

DSP 语音识别电梯控制系统的设计和实现

陈卫兵, 何必都, 黄永坤, 邹豪杰, 张洪波

(湖南工业大学 计算机与通信学院, 湖南 株洲 412007)

摘要: 针对多人乘坐电梯时按键不是很方便的问题, 设计了一种简单的 DSP 语音识别电梯控制系统。以 TMS320C6713 DSP 芯片作为系统运算控制中心, TLV320AIC23B 芯片作为语音输入、输出的模拟前端, EP2C5Q208C8 FPGA 芯片作为系统 IO 扩展, 以这 3 个部分的有机结合实现了非特定人、孤立词、小词汇量的嵌入式语音识别电梯控制。测试结果显示: 所设计的系统结构合理, 语音识别电梯控制准确度达 80%。

关键词: 语音识别; 数字信号处理器; 端点检测; 动态时间弯折算法

中图分类号: TP202; TP273

文献标志码: A

文章编号: 1673-9833(2011)04-0101-04

Design and Implementation of DSP Speech Recognition Controlling System in Elevators

Chen Weibing, He Bidu, Huang Yongkun, Zou Haojie, Zhang Hongbo

(School of Computer and Communication, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract : Aiming at the inconvenience of many people pressing elevator buttons, designed a smart DSP voice recognition controlling system for elevators. With TMS320C6713 as the system operation control center, TLV320AIC23B as the analog end of voice input and output and EP2C5Q208C8 FPGA as the system I/O expanding port, the system was integrated to achieve non-specific people, isolated word and small vocabulary embedded speech recognition elevator controlling. Test results show that the system architecture is logically arranged and speech recognition accuracy of the elevator control reaches 80%.

Keywords : speech recognition; DSP; endpoint detection; DTW

0 引言

语音识别是将原始语音经过预处理后进行特征提取, 再与事先经测试和训练后所得并存储到计算机的标准参考模型进行比较, 最后得出判定和识别结果。20 年来, 语音识别技术取得了较大发展, 它经历了从孤立词、小词汇量、特定人大词汇量、非特定人的发展历程^[1]。但语音识别的计算量较大, 难以实时实现, 此问题制约着它在各个领域内的应用。

目前, 随着数字信号处理 (digital signal processing, DSP) 专用集成电路技术的迅速发展^[2-3], 语音识别, 尤其是计算量较小的非特定人的孤立词识别的实时实现成为可能。

电梯行业中, 传统的电梯控制器要求人们通过按电梯的楼层按钮来确定需要到达的区域。当电梯中人数较多、较拥挤时, 按键很不方便。考虑到语音的非接触传输特点, 将语音识别和控制技术用于电梯中, 将使传统的电梯更加人性化、便捷化。因

收稿日期: 2011-05-05

作者简介: 陈卫兵 (1966-), 男, 湖南湘潭人, 湖南工业大学副教授, 博士, 主要从事嵌入式系统方面的教学与研究,

E-mail: 26895387@qq.com

此,本文拟以TMS320C6713 DSP芯片作为系统运算控制中心,TLV320AIC23B芯片作为语音输入、输出的模拟前端,EP2C5Q208C8 FPGA芯片作为系统IO扩展,将这3个部分有机结合,开发基于非特定人、孤立词、小词汇量的嵌入式语音识别电梯控制系统,以解决多人乘坐电梯时按键不方便的问题。

1 系统设计方案

图1 为电梯语音识别控制系统的结构框图。

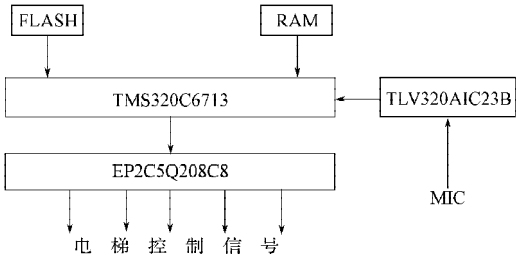


图1 系统总体结构框图

Fig. 1 Block diagram for the overall system

从图1可看出,所设计的语音识别控制系统主要由TMS320C6713 DSP语音识别处理芯片、TLV320AIC23B语音输入/输出的模拟前端芯片、电梯控制接口扩展FPGA芯片和其他辅助芯片(外部存储器SDRAM选用MT48LC4M16A2,大小为64 MB,对应地址为:0x80000000H~0x82FFFFFFH。

