

# 基于神经网络的家用型睡眠呼吸暂停综合症监测系统<sup>①</sup>

孟泽民, 林生生

(浙江工业大学 信息工程学院, 杭州 310023)

**摘要:** 本文介绍了一款新的 SAS 监测系统, 系统在不添加任何设备的情况下, 利用手机调用自行设计的手机软件采集用户的鼾声数据, 并采用 ftp 协议, 利用家庭无线网传输数据, 最终将数据存储于 PC 机中, 并将神经网络算法和语音识别技术加入到系统的核心算法中, 用于识别语音和鼾声, 结合 SAS 的病征实现对 SAS 病症的诊断. 实践证明, 本系统相对于传统的 SAS 监测系统, 具有更高的病症判别率.

**关键词:** SAS; ftp; Mel 倒频系数; K-均值; RBF 神经网络

## Home Sleep Apnea Syndrome Observation System Based on Neural Network

MENG Ze-Min, LIN Sheng-Sheng

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China)

**Abstract:** This paper introduces a new SAS monitoring system. The system utilizes a self-designed app invoked by mobile phones to collect the data of users' snore. Then it transmits the data via home wireless network by means of ftp and finally stores the data in a PC. Moreover, the neural net algorithm and voice-recognition technology have been inserted into core algorithm of the system to identify voice and snore, which can implement the diagnosis of the SAS by combining the analyses of symptoms. The system has a higher disease classification rate than a traditional SAS monitoring system.

**Key words:** SAS; ftp; Mel cepstrum; k-means; RBF neural network

睡眠呼吸暂停综合症(sleep apnea syndrome, SAS)是指每晚 7 小时睡眠中, 口或鼻腔气流持续停止 10s 以上, 并超过 30 次者. 其重要特征为呼吸停顿与睡眠中断交替出现, 是发病较高并且具有一定潜在危险的疾病. 目前, SAS 诊断的标准是多导睡眠图(Ploysomnography, PSG)分析, 但检测过程较为复杂并且要求患者在医院且在专业人员的看护下整晚测量数据. 然而在生活压力日益增加的现代生活中, 对于出现了上述病症的病人很少能及时发现并配合医生进行治疗.

与此同时, 随着科学技术的迅速发展, 家用型健康监测系统的研制越来越受到研究人员的重视. 在多数现有的关于 SAS 的家用型监测系统中, 基本以外加传感器作为主要检测手段, 在少数通过鼾声作为分析依据的研究中多以时域分析为主, 缺少频域特征量的利用, 往往会忽略梦话等语音信号的干扰而形成误判,

使得整个系统的诊断效果不佳.

本文介绍了一种家用型睡眠呼吸暂停综合症监测系统, 在不添加额外硬件设备的基础上, 利用手机终端采集声音信号, 采用家庭局域网将数据传送到电脑端进行数据处理. 通过提取鼾声的有效信息进行神经网络训练, 有效地判别鼾声和梦话, 更加准确地计算鼾声的出现时间及间隔时间, 并识别是否患有 SAS, 且不影响用户的正常睡眠.

### 1 系统核心算法原理

在系统的实现过程中, 涉及的技术和算法有: 手机数据采集、数据传输、PC 数据存储及数据处理. 其中数据处理的算法是整个系统的核心.

数据处理算法如图 1 所示, 其中包括数据的时域处理、频域处理、有效信号提取, 并结合神经网络, 实

<sup>①</sup> 收稿时间:2013-08-10;收到修改稿时间:2013-09-12

现语音和鼾声的区分。

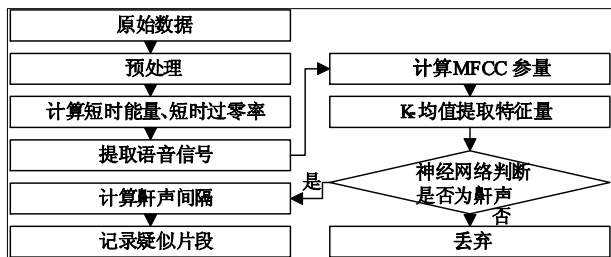


图 1 语音数据处理框图

1.1 时域数据处理

时域数据处理需要完成从环境音中提取有效片段。其处理过程相对比较简单，包括：预加重、加窗分帧、短时过零率和短时能量计算。

其中预加重可以加重语音的高频部分，补偿语音信号受到发音系统所压抑的高频部分，增加语音的高频分辨率，便于频域特征量的提取。在实际操作中我们采用一阶 FIR 高通滤波器实现预加重。

