

基于 HMM 的嵌入式语音交互在 AmI 中的应用

龚晓庆¹, 柯素娟², 李 辉², 华庆一¹

(1. 西北大学信息科学与技术学院, 西安 710069; 2. 北京工业大学嵌入式软件与系统研究所, 北京 100022)

摘 要: 环绕智能的一个重要目的是实现人与环境的交互更加自然。该文通过语音识别技术体现环绕智能环境中的自然交互的理念, 系统采用隐马尔可夫模型实现语音识别, 同时为了保证嵌入式环境下交互的实时性, 采用 IP 核来设计与实现算法, 并进行仿真验证。实验数据表明系统具有较好的性能, 对语音识别在环绕智能环境中的应用研究有一定的理论和实践意义。

关键词: 环绕智能; 语音识别; 嵌入式; 隐马尔可夫模型; IP 核

Application of Embedded Speech Interaction Based on HMM in AmI

GONG Xiao-qing¹, KE Su-juan², LI Hui², HUA Qing-yi¹

(1. College of Information Science and Technology, Northwest University, Xi'an 710069;

2. Institute of Embedded Software and Systems, Beijing University of Technology, Beijing 100022)

【Abstract】 One of the main aims of Ambient Intelligence(AmI) is to make human interaction with the environment more natural. The concept of natural human machine interaction in AmI is embodied through speech recognition technology. In order to ensure the real-time of the interaction, the system implements the algorithms by using IP core technology. The experimental result shows that the system performs well and it is helpful for the application research of speech recognition technology in AmI.

【Key words】 Ambient Intelligence(AmI); speech recognition; embedded; Hidden Markov Model(HMM); IP core

1 概述

近年来,环绕智能(Ambient Intelligence, AmI)逐渐成为一个新的研究热点^[1]。AmI体现在技术上是计算系统的高嵌入性、智能性和高移动性,具体表现形式为嵌入了计算智能、网络通信和多通道人机交互等技术的智能环境,它具有普适计算(ubiquitous/ pervasive computing)、自然交互、智能(intelligence)和上下文感知(context-aware)^[2]等技术特征。

环绕智能包括了多个计算领域的汇合:(1)“普遍存在”(ubiquitous)的或“普遍深入”(pervasive)的计算,这主要归功于网络能力的大幅提高以及大量的低成本计算设备的发展。(2)智能系统的研究,它提供学习算法、模式匹配、语音识别、语言翻译、手势分类和情境估计等。(3)上下文感知(context awareness)和自然交互(natural interaction),主要研究跟踪和定位任何类型的对象,并且提供对象和环境的交互接口。更进一步,还有在 AmI 环境中对象之间的“社会性”交互,因此自然交互是环绕智能中很重要的一种理念。

由于语音是人机交互最自然的方式,语音处理技术在环绕智能的发展中发挥越来越重要的作用。在笔者的研究项目中,建立了一个具有下列特征的环境智能应用系统:

- (1)提供人机交互工具,这样用户就能得到特定的服务。
- (2)能根据语音来控制 AmI 环境中的被控设备。

基于以上目标以及环绕智能的特点,本文提出了一种嵌入式语音交互技术,通过 AmI 采集用户的语音数据,提取特征向量为每个用户建立特征模板库,进而实现 AmI 环境中的语音交互技术。主要解决的问题和分析的重点在于:

- (1)语音识别模型的建立,本文采用隐马尔可夫模型(Hidden Markov Model, HMM)来建立识别模型。

(2)如何在嵌入式环境中提高语音识别的速度,本文采用复用 IP 技术来部分实现语音识别算法解决速度的问题。

本文的实验结果表明:通过嵌入式语音识别技术实现 AmI 语音交互能够为 AmI 环境提供友好的主动服务。AmI 应用场景如图 1 所示。

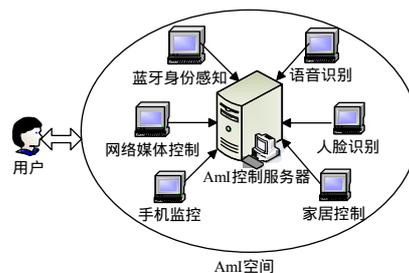


图 1 应用场景

2 技术准备

语音识别系统是一种模式识别系统,按照说话人的限定分为特定人识别和非特定人识别。基于传统DTW算法的语音识别系统由于算法自身特点,对特定人语音识别有较好的识别性能,但在使用前需要对所有词条进行训练,词表一旦发生变化则需要用户重新训练模板,使用非常不方便^[3]。HMM是一种基于统计信号的模型,算法具有良好的识别性能和抗

基金项目: 计算机科学重点实验室开放课题基金资助项目“基于任务的用户界面软件体系结构建模研究”(SYSKF0704)

作者简介: 龚晓庆(1974-),女,讲师、博士,主研方向:人机交互,软件工程;柯素娟,硕士研究生;李 辉,博士;华庆一,教授

收稿日期: 2008-08-13 **E-mail:** gxq@nwu.edu.cn

噪性能，可用于非特定人识别且不需要用户事先训练，被应用到许多优秀的语音识别系统中。

2.1 隐马尔可夫模型

隐马尔可夫模型通过对大量语音数据进行统计，建立要识别词条的统计模型。图 2 表示的是一个从左到右的 HMM 模型图。

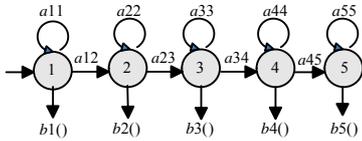


图 2 从左至右的 HMM 模型

从图中可以看出一个具有 N 状态的 HMM 模型可以由一个三元组参数 $\lambda = (A, B, \pi)$ 来定义^[4]。其中， $A: \{a_{ij}\} = P(q_j | q_i)$ 为状态转移概率矩阵，描述了从状态 q_i 转移到状态 q_j 的概率； $B: b_j(k) = P\{V_k | q_j\}$ 描述了状态 q_j 下得到符号 V_k 的概率， $1 \leq j \leq N, 1 \leq k \leq M$ ； π 为初始概率矢量 $\pi(i), 1 \leq i \leq N$ ； $Q = \{q_1, q_2, \dots, q_N\}$ 为时间 $[1, t]$ 内经历 HMM 模型的状态序列； N 为状态数； M 为编码符号集中符号的总数。

2.2 语音识别系统原理

语音识别系统包括前端处理和后端处理 2 个部分^[5]。前端部分完成语音信号处理工作，后端部分完成语音模板的训练和识别，前端处理部分的输出是后端处理部分的输入。

2.2.1 系统前端处理

前端处理部分包括分帧、预加重、加窗、端点检测、特征提取、矢量量化等操作。本系统帧长取 20 ms，帧移为 1/2 帧。采用式(1)对语音信号进行加重：

$$data(n) = S(n) - \alpha S(n-1) \quad \alpha = 0.95 \quad (1)$$

为了消除分帧产生的边界效应，对每一帧做加窗处理，采用汉明窗，定义为下式(L 为帧长)：

$$w(n) = \begin{cases} 0.54 - 0.46 \cos(\frac{2\pi n}{L-1}) & 0 \leq n < L \\ 0 & \text{其他} \end{cases} \quad (2)$$

端点检测是找出语音信号的起点和终点的过程，本系统采用对短时平均幅度和短时平均过零率双门限的方法来进行端点检测，能准确地检测到语音的起点和终点，保证了系统的识别率。

特征提取是要从语音波形中提取出反映语音特征的信息。考虑到嵌入式系统资源有限的特点，在保证系统识别率的同时需要尽可能地减少系统的运算量，本系统中采用 LPCC 算法进行特征提取，提取得到 16 阶 LPCC 系数。