FLASH采用AM29LV800B,大小为2 MB,对应地址为0x90000000H~0x90200000H组成。TMS320C6713为高性能32位浮点DSP,适用于专业音频信号处理,主频达300 MHz,处理速度达2400 MIPS/1800 MFLOPS,能满足快速运算和处理时间的语音识别要求。

TI公司的TLV320AIC23B是一款集成ADC(application data center),DAC(digital analog canverter)于一体的模拟接口电路,采用先进的Sigma-delta过采样技术,可在8 kB~96 kB采样率范围内提供16, 20, 24, 32位采样,ADC和DAC的信噪比可分别达90 dB和100 dB。其与外围音频输入设备的接口电路如图2所示。

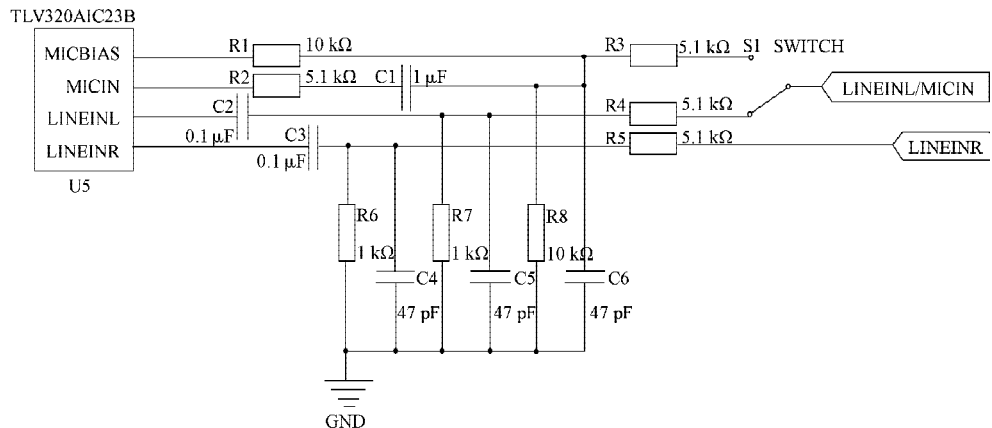


图2 音频输入电路

Fig. 2 Audio input circuit

TMS320C6713与TLV320AIC23B连接方式见图3。

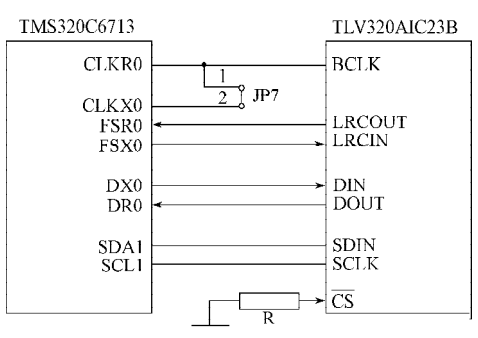


图3 TMS320C6713与TLV320AIC23B硬件连接

Fig. 3 The hardware connection between TMS320C6713 and TLV320AIC23B

TLV320AIC23B的控制口用于设置其工作参数,

采用I²C总线接口实现;数据口用于传输TLC320AD50B的A/D,D/A数据,TLV320AIC23B的数据口与TMS320C6713的McBsp0接口连接,用于芯片间的数据交换。

2 算法及软件实现

2.1 语音识别算法

语音识别系统的总体方案见图4。

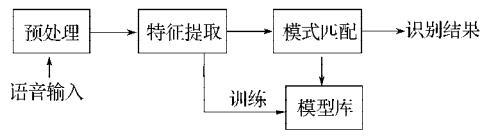


图4 语音识别方案

Fig. 4 The program of speech recognition

语音识别系统首先将收集到的语音信号进行预处理,包括预加重、加窗、端点检测等;然后进行特征提取,即从语音波形中提取出随时间变化的语音特征序列;最后,将其训练为声学模型,并且在模式匹配中运用识别算法进行匹配,得到最佳识别结果。

2.2 端点检测

端点检测之前先对采集到的语音信号进行预加重,以去除语音信号中的低频噪声,然后选用 hamming 窗做加窗处理。主要依据为 hamming 窗主瓣比矩形窗的主瓣宽度大 1 倍,同时其带外衰减也比矩形窗大 1 倍多,因而不会损失信号中的高频成分。加窗后将语音信号分割为帧。