加窗分帧处理是将连续信号分割成短段处理，加窗以增加音框左端和右端的连续性。本系统采用 256 个数据为一帧，采用交叠分段的方法使帧与帧之间平滑过渡，保持其连续性。

然后计算时域数据的短时平均过零率和短时平均能量，并通过双门限法来达到提取有效声音片段的目的。处理结果如图 2 所示。

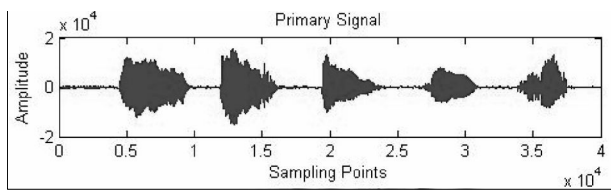


图 2(a) 原语音信号

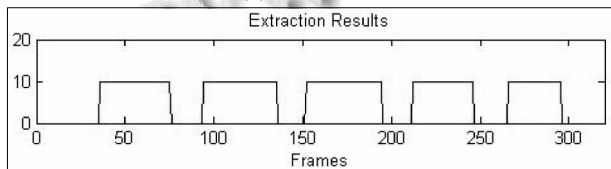


图 2(b) 提取结果

1.2 频域数据处理及 K-均值提取

Mel 频率倒谱系数(MFCC)是语音识别中常用的特征量。这里将被用来识别语音与鼾声。其算法可分以下几步：

① 对上文中的提取结果进行短时傅里叶变换(DFT)得到其频谱和能量分布；

② 加 Mel 滤波器，即用一组三角滤波器在频域对能量进行带通滤波。其中心频率按 Mel 频率的刻度均匀排列(间隔 150Mel, 带宽 300Mel)。Mel 频率计算如式 1 所示，通过 Mel 频率计算再通过三角滤波我们得到对数能量  $E_k$ 。如此处理可以平滑频谱，凸显原先语音的共振峰，将信号的音调和音高滤除，即 MFCC 将不受音调影响。

$$mel(f) = 2595 \times \log_{10}(1 + f / 700) \quad (1)$$

③ 离散余弦转换(Discrete cosine transform, or DCT)，将上述的每帧的 20 个  $E_k$  带入离散余弦转换可以得到 L 阶 Mel-scale Cepstrum，计算如式 2 所示，L 一般取 12。MFCC 系数中 0 阶一般不用，但能反映的是倒谱能量。

$$C_m = \sum_{k=1}^M E_k \cos[\pi(k - 0.5)m / M], \quad m = 1, 2, \dots, L \quad (2)$$

在 MFCC 参数的计算过程中，原信号通过 DFT 转到频域，再通过 DCT 转会到一个叫类似时域，又称 Quefrequency Domain 也就是倒频谱(Cepstrum)。

但由于一个鼾声信号或一个语音信号的开始到结束所包含的帧数不确定，因此它们所包含的 MFCC 参量个数也无法确定，如图 3 所示，但在后续的神经网络判别中需要固定的，因此需要进行进一步的处理，本文中采取 K-均值聚类算法

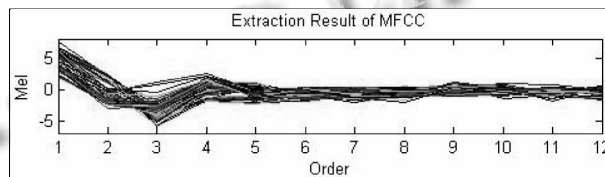


图 3 一个完整语音所有帧的 12 阶 MFCC 参量

K-均值(K-means)算法由 MacQueen 提出，用来解决聚类分析问题，广泛应用于数据挖掘和知识发现领域。这里采用 K-均值算法将杂乱的 MFCC 参量归类，取出能表征的 K 个参量，文中 K 值取 4，其提取效果如图 4 所示

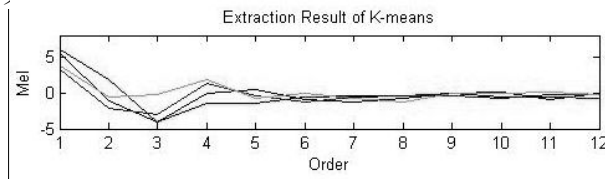


图 4 K-均值提取结果

### 1.3 RBF 神经网络训练

径向基(Radial Basis Function , RBF)神经网络是一种性能良好的前向网络, 它不仅有全局逼近的性质而且具有最佳逼近性能, 在系统辨识和建模方面应用广泛.