矢量量化是为了减少存储量、减轻系统负荷对语音数据进行压缩的技术。本系统采用 LBG 算法来生成码本，初始码本的选择采用随机法。为了平衡码本失真度与码本存储量、搜索计算量的矛盾，本系统采取的码本容量为 128。

2.2.2 系统后端处理

后端处理部分进行模型训练和模型识别。为解决 HMM 模型的 3 个基本问题，HMM 模型提供了 3 个基本算法^[4]。前向后向算法能有效计算出一个序列的概率，Baum-Welch 算法计算出在给定模型下一个观察序列的最大概率，完成模型的训练过程，Viterbi 算法计算得到一个最优状态序列的概率，解决了 HMM 模型的识别问题，3 个算法结合使用实现语音识别过程。在本系统中，为了保证系统的实时性，部分算法采用硬件 IP 核来实现。

3 基于 FPGA 的语音识别系统框架

FPGA 是软硬件协同开发平台，系统的实现包括硬件和软件 2 大部分。下文将具体介绍语音识别系统的硬件平台框架和软件平台框架。

3.1 系统硬件框架

系统的硬件框架如图 3 所示。

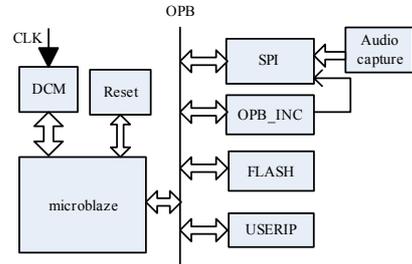


图 3 系统的硬件框架

在系统的硬件框架中，Microblaze 是处理器；OPB 是片上总线，提供分离的 32 位地址总线和 32 位数据总线，用来访问低速和低性能的系统资源；SPI 控制器实现与语音采集模块通信；中断控制器 OPB_INC 作为一个 GPIO 引入系统实现 SPI 的中断服务功能；FLASH 用来存储语音识别系统的启动程序，并存储训练得到的模板；USERIP 是自定义的 IP 核，实现语音识别系统算法。

3.2 系统软件框架

语音实时识别程序包括训练和识别 2 个大部分，对采集到的系统中的语音数据，系统需要判断是进行训练还是识别操作，从而作不同的处理。流程如图 4 所示。

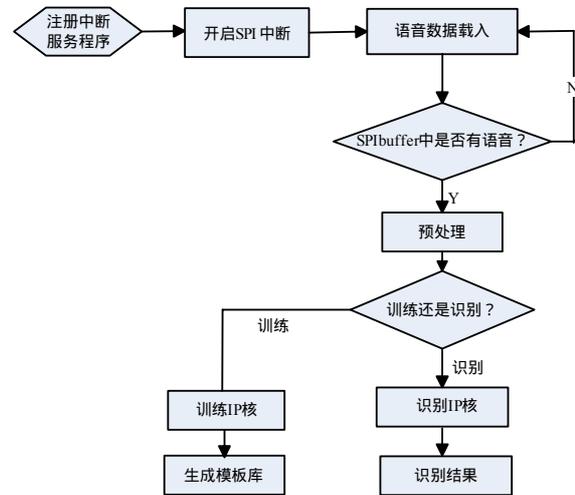


图 4 系统软件流程

各部分功能如下：

- (1)注册中断服务程序：初始化 SPI 中断处理器的初始接口，并把 SPI 的中断挂在总线中断控制器 OPB_INC 上。
- (2)开启 SPI 中断：开启 OPB_INC 中断，开启 SPI 中断。
- (3)语音数据载入：载入语音采集模块采集到的数据到系统中。
- (4)预处理：对输入的语音信号进行处理，并把语音信号的特征数据输入到 HMM 算法的 IP 核中进行计算。
- (5)训练 IP 核：进行训练模块的计算。
- (6)生成模板库：根据训练计算结果生成模板数据。
- (7)识别 IP 核：进行模板匹配识别计算。
- (8)识别结果：得到相应的识别结果。