端点检测就是从含噪声的信号中检测出说话人语音信号的起始点和结束点。只有正确检测出语音信息段才能正确地进行语音处理。端点检测的时域处理方法是:首先在当前环境情况下,采集一段无声语音,求出平均过零率。由于采集声音信号的最初的短时段为无语音段,仅有均匀分布的背景噪声信号。这样就可以用已知为“静态”的最初几帧(一般取 10 帧)信号计算其过零率阈值,以此作为无语音段,当过零率变化时作为语音的起始。语音结束点的获得方法与此相同,从后向前搜索,当超过过零率的变化时作为语音的结束。

2.3 特征提取

特征参数提取^[4]是指从语音信号中抽取有效的语音信号特征,提取算法如下:

- 1) 对信号进行短时傅里叶变换得到频谱。
- 2) 求频谱幅度的平方得到能量谱,再用一组三角形滤波器在频域内对能量谱进行带通滤波;设滤波器数为 M ,滤波后得到的输出为 $X(k)$, $k=1,2,\dots,M$ 。
- 3) 对滤波器组的输出取对数,然后对它做 $2M$ 点逆离散傅里叶变换,得到 Mel 频率倒谱系数 (mel frequency cepstrum coefficient, MFCC)。因为对称性的关系,变换式可简化表示为:

$$C = \sum_{i=1}^M \text{Log} X(k) \cos[\pi(k-0.5)n/M], n=1,2,\dots,L。 \quad (1)$$

式中 L 为 MFCC 系数的个数,本系统取 24 个。

2.4 模板匹配

采用动态时间弯折 (dynamic time warping, DTW) 算法^[5]进行模板匹配:假设参考模板的 MFCC 系数向量序列为 $\mathbf{X}=(x_1, x_2, \dots, x_i)$, 输入语音的 MFCC 系数向量序列为 $\mathbf{Y}=(y_1, y_2, \dots, y_j)$, $i \neq j$ 。DTW 算法就是要寻找一个最佳的时间规正函数,使待测语音的时间

轴 j 非线性地映射到参考模板的时间轴 i 上,因而总的累计差值最小。算法过程如图 5 所示。

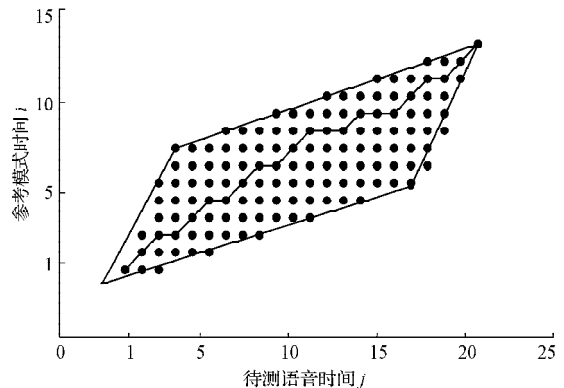


图 5 DTW 算法过程

Fig. 5 DTW algorithm process

图 5 中曲线连接起来的点就是模板与待测语音信号间的距离 $d(x_i(n), y_j(n))$, 亦称为局部匹配距离。DTW 算法就是通过局部优化的方法实现加权距离总和最小,也就是相似度最大,即

$$D = \min \sum_{i=1}^I d(X_{i(n)}, Y_{j(n)}). \quad (2)$$

2.5 软件实现

软件设计以 TI DSP/BIOS 实时多任务操作系统为软件设计基础,采用图像界面配置 DSP/BIOS,在开发环境中自动生成 .cmd 文件。使用 TI 的 TMS320C6000 Chip Support Library API Reference Guide 进行 EMIF, McBsp, PLL 等初始化操作。通过 I²C 总线端口 1 配置 TLV320AIC23B 芯片, McBsp0 口与 TLV320AIC23B 相连实现语音信号采集。端点检测、MFCC 和 DTW 模板匹配用 C 语言编写。系统实现的主流程见图 6。

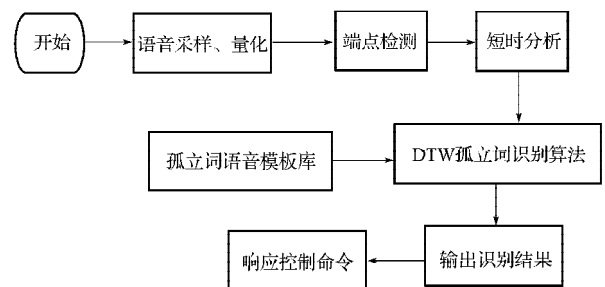


图 6 程序主框图

Fig. 6 The block diagram of main program

TLV320AIC23 内部有 11 个 16 位寄存器,这 16 位控制字中, B[15~9] 为寄存器的地址, B[8~0] 为要写入寄存器的数据。写入 11 个寄存器的数值如下:左声道输入控制=0x17;右声道输入控制=0x17;左耳机通道控制=0x7F;右耳机通道控制=0x7F;模拟音频通道控制=0x1C;数字音频通道控制=0x1;启动控制=0;

