

适合于盲人使用的信息输入方法的研究与实现^①

Research and Implementation of Information Input Method Suitable for the Blind

王 一 滕召胜 温 和 王璟珣 (湖南大学 电气与信息工程学院 湖南 长沙 410082)

摘 要: 提出了 Braille 立体盲文的平面化输入方法,研究了非特定人孤立词语音识别技术,建立了一种新型盲人专用信息输入方法。该方法将 Braille 立体盲文抽象为平面符号,实现盲文的平面化输入及识别,并将连续隐马尔可夫模型的语音识别算法应用于非特定人孤立词语音系统中,利用嵌入式 DSP 系统实现了语音与盲文信号采集、处理和识别,为盲人信息输入提供了新的途径。实验结果表明,该系统语音指令识别率 > 92%,盲文字符识别率 > 91%。

关键词: 盲文识别 语音识别 Braille 立体盲文 隐马尔可夫模型 嵌入式 DSP 系统

随着计算机处理能力的不断提高,语音文字识别技术得到了迅速发展,支持语音文字识别的各种产品纷纷面世^[1]。然而日益高科技化的产品与服务并没有使社会的某些群体得到更好的关怀与照顾。如何使科技发展更好地服务于残疾人,是科技工作者应该努力的重要方向。研究语音识别及盲文识别技术,开发实用的语音识别和盲文识别系统,对于高科技服务残疾人具有十分重要的意义。

本文研究了适合于盲人使用的信息输入方法,提出并设计了一个全音控、带有盲文识别功能的嵌入式 DSP 盲人通讯系统。系统采用连续隐马尔可夫模型(CHMM, Continuous Hidden Markov Model)作为语音识别算法,实现了非特定人孤立词语音识别。系统盲文输入采用触摸板手写方式,利用立体盲文抽象法及组合模板匹配法实现了对于六点式 Braille 立体盲文字符的平面化输入及识别。同时操作无线 GSM 模块,从而实现通讯功能。

1 盲人通信系统的构成

适合于盲人使用的信息输入方法是通过盲人通信系统来实现的,其结构框图如图 1 所示。系统采用 ADI 公司高运算速度的 Blackfin 系列数字信号处理器

ADSP-BF533 作为核心处理器^[2],以 AD1836A 作为音频编解码芯片,四线电阻式触摸屏作为字符输入设备,以及一个 GSM 模块实现无线通讯功能。

盲人用户语音信息通过麦克风和 AD1836A 音频编解码芯片传入 ADSP-BF533 的语音处理模块分析和识别,根据解析所得的语音指令控制字符处理模块及无线 GSM 模块完成通讯。

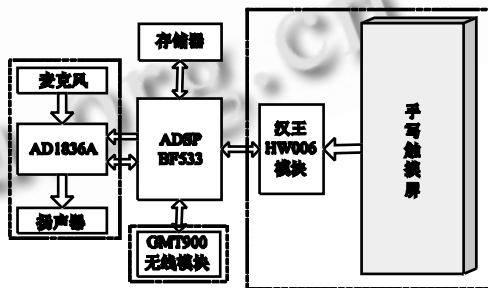


图 1 盲人智能通讯系统结构框图

1.1 盲文处理模块

盲文处理模块的字符输入为触摸屏手写输入方式。系统采用 3.5 寸四线电阻式触摸屏,盲人用户使用手写笔在触摸屏上书写盲文字符后,其坐标数据经 4 根数据线传输至汉王 HW006 模块,此模块可将原

① 基金项目:国家自然科学基金(60872128)

收稿时间:2009-06-02

始坐标数据经 UART 接口送出，系统对这些数据进行盲文识别处理后可实现盲文平面化输入。

1.2 语音处理模块

语音处理模块采用 AD1836A 作为音频编解码芯片，同时选取一个驻极体微型麦克风和扬声器作为语音输入输出媒介。AD1836A 与 BF533 的数据通信采用 SPORT 接口，同时 SPI 接口被用来对 AD1836A 进行配置。

语音数据的采集通过按键来实现。当输入语音信号时，将该键按下，系统检测到按键按下时使能 SPORT 和 DMA 中断，语音采样数据由 DMA 控制器存入开辟好的缓存区。语音输入完毕后松开按键，当系统检测到按键弹起时关闭数据接收中断。这种方式优点是语音信号长度具备较大灵活性，采集完毕后可以马上进行处理。原始语音数据经语音识别处理后形成语音控制指令，控制系统进行下一步操作。

设计中，系统还利用数字录音技术，将一级常用汉字库和 10 个数字及 26 个字母(共 3817 个)存入 NANDFLASH 存储器。并编写了函数实现语音播报，即可轻松完成盲人用户操作时的各种语音提示。

