文章编号:1671-4598(2016)07-0105-04 DOI:10.16526/j.cnki.11—4762/tp.2016.07.028 中图分类号:TP391.4 文献标识码:A

基于音频监控的婴儿智能监护系统设计

杜仲平[1],李一博[1],叶 霆[2]
(1.天津大学 精密仪器与光电子工程学院,天津 300072;2.天津九安医疗电子股份有限公司,天津 300190)

摘要:为解决目前婴儿智能监护产品对婴儿哭声识别不准确的问题,设计了能准确识别婴儿哭声的婴儿智能监护系统;使用RT5350芯片作为处理器,使用PAP7501芯片实现对视频信号和音频信号的采集;对系统移植了Linux内核,使用多线程编程技术编写了应用程序;通过提取梅尔频率倒谱系数作为特征参数,使用动态时间规整算法作为识别算法,实现了对婴儿哭声的准确识别;该系统可以与云服务器通信,实现婴儿哭声报警功能;该系统可以与用户的终端设备通信,实现视频数据和音频数据的传输;该智能监护系统极大地方便了用户对婴儿的监护。

关键词:智能监护;婴儿哭声;梅尔频率倒谱系数;动态时间规整算法

Design of Infants' Intelligence Monitoring System Based on Audio Monitoring

Du Zhongping[1], Li Yibo[1], Ye Ting[2]
(1. School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China;
2. ANDON Health Co., Ltd., Tianjin 300190, China)

Abstract: To solve the problem of the current infants' intelligent monitoring products' low recognition rates for the infants' crying, an infants' intelligent monitoring system which can recognize the infants' crying accurately was designed. The RT5350 chip was used as processor, the PAP7501 chip was used to capture the video signal and audio signal. A Linux kernel was ported to the system, the multi-threaded programming techniques was used to program the application programs. The Mel frequency cepstrum coefficient (MFCC) was extracted as the characteristic parameters, the dynamic time warping (DTW) algorithm was used as the recognition algorithm. The system can communicate with the cloud server to realize the infants' crying alarm function, and it can also communicate with the user's terminal device to realize the transmission of video data and audio data. The users can monitor the infants effectively with the help of the system.

Keywords: intelligent monitor; infants' cry; MFCC; DTW algorithm

0 引言

婴儿的哭声中包含了婴儿大量的生理的和心理的信息,反映了婴儿的某种需要,是婴儿与外界交流的重要手段,是婴儿对外界发出的一种警报,对婴儿哭声进行监控有重要的意义。目前的婴儿智能监护产品有的不能对婴儿哭声进行识别,有的通过计算音频信号频谱的能量分布实现对婴儿哭声的识别,不能取得很好的识别效果[1-4]。为了解决婴儿智能监护产品对婴儿哭声识别不准确的问题,本文提取了在语音识别中被广泛使用的Mel频率倒谱系数作为特征参数,使用动态时间规整算法作为分类算法,实现了对婴儿哭声的准确识别。

为了实现对婴儿的有效监护,本文设计了以音频监测功能为主要功能的婴儿智能监护系统,该系统还具有视频监控和网络通信的功能。当系统监测到婴儿哭声后,会向远端的云服务器发送消息,云服务器收到消息后,会向指定的用户推送消息,实现对婴儿哭声的报警。用户可以使用手机或平板电脑等终端设备通过网络与本智能监护系统通信,实时的对婴儿的状态进行监控。

1 婴儿哭声的识别方法

由于婴儿的发音原理与成年人发音原理相似,本文使用在孤立词语音识别中取得很好效果的动态时间规整算法对婴儿哭声进行识别[5]。使用动态时间规整算法对婴儿哭声进行识别的过程分为两个阶段,第一阶段为训练阶段,通过对训练集的样本提取特征参数建立参考模板库,第二个阶段为识别阶段,提取输入信号的特征参数得到测试模板与参考模板库里的参考模板进行匹配,得到识别结果[6]。婴儿哭声识别原理图如图1。