RBF 由三层组成, 其结构如图 5 所示, 对于输入信号, 隐层节点会产生局部响应. 本文采用高斯函数(见式 5)作为基函数, 其中  $x$  是  $n$  维输入量;  $c_i$  是第  $i$  个基函数的中心;  $\sigma_i^j$  是第  $i$  个感知的变量,  $m$  是感知单元的个数.

$$R_i(x) = \exp\left[-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right], \quad i = 1, 2, \dots, m \quad (3)$$

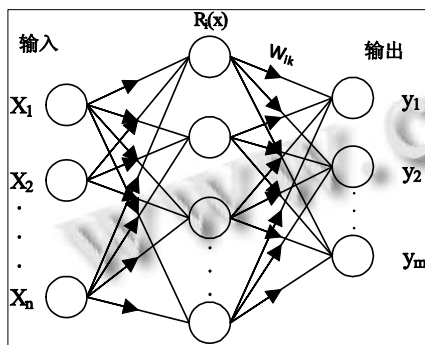


图 5 RBF 神经网络结构示意图

输入层实现从  $x \rightarrow R_i(x)$  的非线性映射, 输出层实现从  $R_i(x) \rightarrow y_m$  的线性映射(式 6), 其中  $r$  是输出节点数,  $w_{ik}$  是权值. 在系统实现过程中, 我们采取对输出值  $y$  进行 2 进制编码来实现分类功能.

$$y_k = \sum_{i=1}^m w_{ik} R_i(x), \quad k = 1, 2, \dots, r \quad (4)$$

## 2 整体方案设计与实现

系统整体方案: 调用手机录音功能采集声音数据, 手机和 PC 机采用 ftp 协议连接, 经家庭路由器(AP)传输到 PC 端, 在 PC 机获取数据后由 C#编写的软件进行数据处理和分析, 并对疑似数据进行存储以供用户参考. 系统框图如图 6 所示.

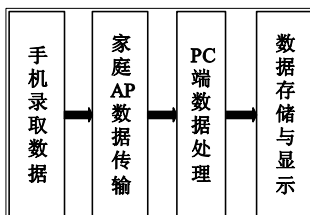


图 6 家用型睡眠呼吸暂停综合症监测系统结构框图

### 2.1 手机端软件实现

手机软件采用 java 语言编写, 通过调用手机自带的录音功能, 采集数据. 为配合 PC 端软件, 手机软件将录入的音频文件重新编码成.wav 格式. 在软件中编写简易 ftp 客户端, 实现与 PC 端相连, 适合 Android 3.2 及以上版本. 操作界面如图 7 所示.



(a) (b)

图 7 (a) 手机软件主界面 (b) 手机简易 ftp 客户端界面

随着手机降噪技术的逐步发展, 环境音对通话质量的影响越来越小. 因此我们通过手机采集到的数据中, 环境音很大一部分已被手机自动滤除. 再经过上文提及的滤波器, 环境音对本系统的影响基本可以忽略. 这也是采用手机采集数据的有利之处. 并且, 上文提及的滤波器将对数据滤波, 消除因手机录音质量不同而带来的误判.

### 2.2 PC 端软件实现

PC 端软件写有简易的 ftp 服务器, 配合手机端接收其传输的数据. 并且 PC 端软件具有播放, 删除文件, 分析声音片段的功能.

#### 2.2.1 ftp 服务器

ftp 协议在局域网传输中运用较为广泛, 其端口号为 20, 21. ftp 具有较全的命令, 在本系统中只需要实现文件的传输和接收, 因此我们在 PC 端软件编写了简易的 ftp 服务器, 配合手机 ftp 客户端进行二进制数据传输, 如图 8 中 5 和 4 两部分所示.

