

# 基于非特定人车载音响语音控制系统的设计与实现

孙保群<sup>1</sup>, 郭恒飞<sup>2</sup>, 王琼<sup>2</sup>

(1.合肥工业大学 汽车工程技术研究院,安徽 合肥 230009;

2.合肥工业大学 计算机与信息学院,安徽 合肥 230009)

**摘要:** 提出一种语音命令控制车载音响操作的设计方案,以德国 Infineon 公司新推出的具有 DSP 和单片机双核的 SoC 语音处理芯片 UniSpeech-SDA80D51 为核心组成非特定人车载音响语音控制系统,并实现了系统样机的研制。该系统在江淮同悦 SL1102C1 型车载音响上进行了语音控制实测,实测数据表明系统语音识别率可达到 95%。

**关键词:** 非特定人语音识别;车载音响语音控制;UniSpeech-SDA80D51;隐马尔可夫模型

中图分类号: TP29

文献标识码: A

文章编号: 1674-7720(2011)05-0105-04

## Design and realization of automotive audio control system based on speaker-independent speech recognition

Sun Baoqun<sup>1</sup>, Guo Hengfei<sup>2</sup>, Wang Qiong<sup>2</sup>

(1. Automotive Engineering Research Institute, Hefei University of Technology, Hefei 230009, China;

2. School of Computer & Information, Hefei University of Technology, Hefei 230009, China)

**Abstract:** This paper proposes a design of speech control for audio system of automotive. The Infineon's new Soc speech processing chip UniSpeech-SDA80D51 with two centers, DSP and MCU, which as the center of automotive-audio speech control system, realize the speaker-independent speech recognition and control of automotive-audio SL1102C1 on TonYue JAC. After testing the system in the relatively quiet environment, speech recognition accuracy rate can reach 95%.

**Key words:** speaker-independent; speech recognition; audio system of automotive; UniSpeech-SDA80D51; HMM

随着现代电子技术的发展,越来越多的车载电器被加入到车身电子行列中,其在使得汽车性能改善的同时,增加了汽车驾驶操作的复杂度,也给行车过程带来了不安全的隐患。随着语音识别算法的改进和新一代专用语音处理芯片的问世,出现了使用语音控制代替手动控制的车载电器,从而减轻驾驶员手动操作负担,大大提高行车安全系数。

目前我国的车身电子语音控制主要集中在汽车导航系统的应用上,没有充分发挥语音识别技术在车身电子中的应用价值。本文首次提出了一种的以专用语音处理芯片 UniSpeech-SDA80D51 为核心组成非特定人车载音响语音控制系统的设计方案,并实现了系统样机的研制。该系统在江淮同悦 SL1102C1 型车载音响上进行了语音控制实验,实验数据表明系统语音识别率可达到 95%,为下一步项目产品化开发奠定了基础。

### 1 车载音响语音控制系统

非特定人车载音响语音控制系统结构框图如图 1 所示。

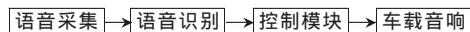


图 1 非特定人车载音响语音控制系统结构框图

系统的主要功能是:语音采集模块(由定向拾音器组成)用于采集驾驶员发出的语音命令信号,由语音识别模块实现信号的 A/D 转换,并对转换的数字信号进行语音识别处理,最终输出与语音命令相对应的词条编码信息。控制模块对接收的词条编码信号进行逻辑分析与处理并产生对应的控制信号,通过系统 I/O 接口驱动车载音响动作,完成驾驶员的语音命令。

#### 1.1 语音识别模块

语音识别模块主要由 UniSpeech-SDA80D51 芯片及外围电路组成。

SDA80D51 是德国 Infineon 公司专为语音识别和语音处理应用领域新推出的专用芯片,采用高集成度的 SoC 系统结构,以  $0.18\ \mu\text{m}$  半导体工艺制造,SDA80D51 的基本结构如图 2 所示。