1.3 无线 GSM 模块

系统通讯部分采用华为技术有限公司高性能高稳定工业级 GSM 无线模块 GTM900。GTM900 是一款三频段 GSM 的无线模块，它支持标准的 AT 命令及增强 AT 命令，提供丰富的语音和数据业务等功能，是 GSM 通信应用的理想解决方案。在盲人智能通讯系统架构内，GTM900 模块是实现电话拨打接听、短消息接受发送等通讯功能的直接载体。

2 Braille立体盲文的平面化识别

本系统所采用的盲文为国际通用的 Braille 盲文。此种盲文以 6 个圆点为基本结构组成，这 6 个点在纸面上有的凸起，有的不凸起，形成 64 种变化，即 64 种符形。每个符形有左右两列，每列各三个点，从左边自上而下叫做 1、2、3 点，从右边自上而下叫做 4、5、6 点。盲人可以通过触摸感知凸起点从而进行阅读或书写^[3]。本系统对盲文字符的处理流程如图 2 所示。

2.1 平面化处理

由于盲文是一种特殊的立体式文字，因此在识别过程中必须对其进行特定处理。本设计中约定“+”代表凸起点，“O”代表未凸起点，因此盲人用户只要在触摸屏上使用手写笔输入这两种字符，便可像正常人般完成文字信息的计算机输入。例如在输入字母“L”时只需要在触摸板上使用手写笔输入如图 3 所示字符，这样就完成了 Braille 立体盲文的平面化处理。



图 2 盲文字符识别流程图

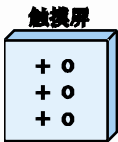


图 3 盲文字母“L”在触摸屏上输入示意图

2.2 特征提取

对盲文进行平面化处理后，系统将首先对“+”和“O”进行特征参数提取。

系统采用的特征参数提取过程如下：①根据触摸屏返回的坐标数据，提取上、下、左、右四个端点，从而可以确定一个包围了手写字符的矩形区域；②将 X、Y 坐标区间五等分，从而把矩形区域分为 5×5=25 个小矩形区域，对应一个特征数组；③判断原始坐标数据落于哪些小矩形区域内，如果某个区域内有触摸数据，在特征数组中对应处置 1，反之则置 0，从而得到一组特征向量。

2.3 匹配识别

系统对字符采用组合模板匹配识别方式。预先通过多次手写盲文字符，对每个字符得到多个特征向量模板。实验表明当字符模板数超过 10 个后，识别率没有显著提高，因此综合识别率和计算时间，系统设计中每个字符采用 10 个特征向量模板。

当进行识别时，输入的盲文字符经处理转化为新的特征向量后，系统对这个特征向量分别与模板库中各特征向量作欧式距离判断。得到距离最小的一个特

征向量后，即可以判断出该字符是“+”还是“O”。由于平面化盲文是这两个字符的固定组合，因此可再通过综合判断得到对应的字母。

3 基于CHMM的语音识别

语音识别涉及到的内容非常广泛，设计中必须根据不同的需要来定义适当的语音识别范围。本系统中，语音识别主要实现小词汇量、非特定人、孤立词的识别。因此，系统选用适合非特定人语音识别的连续隐马尔可夫模型，事先采集足够多且多样化的数据将模型训练好，任何用户都可以不需训练直接输入语音指令进行操作。系统的语音指令处理流程如图 4 所示。

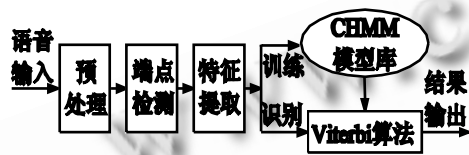


图 4 语音识别流程图

语音识别首先要求系统检测出语音信号的起始点和结束点。常见方法有时域、频域和语谱分析等^[1]。考虑到端点检测的实时性要求，本系统采用时域分析方法，即根据声音信号的短时平均幅度和短时过零率，经过信号预处理后采用双门限法来实现语音信号和噪声的区分。

语音信号特征参数的提取是实现语音识别的一个关键点。系统选用 MFCC 参数作为语音信号的特征参数。大量实际验证表明，选取 MFCC 参数作为特征参数进行语音识别效果最好，且 DSP 强大的运算性能完全可以克服 MFCC 参数计算量大的缺点^[4]。因此，系统选用 MFCC 参数作为语音信号的特征参数。

运用 CHMM 进行孤立词的识别时，首先要为每一个孤立词准备一个隐马尔可夫模型，这需要通过模型训练来实现。在得到一个未知语音特征向量转化成的观察值序列后，采用 Viterbi 解码算法计算这组观察值序列在每个 CHMM 上的输出概率，输出概率最大的 CHMM 对应的孤立词就是识别结果。系统中，CHMM 的训练是在上位机，用 MATLAB 进行的。首先是每个语音指令语料的准备。通过特征参数计算得到每个命