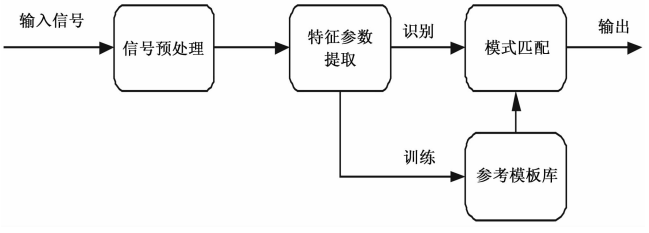

图1 婴儿哭声识别原理图

1.1 Mel频率倒谱系数的提取

Mel频率倒谱系数是目前使用最广泛的语音特征参数之一,它是着眼于人耳的听觉特性的特征参数[7]。人耳对不同频率的声音具有不同的感知能力,实验发现,在1 000 Hz以下,感知能力与频率成线性关系;在1 000 Hz以上,感知能力与频率成对数关系[8]。在频域里,有一种Mel频率尺度,Mel频率尺度与人耳的感知特性成线性的关系。频率 f 与Mel频率B之间的转换公式如公式(1)。

$$B = 2\ 595\ \lg\left(1+\frac{f}{700}\right) \tag{1}$$

Mel频率倒谱系数是基于上述Mel频率概念而提出的,提取计算过程如图2。

收稿日期:2015-12-30; **修回日期**:2016-02-24。
基金项目:天津科技支撑计划项目(14ZCZDGX00003);天津市海洋经济创新发展区域示范项目(2015120024000473);天津市自然科学基金一般项目(13JCYBJC18000)。
作者简介:杜仲平(1990-),男,吉林通化人,硕士研究生,主要从事嵌入式系统方向的研究。

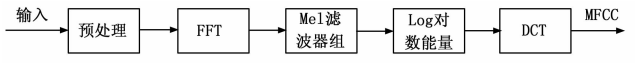

图 2 Mel 频率倒谱系数提取过程

Mel 频率倒谱系数的提取过程如下：

(1) 原始语音信号 $s(n)$ 经过归一化、预加重、分帧、加窗、端点检测处理，得到每个语音帧的时域信号 $x(n)$。归一化是为了减小音量大小不同对识别结果的影响。预加重的目的是提升高频部分，使信号频谱变得平坦，对信号的预加重通过使用传递函数为 $H(z)=1-az^{-1}$ 的预加重滤波器实现的，a 为预加重系数，本文中 a 的值为 0.975。由于语音信号具有短时稳定性，可以将 10 ms～30 ms 左右的语音信号看成是稳定的信号，本文的 AD 转换器的采样率为 8 kHz，将 32 ms 的信号分为一帧，一帧信号有 256 个采样点，相邻帧的帧移设置为 64 点。为了降低由于截断产生的频谱泄漏，对每一帧信号加汉明窗。使用双门限法对信号进行端点检测，找出语音段的起始点与终止点。

(2) 将时域信号 $x(n)$ 经过快速傅里叶变换 (FFT) 后得到线性频谱 $x(k)$。

(3) 将上述线性频谱通过 Mel 频率滤波器组得到 Mel 频率。

Mel 滤波器组为在语音的频谱范围内设置的若干个的带通三角形滤波器，每个滤波器的传递函数 $H_m(k)$ 计算公式如公式 (2)。

$$H_m(k)=\begin{cases}0(k<f(m-1))\\ \dfrac{k-f(m-1)}{f(m)-f(m-1)}(f(m-1)<k\leqslant f(m))\\ \dfrac{f(m+1)-k}{f(m+1)-f(m)}(f(m)<k\leqslant f(m+1))\\ 0(k>f(m+1))\end{cases}$$
$$(0\leqslant m<M) \tag{2}$$

M 为滤波器组中滤波器的个数，通常取值在 24～40 之间，本文取 24。

(4) 计算 Mel 频率的对数能量 $S(m)$。

(5) 将上述频谱的对数能量 $S(m)$ 做离散余弦变换 (DCT) 就得到了 Mel 频率倒谱系数 $C(n)$。

$$C(n)=\sum_{m=0}^{M-1}S(m)\cos\left(\frac{\pi n\left(m+\frac{1}{2}\right)}{M}\right)(n=1,2,\cdots,l) \tag{3}$$

式中，l 表示 Mel 频率倒谱系数的阶数。实际应用中，通常取 12～16 阶的 Mel 频率倒谱系数，本文使用 12 阶的 Mel 频率倒谱系数进行实验。分别对判定为语音段的每一帧信号提取特征参数，同一语音段的不同帧信号的特征参数构成一个特征参数矢量序列。