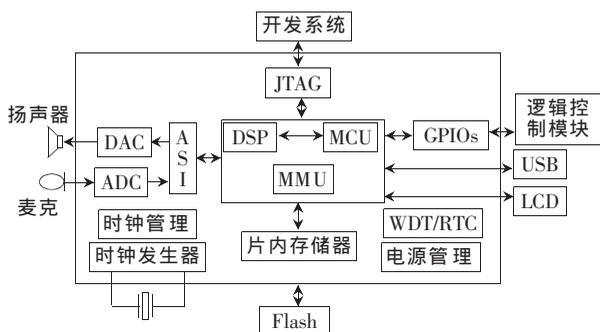


图 2 SDA80D51 基本结构图

SDA80D51 片内集成了直接双访问快速 SRAM、2 路 ADC 和 2 路 DAC、多种通信接口和通用 GPIO 等部件。SDA80D51 工作方式以 M8051 为主控制芯片,主要完成系统配置和 SPI、PWM、I<sup>2</sup>C、GPIO 等接口的控制以及语音数据的传输工作;DSP 核心 OAK 为协处理器,完成语音识别算法、语音编解码算法等语音处理工作。

非特定人语音信号由定向拾音器输入,经过 SDA80D51 内部的数据采集模块进行 A/D 转换,再经过识别程序的预处理、端点检测、特征参数提取、模板匹配等处理,选择识别词表中最接近的词条序号作为识别结果,识别结果通过 GPIO 口输出。

### 1.2 控制模块

控制模块由 MCU 和模拟开关电路构成,本模块主要完成对语音识别模块输出的词条编码信号进行逻辑分析和处理,通过模拟开关电路产生对应功能的控制信号控制音响的动作。其中 MCU 选用美国 ATMEL 公司产品 AT89S51,综合 AT89S51 输出 I/O 信号电压特性和 SL1102C1 音响控制面板电阻式分流键盘电路的特点,确定使用继电器模拟 SL1102C1 控制面板按键的闭合和断开动作。AT89S51 和继电器

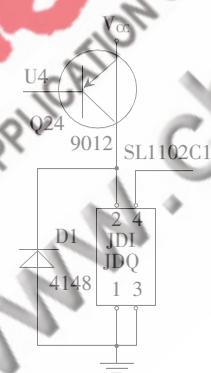


图 3 AT89S51 与模拟开关电路原理图

### 1.3 音响模块

本设计是基于 SL1102C1 型汽车音响。SL1102C1 是专门为中档轿车设计的汽车音响,具有 MP3 播放、收音机和显示时间等功能,目前大量使用在江淮同悦轿车上。SL1102C1 前板共有开关机/静音、音效、播放/暂停等 15 个按键和一个用来调节音量的编码开关。

SL1102C1 前板上的按键为电压采样识别方式,按键包含短按和长按两种动作,AT89S51 输出电压为 TTL 电平,直接采用 I/O 信号驱动音响按键动作容易引起误识

别,造成系统误操作,因此本文采用图 3 所示的模拟开关电路,很好地解决了上述问题。当 AT89S51 接收到语音编码信号后,立即进行逻辑分析并输出对应的控制信号驱动相应继电器吸合模拟按键动作,按键的短按和长按功能是通过软件实现的。

模拟开关电路还适用于 SL1102C1 前板上的编码开关,编码开关具有音量调节功能,其工作原理如图 4 所示。

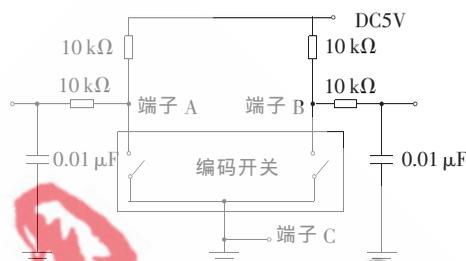


图 4 编码开关原理图

由图 4 可知,编码开关上有 A、B、C 三个端子,开关旋钮左、右旋转时,A、B 端子输出对应的脉冲信号。当 MCU 收到操作编码开关的语音命令信号后,驱动继电器动作,控制端子 A、B 输出信号,模拟开关旋钮功能。

## 2 系统软件设计

系统的软件包括非特定人语音识别模块和逻辑控制模块。

### 2.1 非特定人语音识别模块

非特定人语音识别模块基于 HMM 模型算法。该算法通过对大量语音数据进行数据统计,建立识别词条的统计模型语音库,然后从待识别语音中提取特征,与模型库进行匹配,由比较匹配分数得到识别结果,并通过 SDA80D51 的 GPIO 口输出识别结果对应的词条编码信号。语音识别模块主要由信号预处理、特征参数提取、模型匹配和 Viterbi 算法部分组成,非特定人语音识别模块框图如图 5 所示。

