为训练序列时，检测结果的漏检率最高[2]；Alain 等人通过将频率域分成 N 等分，计算每份的能量值表征声音信号，利用 GMM、HMM 来检测识别噪声环境下的突发声音，实验结果得到了很好的识别率，在信噪比 70 dB 的时候识别率能达到 98%，在 0 dB 的时候也能达到 80%以上，但是他们所用的测试环境仅限于噪声是高斯白噪声[3]，而且这个方法的算法复杂度很高，不易于在硬件中实现。综合考虑算法复杂度和识别率等多方面内容，本文提出了一种用于公共场所中枪声的多级检测识别方法。

1 算法介绍

非语音识别系统与语音识别系统类似，都基本由特征参数提取算法和模式匹配算法构成。

1）特征参数提取

用于声音分类的特征参数[4]很多，可以归纳为 3 大类：时域特征参数、频域特征参数，同态（倒谱）特征参数。时域特征参数的特点是提取算法都不复杂，缺点是对信号的可鉴别能力有限。频域特征参数与人类听觉系统有一定的关系，但是频域特征参数仅适用于加性信号，对于复杂的乘积性组合信号处理能力不好。同态特征参数的优点正是在于这种非线性系统的处理，也就是对其进行同态分析，设法将非线性问题转化为线性问题来处理。

2）模式匹配及模型训练技术

模型训练是指按照一定的准则，从大量已知模式中获取表征该模式本质特征的模型参数，而模式匹配则是根据一定准则，使未知模式与模型库中的某一个模型获得最佳匹配。语音识别所应用的模式匹配和模型训练技术主要有动态时间归正技术（Dynamic Time Warping，DTW）、隐马尔可夫模型（hidden Markov model，HMM）和人工神经元网络（Artificial Neural Networks，ANN）。在这 3 种技术中，DTW 是较早的一

收稿日期：2013-03-12 稿件编号：201303147
作者简介：张 涛（1975—），男，天津人，博士，副教授。研究方向：音视频编解码，DSP 应用，多媒体处理器结构设计等。

种模式匹配和模型训练技术，它应用动态规划方法成功解决了声音信号特征参数序列比较时时长不等的难题，它的算法复杂度低而且识别率针对某些特定方面也有很好的表现，尤其在孤立词语音识别中获得了良好性能。

对于突发事件的声音检测，如枪声，输入信号类似于语音中的孤立词，而且系统所需要的匹配模板较少。用于此类识别，DTW 算法与 HMM 算法在相同的环境条件下，识别效果相差不大，但 HMM 算法要复杂得多，主要体现在 HMM 算法在训练阶段需要提供大量的语音数据，通过反复计算才能得到的模型参数，而 DTW 算法的训练中几乎不需要额外的计算。所以 DTW 算法对这种输入信号比较短促，类似于单音信号而且模板又比较少的声音进行识别时，在算法复杂度和识别率方面都很适合，能获得良好的效果。故本实验采用 DTW 算法。

3)在本算法中，用到的特征参数是短时能量、短时平均过零率和 Mel 频率倒谱系数。

①短时平均过零率

对于短时平均过零率[5]这个参数，由于在无声段噪声使声音波形在 0 值附近来回摆动，如果只考虑符号变化这一单一条件会导致计算出的过零率和有声段的区别并不十分明显；在本算法中还加入另一条件就是设定一个差的阈值 δ，使不仅 $s_n(m)*s_n(m+1)<0$，还要 $|s_n(m)-s_n(m+1)|>\delta$。在本系统中经多次试验取定 $\delta=250$。

②Mel 频率倒普系数(MFCC)算法

短时能量和短时平均过零率都只是时域上的一些特征参数，这种参数没有充分利用人耳的听觉特性。人的听觉系统是一个特殊的非线性系统，它响应不同频率表信号的灵敏度是不同的，基本上是一个对数的关系。Mel 频率倒谱系数(Mel frequency cepstrum coefficient，MFCC) 充分利用人耳这种特殊的感知特性参数。

MFCC 参数是按帧算的。首先通过 FFT 得到该帧信号的功率谱 $S(n)$，然后通过 Mel 频率滤波器 Hm(n)。滤波器在频域上为简单的三角波，其中心频率为 fm，它们在 Mel 频率轴上是均匀分布的。在线性频率上，当 m 较小时，相邻的 fm 间隔很小，随着 m 的增加，相邻的 fm 间隔逐渐被拉开。如图 1 所示。