1.2 动态时间规整算法及改进

1.2.1 动态时间规整算法

动态时间规整算法是把时间规整和距离测度计算结合起来的一种非线性规正技术[6]。DTW 原理图如图 3。由训练集样本提取出来的特征参数矢量序列叫参考模板，由测试信号提取出的特征参数矢量序列叫做测试模板，参考模板用 $R=\{R(1),R(2),\cdots\cdots,R(m),\cdots\cdots R(M)\}$ 表示，测试模板用 $T=\{T(1),T(2),\cdots\cdots,T(n),\cdots\cdots,T(N)\}$ 表示。其中，m、n 分别是参考模板和测试模板的帧号，M、N 分别为参考模板和测试模板的帧数。$R(m)$ 为参考模板的第 m 帧特征矢量，$T(n)$ 为测试模板的第 n 帧特征矢量。分别取参考模板和测试模板的第 m 帧和第 n 帧，用 $d(R(m),T(n))$ 表示这两帧之间的矢量失真度，本文使用欧式距离测度计算失真度，用 $D(R(m),T(n))$ 表示到当前点的最优路径的累计失真度。本文中对搜索路径作如下限制：可以到达坐标为 (m,n) 的点的前一节点只能是 $(m-1,n)$，$(m-1,n-1)$，$(m-1,n-2)$ 中的一点，此时，最优路径的累积失真度的递推公式如下：

$$D(R(m),T(n))=d(R(m),T(n))+\min\{D(R(m-1),T(n)),D(R(m-1),T(n-1)),D(R(m-1),T(n-2))\} \tag{4}$$

从 (1，1) 点出发，反复递推到 (M,N) 点，就可以得到最优路径的累计失真度 $D(R(M),T(N))$。累计失真度表明测试模板与参考模板的相似程度，累计失真度越小，两段信号的相似程度越高。

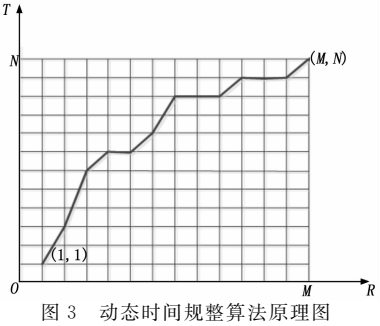

图 3 动态时间规整算法原理图

1.2.2 动态时间规整算法的改进

上述算法需要对参考模板和测试模板的每一个特征矢量都进行欧式距离计算，需要进行 $M*N$ 次欧式距离计算。为了减少不必要的计算，对算法的搜索范围进行限制，删去在任意一个轴的方向上过分倾斜的路径，把搜索路径限制在一个平行四边形里，相邻边的斜率设定为 0.5 和 2。动态时间规整算法是一种模板匹配算法，它的识别效果依赖于端点检测的效果，为了降低算法对端点检测的依赖，放松算法对搜索路径的起始点和结束点的限制，不严格的要求测试模板和参考模板的端点对齐，即搜索路径的起点在横纵两个方向上放宽 2 帧，起点可以在 (1，1)，(1，2)，(1，3)，(2，1)，(3，1) 中选择，搜索路径终点的约束也做类似的放松[9]。改进的动态时间规整算法原理图如图 4。

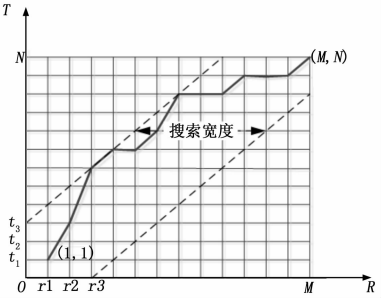

图 4 改进的动态时间规整算法原理图

2 系统硬件设计

本实验的硬件原理图如图 5 所示。本文设计的智能监护系统由视频采集模块，音频采集模块，电机控制模块和网络通信

模块等功能模块组成。本文使用 Ralink 公司的 RT5350 芯片作为处理器，它集成了 MIPS 24Kc 360 MHz 处理器；支持 16－bit SDRAM，最高可扩展 64 MB 的 SDRAM；具有 USB 2.0 Host/Device 接口；集成五端口百兆以太网交换机；具有 GPIO，SPI，I²C，I²S，PCM，UART 及 JTAG 接口；集成 2.4 GHz 射频单元，集成 802.11n 1×1 MAC/基带处理器，支持 150 Mbps 无线数据带宽，能满足数据传输的要求。