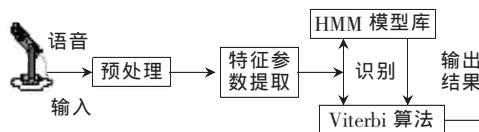


图 5 非特定人语音识别模块框图

#### 2.1.1 信号预处理

信号预处理部分主要完成输入语音信号的采样、模/数转换功能。A/D 变换由 SDA80D51 内嵌 12 位 A/D 变换器实现,采样频率固定为 8 kHz。

#### 2.1.2 特征参数提取

特征参数提取基于语音帧,采用分帧提取特片。先对语音信号进行重叠分帧,前一帧和后一帧重叠一半(帧信号重叠是体现相邻两帧数据之间的相关性),帧长为 25 ms,对每帧提取一次语音特片。

语音信号是声道响应和声门激励信号的卷积。分别求声道传输函数和声门激励信号的对数频率响应,由于声门激励信号的频率响应和声道传输函数在频谱的变化快慢不同,如将频率轴视为时间轴,则声门激励信号的频率响应对应于“高频”区;而声道传输函数对应于“低频”区,处在不同区域就易于分辨。

MFCC 参数属于感知频域倒谱参数,反映了语音信号短时幅度谱的特征。 $p$  维 MFCC 参数的具体计算提取过程如下:

(1)用 DFFT 对每帧  $s(n:m)$  计算线性频谱,计算频谱模的平方为功率谱;

(2)功率谱经过 Mel 滤波器组获得  $D$  个参数  $X(i)$ ,  $D$  是 Mel 滤波器组中三角形滤波器的数量;

(3)对  $X(i)$  做对数运算和离散余弦变换,余弦变换计算公式如下:

$$MFCC_p = \sum_{i=0}^{D-1} Y(i) \cos\left(\frac{\pi(i+0.5)p}{D}\right)$$

式中的  $Y(i)$  是第  $i$  个 Mel 滤波器的对数能量输出,  $i=1, 2, \dots, D$ 。

### 2.1.3 HMM 语音识别算法

隐马尔可夫采用概率统计模型描述语音信号, HMM 模型建立在 Markov 链基础上,使用 Markov 链来模拟语音信号统计特性的变化。HMM 模型为双重随机过程,其一是 Markov 链,由  $(\pi, A)$  描述状态的转移,输出为状态序列;另一个是随机过程,由  $B$  描述,在统计意义上  $B$  反映了状态和观察值之间的对应关系,输出为观察值矢量序列。Markov 链中状态和时间参数都是离散的 Markov 过程。

Viterbi 算法是一种帧同步动态规整算法,在给定观察值序列和模型时, Viterbi 算法给出了一个概率密度  $P(Q, O|\lambda)$  最大的状态序列。Viterbi 算法包括初始化、递推、终止、路径回溯和确定最佳状态序列。

(1)初始化

$$\delta_1(i) = \pi b_i(O_1) \quad 1 \leq i \leq N$$

$$\varphi_1(i) = 0 \quad 1 \leq i \leq N$$

(2)递归

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T, 1 \leq j \leq N$$

$$\varphi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T, 1 \leq j \leq N$$

(3)终止

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)]$$

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)]$$

(4)状态序列

$$q_t^* = \varphi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1$$

(5)Viterbi 评分

$$P^* = \max_Q (P(Q, O|\lambda))$$

对于语音处理而言,因  $Q$  的变化,  $P(Q, O|\lambda)$  取值范

围很大,而  $P(Q, O|\lambda)$  的最大值占了全部  $P(Q, O|\lambda)$  的很大的成分,所以可以用 Viterbi 算法来计算  $P(O|\lambda)$ 。

### 2.2 控制模块

控制模块的主要功能是:在 AT89S51 查询到语音词条信号后,查表获得词条编码,根据编码判断对应按键是长按或短按,分别进入相应的子程序处理。在子程序中,输出语音命令所对应的 I/O 控制信号驱动继电器吸合模拟按键或编码开关动作,并及时复位 I/O 口。控制模块还具有完全兼容手动控制的功能,在语音控制操作的同时也可以进行手动操作,手动的优先级高于语音命令,这样可以避免语音控制和手动控制之间发生冲突。