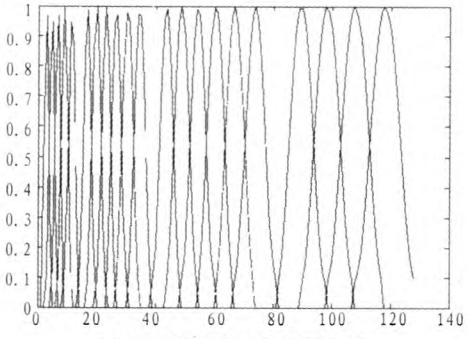

图 1　Mel 频率滤波器示意图

Fig. 1　Schematic diagram of the Mel frequency filter

Mel 频率和线性频率的关系：　$fmel=2595\log(1+f/700)$

声音信号的 MFCC 提取及计算过程如图 2 所示[6]。

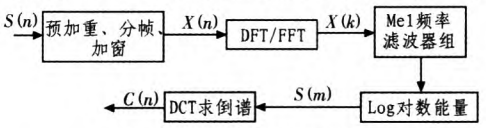

图 2　MFCC 提取过程框图

Fig. 2　Structure diagram of the MFCC's extraction process

4)本算法的实现步骤

通过前面的分析，本系统所采用的特征参数是短时能量、短时平均过零率和 MFCC 的结合，所采用的时域参数实现容易且计算量小但是没有考虑人耳的听觉特性，所以识别率不高；而 MFCC 正好弥补了这一点，但是计算量很大；所以本算法采用一种多级的模式，将这个几个参数结合在一起，彼此起到互补的作用，一级一级逐步缩小目标，最后检测出枪声的位置。整个算法可以分成 3 级，算法的分级结构框图如图 3 所示。

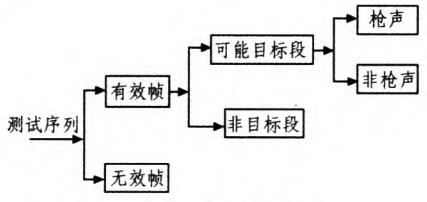

图 3　算法分级结构框图

Fig. 3　Structure diagram of the algorithm classification

在第一级中对每帧(1 024 点)计算其短时能量和短时平均过零率；然后根据设定好的门限对每帧的能量和过零率进行判断，这两个参数只要有一个参数满足条件，就将这帧认为是有效帧。其中短时能量门限 EN_MIN=53，短时过量率的上下门限分别为 ZCR1=65、ZCR2=100。

在这一级的判断中，还采用了一个平滑机制。采用这个平滑机制是为了减小漏检率，由于实验所选用枪声模板为 11 帧，所以为了提高检测速度和检测准确率，最小的可能目标段要求至少 6 帧。为了避免将一个可能目标段由于中间一两帧的无效帧被分成两个非目标段，导致检测遗漏，本实验中采取了这个 3 帧的平滑机制，也就是两个可能目标段之间的无效帧小于等于 3 帧时，这两段会被认为是一段。在程序中的具体思想是，如果当前帧根据短时能量和过零率判断时，被判断为无效帧；那么如果它的前三帧中有一帧是根据这两个条判断为的有效帧，则当前帧也被平滑为有效帧。

在第二级中，对连续有效帧组成的各个段进行判断。由于本算法中选择的枪声模板长度是 11 帧，所以当连续有效帧小于 6 帧时，认为是非目标段，直接舍弃，不进行下一步分析；如果连续有效帧达到 17 帧时还没有遇到无效帧，那么将前 11 帧作为一个可能目标段进行下一步的分析，将后 6 帧作为下一段的前 6 帧继续计数；如果连续有效帧帧数在 6 和 17 之间，则直接将这一段作为一个可能目标进行下一步分析。

在第三级中，对可能目标段进行 MFCC 参数提取，然后将其与模板的 MFCC 参数进行 DTW 匹配，如果计算出的最后累计距离值小于设置的阈值，就确定这段为枪声，输出该

段的起止帧号。

2 实验结果

本实验所采用的输入文件都是.WAV 文件,声道是单声道,采样率为 48 kHz,采样值是 16 bit 的 PCM 量化编码。在实验时,将帧长定为 FRAME_LEN=1 024。实验步骤及结果如下:

首先要确定模板,然后利用各种枪声文件训练出 DTW 匹配时所需要的阈值 GUN_MAX。选定的模板 gun_template 是只有一声枪声文件,模板包含 11 帧。通过对各种环境下包含枪声的序列进行训练,可以将所需要阈值 GUN_MAX 确定为 4 525。

在模板和阈值都训练成功后,利用训练出的结果开始对待测文件进行检测,待检测文件 gun_test 和训练时用的文件的采样率等参数是完全一样。待测文件是一个复杂的环境,有背景音乐,有人说话等各种声音,部分检测结果示意图如图 4 所示。实线框代表手工标注结果;虚线框代表算法检测结果。该测试文件的总帧数为 1953 帧。其中从 100 帧到 400 帧的手工标注结果和算法检测结果如表 1 所示。通过漏检率和错检率来表示检测准确率,其中漏检率 $\alpha=\frac{\text{漏检帧数}}{\text{总帧数}}\times 100\%$,错检率 $\beta=\frac{\text{错检帧数}}{\text{总帧数}}\times 100\%$。