#### 2.2.2 声音文件操作

系统通过添加 Windows 的 Media Player 控件和调用媒体播放相关的 dll 库: axinterop.wmplib.dll 和 Interop.WMPLib.dll, 实现音频播放, 声音波形显示, 暂停播放, 调节音量, 静音等功能, 如下图 8 中 1、2、3 部分所示. 结合列表控件我们可以添加播放列表功

能,用户可以选定一个文件夹,软件能自动识别该文件夹内的.wav 文件并将其添加到播放列表。



图 8 软件主界面

### 2.2.3 声音片段分析

wav 文件作为多媒体声音文件格式之一,由若干个 Chunk 组成的.其中 Data Chunk 是真正保存 wav 数据的地方.通过读取 Data Chunk 部分的数据我们可以得到原始的声音数据.接下来对要做的就是对获得声音片段进行数据分析.分析过程的基础是章节一中描述的核心算法,这里不做过多介绍.分析的结果我们以分析报告的形式展示给用户,如图 9 所示.



图 9 分析报告

## 3 实验结果和分析

为了验证系统的可行性,我们征集了 100 个志愿者的语音和鼾声样本(均来自不同的环境),由于病患样本的缺失,我们采用模拟病患数据进行了实验.我们从中随机抽选 60 个语音和鼾声样本对 RBF 神经网络进行训练,利用剩余的 40 组样本对网络训练结果的验证,实验结果如表 1 所示.

结合表 1 我们可以得出以下结论:

- ① 神经网络不能做到零误差;
- ② 隐藏节点数的增加可以提高系统对语音和鼾声的识别率,但随着节点的增加,识别率将收敛于某个值.

而节点数的增加将导致训练时间变长,系统变复杂.根据此结果,我们认为本系统的最佳隐藏节点数为 10 个;

表 1 RBF 神经网络训练效果对照表

实验序号	隐节点数	训练次数	识别个数		识别率	
			语音	鼾声	语音	鼾声
1	3	1000	15	20	0.375	0.500
2	5	1000	17	19	0.425	0.475
3	8	1000	34	32	0.850	0.800
4	10	1000	36	37	0.900	0.925
5	12	1000	35	34	0.875	0.850
6	10	5000	38	37	0.950	0.925
7	10	6000	35	36	0.875	0.900

③ 训练次数的增加也可以降低系统的误差,随着训练次数的增加,系统的误差将趋于某个值上下浮动.

④ 虽然神经网络的训练占据的时间比较长,但系统的网络一旦建立,便能迅速进行工作.为了使网络能时时更新,用户可根据自己的需求,建立鼾声数据库,并定期更新.

另外 K 均值算法中聚类个数 K 对系统的性能也存在较大的影响,表 2 所示为网络隐节点数为 10,训练次数为 5000 次时,不同聚类个数对识别率的影响.

表 2 K 均值算法中 K 值对识别率的影响表

实验序号	K 值	识别率	实验序号	K 值	识别率
1	2	0.425	4	5	0.925
2	3	0.625	5	6	0.895
3	4	0.9375	6	10	0.875

由表 2 可知在本系统中,聚类数为 4 时效果最佳.

当然,除了上述两点对识别率有影响外,输出节点的编码方式(即模式的种类)对识别率也有着很大的影响.实验表明笼统地将模式分为鼾声和非鼾声的效果不及将模式分为鼾声、语音和环境音.

## 4 结语

本系统在不添加任何设备的情况下,利用手机采集数据,家庭无线网传输数据,PC 机存储数据,并将神经网络应用于鼾声和语音的识别中,实现 SAS 病症的识别与检测.由于系统采用神经网络算法,虽然在算法上大大简化了病症的检测,但同时也给系统带来

(下转第 131 页)

其中,  $\Delta Z_b$  表示相邻的第  $b$  个像素点相对中心像素点的灰度变化量;  $w_b$  表示第  $b$  个像素点的距离权值, 即

$\sum_{b=1}^8 w_b = 1$ . 因此, 清晰度  $V$  的范围为  $[0, 255]$ .