令的 MFCC 参数后，开始进行模型训练。训练算法采用的是 Baum-Welch 算法。Baum-Welch 算法利用递归的思想，使用大量训练数据，重复由重估公式得到的新模型，而每一个新模型的参数总要好于上一次训练出的模型参数，直到新模型参数产生的概率不再明显增大，即认为模型已经训练好^[5,6]。由上位机训练出模型后，将参数传递给 DSP 核心处理器。

DSP 识别部分主要是 Viterbi 解码算法的实现，即利用训练好的 CHMM，对于一个新输入的语音信号，首先得到它的特征向量，然后计算出它经过每一个 CHMM 后 Viterbi 算法解码的输出概率。选取输出概率最大的一个作为识别结果。由于基于连续高斯混合密度 HMM 的语音识别需要很高的数据精度，因此，系统实现中将 DSP 的 DOUBLE 型变量设置为 64 位型。为了保证识别速率，经过实际验证，仅在计算 3 个高斯混合元参数和加权混合概率密度时采用了部分 DOUBLE 型变量。

4 实验结果

4.1 语音识别实验

孤立词的识别是语音识别的一个基本问题，而由孤立词组成的一系列语音控制指令成为语音识别的一个重要应用。

在本系统测试实验中，采用系统 25 个语音指令作为测试对象，共用 18 男 9 女在干净环境下的语音，从中选取 9 男 9 女作为圈内说话者，每人每音发 20 遍，前 10 遍作训练集，后 10 遍作圈内测试集；另外 9 男作为圈外说话者，每人每音发 10 遍，只作圈外测试集。其部分实验结果如表 1 所示。

表 1 语音识别率统计表

语音控制指令	圈内测试集	圈外测试集
写短信	99.8%	93.3%
重写	94.6%	90.3%
退格	95.7%	93.7%
完成	95.3%	92.2%
发送	91.2%	88.7%
返回	95.9%	94.1%

其中语音信号以 8kHz 速率采样，量化精度为 16bit，帧长 256 点，帧移 80 点，特征参数取 12 维

美尔频率倒谱系数、12 维美尔频率差分倒谱系数共 24 维，CHMM 采用状态数为 5、高斯混合度为 3 的左右无跳转结构。

从表 1 中可以看出，由于圈外测试集的说话者未参与训练，其识别性能较之圈内说话者有所下降，但系统整体识别率仍在 92%以上。

4.2 盲文识别实验

在盲文抽象符识别实验中，共用 10 男 9 女的笔迹，从中选取 5 男 5 女作为圈内输写者，每人每符号 11 次，其中 1 次共 20 个笔迹形成 20 个模板，剩余 10 次作为圈内测试集；另外 5 男 4 女作为圈外输入者，每人每符号 10 次，形成圈外测试集。其实验结果如表 2 所示。

表 2 盲文抽象符识别率统计表

盲文抽象符	圈内测试集	圈外测试集
+	98.7%	98.6%
0	98.5%	98.5%

从表 2 可以看出盲文抽象符识别不同与语音识别，因为符号笔画简洁，不同人书写差异性不大。但 Braille 盲文体系中，一个盲文字母由 6 个抽象符号组成，故系统整体盲文识别率为 $(98.5\%)6=91.3\%$ 以上。

5 结论

本文提出了 Braille 立体盲文的平面化输入与识别方法，阐述了基于连续隐马尔可夫模型的语音识别

算法在非特定人孤立词语音系统中的应用，建立了适合于盲人使用的盲文信息输入平台，并给出了盲人智能通讯系统的多信息(语音、盲文和触觉感知等)处理方法，利用嵌入式 DSP 系统研制了多信号采集装置，进一步推动了盲人智能通讯设备的产业化进程。

本设计已获得 ADI 中国大学生创新设计竞赛“二等奖”和“最佳原创设计奖”。系统语音指令识别率高于 92%，盲文字符识别率高于 91%。系统设计产业化后可为盲人群体提供接拨电话、接发短信、信息查询等便捷的通讯服务，具有良好的发展前景。

参考文献

1 韩纪庆,张磊,郑铁然.语音信号处理.北京:清华大学出版社, 2004.

2 ADSP-BF533 Blackfin Processor Hardware Reference. Analog Devices, Inc.2006.201 – 584.

3 杨淑莹.模式识别与智能计算.北京:电子工业出版社, 2008.

4 张雄伟,陈亮,徐光辉.DSP 芯片的原理与开发应用.北京:电子工业出版社, 2004.

5 Kanevsky D. Extended Baum transformations for general functions. Proceedings of ICASSP, 2004,10 (1):38 – 45.

6 赵博.MATLAB 在语音分析中的应用.计算机系统应用, 2005,14(2):34 – 37.