表1 部分待测序列的手工标注和算法检测结果统计

Tab. 1 Statistics of the test sequences' manually labeled and algorithm detected results

手工标注枪声位置	检测出枪声位置	漏检帧数	错检帧数
	99 to 108	0	10
136–150	136 to 146	0	0
	147 to 155	0	5
174–203	177 to 187	3	0
	188 to 193	10	0
	232 to 242	2	0
	243 to 253	0	0
230–273	254 to 268	5	0
	289 to 294	0	6
	300 to 309	0	10
330–364	347 to 353	17	0
	355 to 360	5	0
398–403	403 to 408	5	5

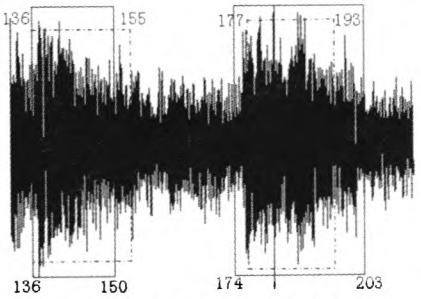

图 4 部分测试结果示意图

Fig. 4 Schematic diagram of the part of the test results

对检测结果进行统计,可计算出总的漏检帧数为 87,则漏检率 $\alpha=\frac{87}{1\,953}\times 100\%=4.45\%$;总的错检帧数为 237,则错检率 $\beta=\frac{237}{1\,953}\times 100\%=12.14\%$。

漏检率和错检率是一对此消彼长的参数,两者不能同时达到最优,只有根据具体情况,选择一个最适合当前情况的折中参数。对于本实验中的参数,如果将短时能量和短时过量率的门限值都减小(EN_MIN=55,ZCR1=68,ZCR2=95),同时将 GUN_MAX 设定为 4520,则所检测出的结果漏检率 α 会变大,错检率 β 会变小。部分检测结果示意图如图 5 所示。

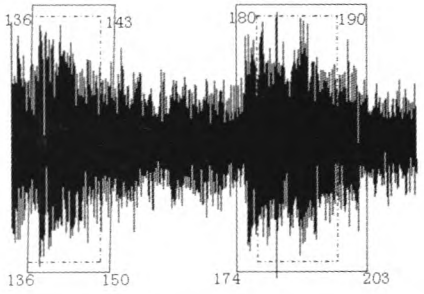

图 5 漏检率变大,误检率变小检测结果示意图

Fig. 5 Schematic diagram of the higher undetected rate and the lower false detected rate

对检测结果进行统计,可计算出总的漏检帧数为 203,则漏检率 α=10.39%;总的错检帧数为 152,则错检率 β=7.78%。

与上述情况相反,如果将检测条件放宽松,设定 EN_MIN=50、ZCR1=60、ZCR2=105,同时将 GUN_MAX 设定为 4530,那么所检测出的结果漏检率会变小,错检率会变大。部分检测结果示意图如图 6 所示。

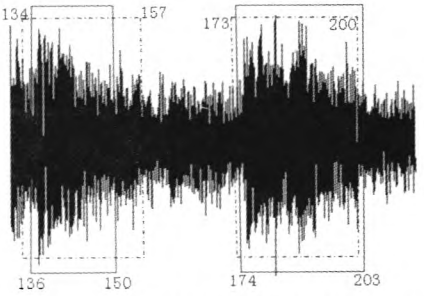

图 6 漏检率变小,错检率变大检测结果示意图

Fig. 6 Schematic diagram of the lower undetected rate and the higher false detected rate

对检测结果进行统计,可计算出总的漏检帧数为 82,则漏检率 α=4.20%;总的错检帧数为 268,则错检率 β=13.72%。

从上述的实验结果可以分析出,参数阈值的大小的稍稍变化会引起检测结果的大幅度变化,但是漏检率和错检率不可能同时达到最优,所以要根据实验的具体要求,选择适当的参数,使二者达到一个折中的平衡。

(下转第 61 页)