表 1 给出了本文方法和 SIFT 算法的定量分析结果. 综合评价指数(CEI)评价更客观, 它与主观的视觉感受更加一致. 综合评价指数越高, 图像质量就越高, 配准的精度就越高. 清晰度( $V$ )的范围为  $[0, 255]$ , 其值越大, 图像越清晰. 本文方法的 CEI 和  $V$  值相对 SIFT 算法的较高, 说明本文方法的配准效果更好, 与人眼的主观评价结果均一致.

表 1 不同配准方法的图像评价结果

配准方法	评价指标				
	MSE	PSNR	ENTROPY	CEI	V
本文	42.8436	68.7749	16.9978	27.2858	49.8566
SIFT	48.5842	63.47	17.8593	23.3312	47.4658

## 6 结语

本文主要通过解决图像配准阶段的各种畸变问题, 来提高后续图像处理的效果. 基于相机标定法得到的变换矩阵, 很好的解决了图像的各种畸变问题, 较准确的锁定了图像重叠范围. 实验表明, 结合相机标定的图像配准方法配准能力较高, 通过指标定量评价与人眼主观评价一致, 很有说服力, 说明此种配准方法实用且有很好的应用前景.

(上接第 223 页)

了一些不可避免的误差. 为了解决这一误差, 我们需要大量的临床实验, 并建立数据库, 通过时时更新的病症数据, 实现神经网络的自学习和自适应, 从而实现系统的高病症判别率.

总的来说, 在网络发达, 手机普及率高的当今社会, 人们将更加关注自身和家人的健康. 本系统不产生额外的硬件, 经济实用, 是一个值得推广的家用医疗系统.

## 参考文献

- 1 Chen S, Cowan CFN, Grant PM. Orthogonal least squares learning algorithm for radial basis function networks. *Neural Networks, IEEE Trans. on*, 1991, 2(2): 302-309.
- 2 Krishna K, Narasimha Murty M. Genetic k-means algorithm. *IEEE Trans. on Systems, Man, and Cybernetics Part B: Cyber-*

## 参考文献

- 1 Knops ZF, Maintz JBA, Viergever MA, et al. Normalized mutual information based registration using kmeans clustering and shading correction. *Medical Image Analysis*, 2006, 10(3): 432-439.
- 2 Ville O, Janne H. Image registration using blur invariant phase correlation. *IEEE Signal Processing Letters*, 2007, 14(7): 449-452.
- 3 Lucchese L, Leorin S, Cortelazzo GM. Estimation of two dimensional affine transformations through polar curve matching and its application to image mosaicking and remote sensing data registration. *IEEE Trans. on Image Processing*, 2006, 15(10): 3008-3019.
- 4 周亚平. 图像配准技术研究. 西安: 西安电子科技大学. 2008.
- 5 赵宣铭. 一种基于 Tsai 法的摄像机改进标定法. *西安工程大学学报, ISTIC*, 2011, 8: 562-563.
- 6 张虎. 机器视觉中二维图像的三维重建[硕士学位论文]. 北京: 北方工业大学, 2006: 37-38.
- 7 尚倩, 阮秋琦, 李小利. 双目立体视觉的目标识别与定位. *智能系统学报*, 2011, 6(4): 303-311.
- 8 Kuglin CD, Hines DC. The phase correlation image alignment method. *IEEE Conference on Cybernetics and Society*. New York. 1975. 163-16.
- netics, 1999, 29(3): 433-439.
- 3 Bilmes JA, Bartels C. Graphical model architectures for speech recognition. *Signal Processing Magazine, IEEE*, 2005, 22(5): 89-100.
- 4 王昊, 刘静. 睡眠障碍的手机检测方法研究. *北京生物工程*, 2010, 29(3): 270-277.
- 5 包亚萍, 郑骏, 武晓光. 基于 HMM 和遗传神经网络的语音识别系统. *计算机工程与科学*, 2011, 33(4): 139-144.
- 6 刘靖明, 韩丽川, 侯立文. 基于粒子群的 K 均值聚类算法. *系统工程理论与实践*, 2005, 6(3): 55-58.
- 7 张雪英. 数字语音处理及 MATLAB 仿真. 北京: 电子工业出版社, 2010.
- 9 余华, 杨露菁, 李启元. 基于径向基神经网络的语音识别技术. *控制工程*, 2009, 16(S1): 90-93.