数字音频格式=0x4F; 样本速率控制=0x3F; 数字界面激活=0x01; 初始化寄存器=0。设置完成后, 启动 A/D 转换, 将转换后的数据存储在 DSP 的内部存储器中, 每次采样 128 点。数据采集流程见图 7。

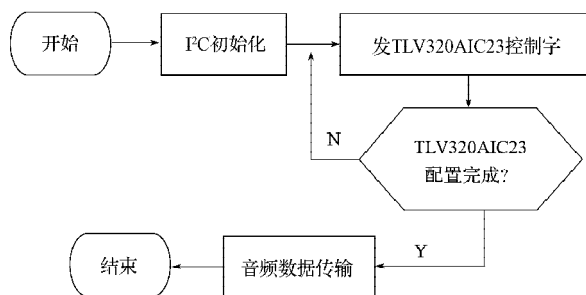


图7 音频数据采集

Fig. 7 Audio data acquisition

3 测试结果

每次测试的采样数为 128 点, 采样频率设为 44.1 kHz, 样本大小为 16 位。在类似电梯的环境中进行测试, 所得结果见表 1。

表1 语音识别电梯控制测试结果

Table 1 Test results of elevator speech recognition control

语音信号	测试样本		识别样本	
	男 10 人, 5 次 / 人	女 10 人, 5 次 / 人	男	女
1 层	50	50	41 (82%)	42 (84%)
2 层	50	50	40 (80%)	41 (82%)
3 层	50	50	39 (78%)	41 (82%)
4 层	50	50	40 (80%)	42 (84%)
5 层	50	50	41 (82%)	42 (84%)
6 层	50	50	40 (80%)	41 (82%)
上	50	50	41 (82%)	43 (86%)
下	50	50	41 (82%)	42 (84%)

从表 1 中可以看出, 所设计的控制系统总的平均识别率大于 80%。且女性的测试结果普遍比男性好, 这应该归属于男性声音中浊音分量比女性重, 导致在系统识别时难度加大。

4 结语

本研究是在以 TMS320C6713 为控制核心, TLV320AIC23B 芯片为语音输入、输出的模拟前端, EP2K5Q208C8 FPGA 芯片为系统 IO 扩展的情况下, 设计和实现了非特定人、孤立词、小词汇量的嵌入式语音识别电梯控制系统。测试结果表明: 所设计的系统的识别和控制效率达 80% 以上, 该系统具有较好的应用前景。

参考文献:

- [1] 赵 力. 语音信号处理[M]. 北京: 机械工业出版社, 2009: 114-117.
Zhao Li. Speech Signal Processing[M]. Beijing: Mechanical Industry Press, 2009: 114-117.
- [2] 周 霖. DSP 通信工程技术应用[M]. 北京: 国防工业出版社, 2004: 145-189.
Zhou Lin. DSP Communications Engineering Technology [M]. Beijing: National Defence Industry Press, 2004: 145-189.
- [3] 邹 彦. DSP 原理及应用[M]. 北京: 电子工业出版社, 2005: 114-117.
Zou Yan. DSP Principles and Applications[M]. Beijing: Electronic Industry Press, 2005: 114-117.
- [4] 侯雪梅, 田 磊. 基于 Mel 倒频特征和 RBF 网络的孤立词语音识别方法[J]. 西安邮电学院学报, 2008, 13(3): 114-117.
Hou Xuemei, Tian Lei. Speech Recognition Method of Isolated Words Based on Mel Cpectrum Feature and RBF Neural Network[J]. Journal of Xi'an University of Post and Telecommunications, 2008, 13(3): 114-117.
- [5] 万 春. 基于 DTW 的语音识别应用系统研究与实现[J]. 集美大学学报, 2002, 7(2): 104-108.
Wan Chun. Research and Application of DTW-Based Speech Recognition[J]. Journal of Jimei University, 2002, 7 (2): 104-108.

(责任编辑: 廖友媛)