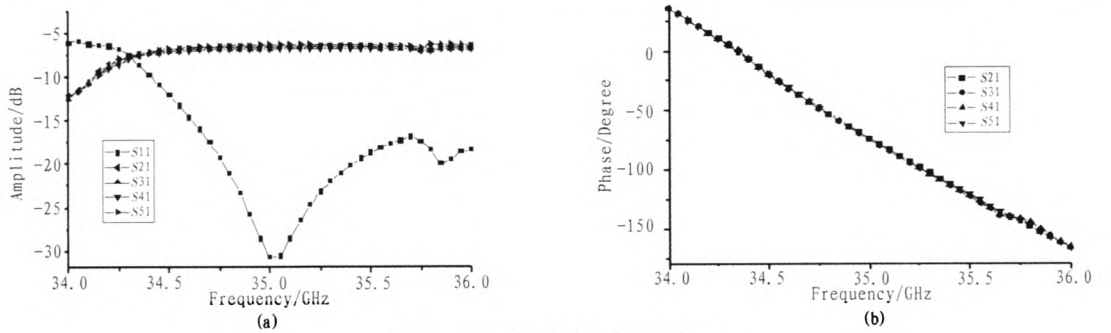

图 5 功分器幅度/相位仿真曲线

Fig. 5 The power combiner amplitude/phase simulation curves

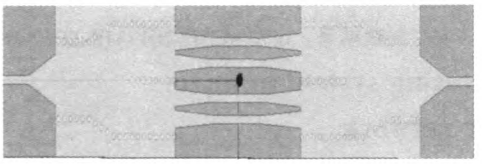

图 6 四路功分/合成器

Fig. 6 The four-way power divider/combiner

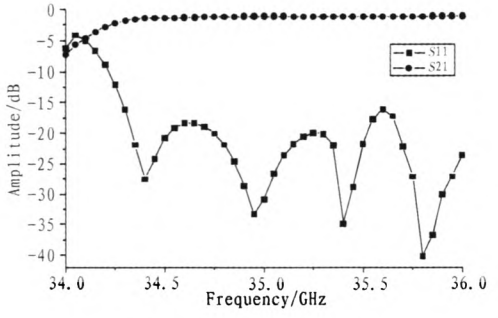

图 7 功分/合成器仿真结果

Fig. 7 The power divider/combiner simulation result

参考文献：

[1] YAN Li,HONG Wei. Investigations on the propagation characteistics of the Substrate Integrated Waveguide based on the Method of Lines[J]. IEE Proceedings-H:Microw Antennas and Propag,2005,152(1): 35–42.

[2] 刘冰，洪伟，郝张成，等. 基片集成波导梳状交替相位功分器[J]. 电子学报，2007，35(6):1061–1064.
LIU Bing,HONG Wei,HAO Zhang-cheng,et al. Alternate phase substrate integrated waveguide (SIW) power divider[J]. Chinese Journal of Electronics,2007,35(6):1061–1064.

[3] 周翼鸿. 基于波导的紧凑型功率合成技术研究[D]. 成都：电子科技大学，2010.

[4] 金海焱. 新型微波毫米波空间功率合成技术研究 [D]. 成都：电子科技大学，2010.

[5] 兰尧. 基片集成波导的应用和仿真研究[D]. 成都：电子科技大学，2006.

[6] 王哲. 基片集成波导的研究与应用[D]. 南京：南京邮电大学，2011.

(上接第 58 页)

3 结 论

文中提出了一种多级的特定突发声音枪声的检测算法，首先利用时域参数短时能量和短时平均过量率，其中还加入了平滑机制来检测出可能目标段；然后再结合倒谱参数 MFCC 和 DTW 对可能目标段进行进一步检测，确定其是否为要检测的目标。经过实验测试，取得了令人满意的结果。采用本算法，不但可以加快运算速度，易于在 DSP 和 ARM 上的移植和实现，而且所采用的测试文件背景环境很复杂，说明本算法在信噪比低的情况下，具有一定的鲁棒性，来确保检测的准确性和有效性。

参考文献：

[1] 詹新明，黄南山，杨灿. 语音识别技术研究进展[J]. 现代计算机:专业版，2008(9):12.
ZHAN Xin-ming,HUANG Nan-shan,YANG Can. Research progress of speech recognition technology[J]. Modern Computer: Professional Edition,2008(9):12.

[2] Clavel C,Ehrette T,Richard G. Events detection for an audio-based surveillance system[C]//IEEE International Conference on Multimedia and Expo,2005.

[3] Dufaus A,Besacier L,Ansorge M,et al. Automatic sound detection and recognition for noisy environment [C]//European Signal Processing Conference,Finland,2000(9):1033–1036.

[4] 刘华平，李昕，徐柏龄，等. 语音信号端点检测方法综述及展望[J]. 计算机应用研究，2008，25(8):2278–2283.
LIU Hua-ping,LI Xin,XU Bo-ling,et al. Review and prospect of the speech signal's endpoint detected method[J]. Application Research of Computers,2008,25(8):2278–2283.

[5] 徐大为，吴边. 一种噪声环境下的实时语音端点检测算法[J]. 计算机工程与应用，2003(1):115 –117.
XU Da-wei,WU Bian. A real-time speech endpoint detection algorithm in a noisy environment [J]. Computer Engineering and Applications,2003(1):115–117.

[6] 王炳锡，屈丹，彭煊. 实用语音识别基础[M]. 北京：国防工业出版社，2005.