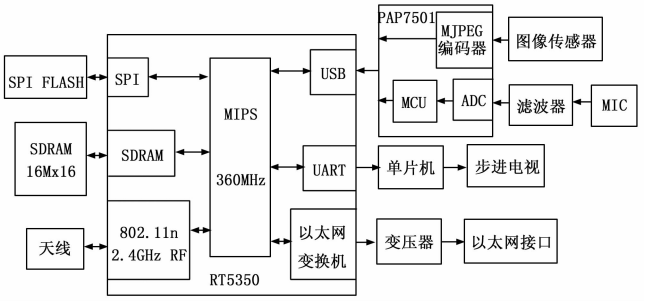

图 5　系统硬件原理图

通过使用 PAP7501 芯片实现对视频信号和音频信号的采集。PAP7501 是一款高速 USB2.0 摄像头控制器，支持 USB 视频类 1.1 和 USB 音频类 1.0 标准，内部集成了 MJPEG 编码器和 ADC 转换器，可以将采集到的视频信号编码成 MJPEG 格式，将音频信号编码成 PCM 格式，并通过 USB 接口将视频信号和音频信号传输到 RT5350 进行处理。本实验中，AD 转换器的采样率设置为 8 kHz，为了防止对音频信号采样时发生混叠，需要在采样前使用防混叠滤波器对信号进行滤波，本文使用的是 4 阶巴特沃斯低通滤波器，截止频率为 4 kHz。防混叠滤波器原理图如图 6。

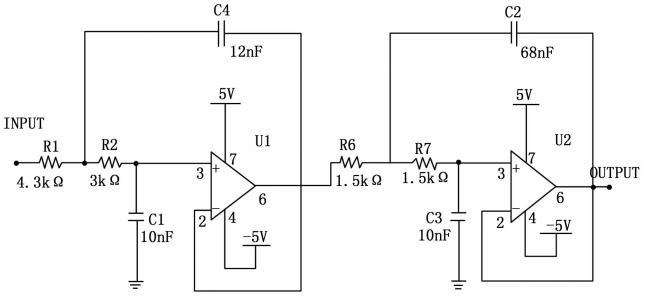

图 6　防混叠滤波器原理图

为了提高系统的性能，通过 SPI 接口扩展了 8 Mb 的 Flash 用于存放程序代码，还扩展了 16 M x 16 位的 SDRAM 用于存放数据。RT5350 内部集成 2.4 GHz 射频单元，增加天线可以起到增强发射信号的作用。RT5350 内部集成了百兆以太网交换机，需要为它添加一个以太网接口，为了实现电压的匹配，使用变压芯片 VP8901 实现电压转换。为了控制云台转动，使用 78F9222 单片机控制步进电机转动，RT5350 通过串口与单片机通信。

3　系统软件设计

本文使用 Ralink 公司提供的 RT288x _ SDK 编译 uboot、系统内核以及根文件系统，搭建好嵌入式 Linux 开发环境。

系统软件采用多线程编程技术实现，系统软件原理图如图 7。由主进程分别创建音频采集线程、视频采集线程、音频处理线程、用户消息处理线程、串口通信线程、与云服务器通信线程、音频发送线程和视频发送线程。