控制模块部分程序代码如下:

```
void main()
{
    WDTRST=0x1E;
    WDTRST=0xE1; //初始化看门狗
    P1=0xff;
    P2=0xff;
    P3=0xff;
    date=P0;
    while(1)
    {
        WDTRST=0x1E;
        WDTRST=0xE1; //WD 指令
        P1=0xff;
        P2=0xff;
        YXSY=1;
        while(P0==0x00) //等待语音信号
        {
            WDTRST=0x1E;
            WDTRST=0xE1; //WD 指令
            YXSY=0;
        }
        YXSY=1;
        if(date!=P0)
        {
            delays(6);
            date=P0;
            if(date==1||date==2) //开机、静音
            {
                PWR=0;
                delays(200);
                PWR=1;
                .....
                P1=0xff;
                P2=0xff;
            }
        }
    }
}
```

### 3 系统实测结果

本系统在江淮同悦 SL1102C1 型车载音响上进行了非特定人语音识别率和模拟开关动作准确率测试。由于汽车音响的语音词条为 2 到 4 个字,语音识别率实验内容为车载音响常用 2 字词条指令 18 条、3 字词条指令 12 条、4 字词条指令 10 条,实验对象为 6 人(4 男、2 女,普通话和方言),实验环境为实验室环境。为了提高系统的识别率,系统采用奥林巴斯 ME52 定向麦克风,提高了麦克风接收范围,系统测试结果如表 1 所示。

表 1 系统识别率测试结果

测试对象	麦克风	样本数	错识别数	识别率/%
普通话男声	定向	40	1	97.5
普通话男声	普通	40	3	92.5
普通话女声	定向	40	2	95
普通话女声	普通	40	3	92.5
方言男声	定向	40	2	95
方言男声	普通	40	4	90
方言女声	定向	40	3	92.5
方言女声	普通	40	5	87.5

由表 1 可知, 系统的识别率与语音指令词条字数、麦克风接收距离、说话人方言有关。男声和女声的识别率接近。

在系统控制电路实验中, 模拟开关动作达到了较高的准确率, 测试结果为 98% 以上, 只要控制程序运行正常, 各路继电器就能按照程序安排执行闭合和断开模拟手动开关操作。

实现汽车电器的语音控制是未来车载电器的发展趋势, 越来越多的解决方案被提出和验证。本文设计在 SL1102C1 型车载音响上使用 SDA80D51 芯片, 实现了车载音响非特定人的语音识别与控制。由于该芯片集成度高, 需要外围模块少, 所以设计的硬件电路简单, 便于调试检测。该设计得到的样机, 有较高的识别率, 工作稳定、可扩展性强, 达到预期的设计目标, 整个设计方案和实现方法是可行的。由于语音识别率随着环境、说话人不同而变化, 虽然 HMM 算法在噪声很小的环境下可以

获得很高的识别率, 但当测试语音或者环境中含有不同程度的噪声污染时, 语音识别系统的性能会有所下降。提高系统的抗噪性和鲁棒性是语音识别系统走向实用化的关键之一。

#### 参考文献

- [1] 杨行峻, 迟惠生. 语音信号数字处理[M]. 北京: 电子工业出版社, 1995.
- [2] Infineon. UniSpeech2V2.0 Functional Specification [Z]. Infineon Technologies AG, 2002.
- [3] 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.
- [4] 王海青. 基于 CDHMM 的口令式语音识别系统及其 DSP 实现[D]. 合肥: 中国科学技术大学, 2003.
- [5] BURCHARD B, ROMER R, FOX O. A single chip phoneme based HMM speech recognition system for consumer applications[J]. Consumer Electronics, IEEE Transactions on, 2000, 46(3): 914-919.
- [6] Masao Namiki, Takayuki Hamamoto, Seiichiro Hangai. Spoken word recognition with digital cochlea using 32 DSP-boards. IEEE Trans. on Acoust, Speech, Signal Processing, 2001, 2: 969-972.

(收稿日期: 2010-10-10)

#### 作者简介:

孙保群, 男, 1956 年生, 教授, 主要研究方向: 汽车车身电子控制系统, 车载电子电器系统, 车辆自动变速传动等。