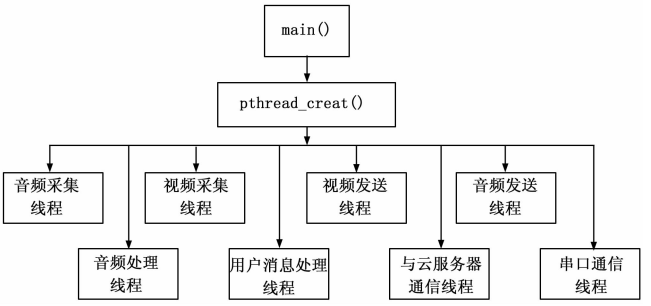

图 7　系统软件原理图

音频线程实现对音频数据的采集，使用 ALSA 框架实现。视频采集线程实现对视频数据的采集，使用 V4L2 框架实现。用户消息处理线程用来接收用户的命令，并把命令通过消息队列的方式传递给其它线程，其它线程根据收到的消息，完成相对应的动作，用户的命令主要有开启/关闭音频报警功能，接收视频、音频数据，控制云台转动。串口通信线程接收用户消息处理线程发过来的指令，通过串口通信，发送消息给单片机，由单片机驱动步进电机控制云台转动，方便用户对婴儿的状态进行监控。与云服务器通信线程使用 HTTP 协议实现智能监护系统与云服务器的通信，当智能监护系统监听到婴儿哭声后，与云服务器通信线程会向远端的云服务器发送消息，云服务器收到消息后会向指定的用户推送报警信息，完成一次报警。音频发送线程和视频发送线程分别用来实现视频信号和音频信号的发送，这里通过使用物联智慧公司的 P2P 解决方案，实现智能监护系统与用户的终端（如手机、平板等）在广域网内的视频、音频数据的传输，而不需要复杂的网络配置[10]。

音频处理线程是本智能监护系统的核心，音频处理线程的流程图如图 8。音频处理线程首先查看消息队列里是否有开启或关闭音频监控功能的命令，根据收到的命令，开启或关闭音频监控功能。音频监控模块开启后，首先检查音频采集模块是否开启，如果音频采集模块没有开启，则通过消息队列向音频采集线程发送消息，开启音频采集模块。在音频处理线程中设置了 30 秒钟不报警模块，避免连续的触发报警。在 30 秒钟不报警模块没有被使能的情况下，进入声音处理模块，判断输入的信号是不是婴儿哭声信号，如果输入的信号被判定为婴儿哭声信号，则触发报警，音频处理线程会向与云服务器通信线程发送消息，并使能 30 秒钟不报警模块。

声音处理模块流程图如图 9，音频处理线程使用音频采集线程采集到的数据，对信号进行预处理并进行端点检测，对判定为声音段的信号提取 Mel 频率倒谱系数得到测试模板，使用动态时间规整算法对测试模板与参考模板进行匹配的到识别结果，如果信号被判定为婴儿哭声信号则触发报警。

4　实验结果与分析

为了测试算法对婴儿哭声的识别率，选择婴儿哭声和日常生活中常见的声音信号，关门声，咳嗽声和人的说话声进行实验。本实验中录制了 50 段婴儿哭声样本，婴儿哭声样本来自 5 名 1 岁以下的婴儿各 10 段；录制了 50 段关门声，来自于铁门和木质门各 25 段；录制了 50 段咳嗽声，来自于 25 名成年男性和 25 名成年女性的模拟发声；录制了 25 名成年男性和

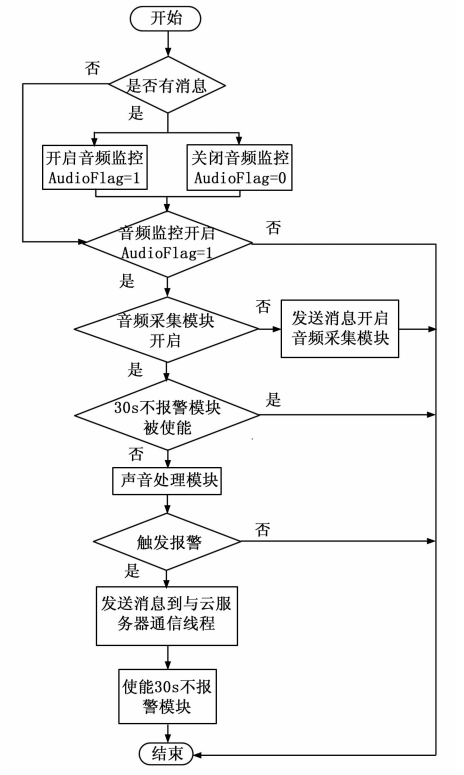

图 8 音频处理线程流程图

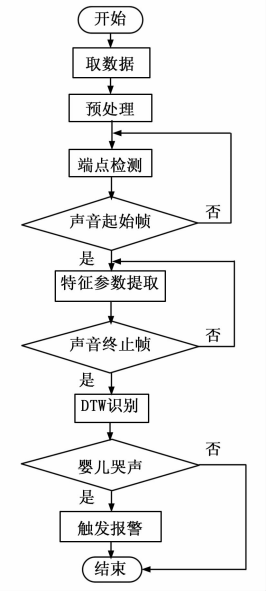

图 9 声音处理模块流程图

25 名成年女性对阿拉伯数字 1，2，3，4，5 的发音各 50 段。分别将这些样本分为两组，第一组 5 个样本，用来提取特征参数，组成参考模板库，这组样本要由来自于不同来源的样本组成；第二组 45 个样本，用来测试算法的识别率。

本实验的 AD 转换器的采样率设置为 8 kHz，采样精度为 16 位，预加重系数为 0.975，帧长为 256 点，帧移 96 点，对每一帧信号加汉明窗，Mel 滤波器组取 24 个 Mel 滤波器，提取 12 阶 Mel 频率倒谱系数作为特征参数。

第一组实验提取 12 阶 Mel 频率倒谱系数作为特征参数，使用动态时间规整算法作为识别算法，实验结果如表 1。

表 1 动态时间规整算法识别率

	婴儿哭声	关门声	咳嗽声	1	2	3	4	5
样本个数	45	45	45	45	45	45	45	45
报警次数	40	3	4	3	2	3	3	2
报警率	88.9%	6.7%	8.9%	6.7%	4.4%	6.7%	6.7%	4.4%

实验表明，该系统可以对婴儿哭声进行有效的识别，对关门声、咳嗽声和人的说话声有较低的误报率。

第二组实验提取 12 阶 MFCC 作为特征参数，使用改进动态时间规整算法作为识别算法，实验结果如表 2。

表 2 改进的动态时间规整算法识别率

	婴儿哭声	关门声	咳嗽声	1	2	3	4	5
样本个数	45	45	45	45	45	45	45	45
报警次数	42	2	2	3	1	2	2	2
报警率	93.3%	4.4%	4.4%	6.7%	2.2%	4.4%	4.4%	4.4%

实验表明，使用改进的动态时间规整算法，放宽对搜索路径的起点和终点约束，提高了测试模板与参考模板的对准精度。对婴儿哭声的识别率由 88.9%提高到 93.3%，有效的提高了算法对婴儿哭声的识别率；对关门声，咳嗽声，和阿拉伯数字 2、3、4 的发音的误报率都有所降低，有效的降低了算法的误报率。该系统对婴儿哭声有较高识别率。

5 结束语

本文通过使用 RT5350 芯片作为处理器，设计了一种能对婴儿哭声进行识别的婴儿智能监护系统。通过提取 Mel 频率倒谱系数作为特征参数，使用动态时间规整算法作为识别方法，实现了对婴儿哭声的准确识别。该系统可以与云服务器以及用户的终端设备通信，将报警消息及时的推送给用户，并能将视频数据和音频数据通过互联网实时的传送给用户，极大的方便了用户对婴儿的监护，有较高的实用价值。

参考文献：

[1] 袁 勇. 多功能婴儿监护系统设计 [J]. 电子世界，2014 (9)：121 - 122.

[2] 杨振雷. 智能婴儿监护系统设计 [J]. 电子世界，2012 (22)：127 - 128.

[3] 李国城，陈佳明，李伟林，等. 基于物联网的婴儿实时监控系统的设计 [J]. 电子设计工程，2015，18：186 - 189.

[4] 李艳雄. 一种能识别婴儿哭声的婴儿监护器及婴儿哭声识别方法：中国，103489282A [P]. 2014－01－01.

[5] 魏丽娜. 婴儿情绪信息的模式识别技术研究与实现 [D]. 上海：复旦大学，2012.

[6] 韩纪庆，张 磊，郑铁然. 语音信号处理 [M]. 北京：清华大学出版社，2004.

[7] 周璐璐，邓江洪. 一种机器人智能语音识别算法研究 [J]. 计算机测量与控制，2014，22 (10)：3267 - 3269.

[8] 周 萍，李晓盼，李 杰，等. 混合 MFCC 特征参数应用于语音情感识别 [J]. 计算机测量与控制，2013，21 (7)：1966 - 1968.

[9] 徐利军. 基于 DTW 的孤立词语音识别研究 [J]. 软件导刊，2012. 11 (2)：137 - 139.

[10] 高 进. TUTK 开启云连线新时代 [J]. 计算机与网络，2014，40 (3)：47.