嵌入式语音识别系统特征参数提取研究

朱 宇,宋 艳
(西安科技大学 计算机学院,陕西 西安 710054)

摘 要:在噪声环境下能准确有效地提取语音信息是语音识别的重点难点,将其应用于嵌入式系统中,有一定的研究意义。通过比较分析传统的语音特征参数提取的方法:线性预测倒谱系数,Mel频率倒谱系数,提出了一种新的方法,采用Mel频率倒谱系数与一阶差分Mel频率倒谱系数(MFCC+△MFCC)相结合的方法提取语音特征参数,结合双门限检测法进行端点检测和HMM模型进行模型匹配,并进行了以ARMS3C2410为核心硬件与软件的系统设计。该方法较传统方法提高了系统的鲁棒性、识别的准确率和系统效率,适用于噪声环境下的语音识别。

关键词:语音识别;线性预测倒谱系数;Mel频率倒谱系数;ARMS3C2410
中图分类号:TP31 **文献标识码**:A **文章编号**:1673-629X(2011)06-0246-04

Research of Characteristic Parameters Extraction Based on Embedded Speech Recognition System

ZHU Yu, SONG Yan
(School of Computer Science & Technology, Xi'an University of Science & Technology, Xi'an 710054, China)

Abstract: The key points and difficulties of speech recognition is the technology of extracting the voice information accurately and efficiently in noisy environment. Applying this technology in embedded systems has some research significance. Through the comparative analysis of the traditional phonetics characteristic parameters extraction methods which are linear forecast cepstrum coefficients and Mel frequency cepstrum coefficients, proposed a new method in which through combining the method of Mel frequency cepstrum coefficients with first-order differential Mel frequency cepstrum coefficients (MFCC + △MFCC) to extract phonetic features parameters firstly, then take advantage of double threshold method to detect endpoint and use HMM model to do model matching, and the system design is carried on the ARMS3C2410 as the core of hardware and software. The experiments show that the system robustness, identification accuracy and efficiency of the system in this method got improved compared with the traditional method, and this method is suitable for noise environment of speech recognition.

Key words: speech recognition; linear predictive cepstral coefficient; Mel frequency cepstral coefficient; ARMS3C2410

0 引 言

近十年以来语音识别技术得到了很大的发展,语音识别技术在便携式系统中得到广泛的应用。语音特征参数提取是语音识别中很重要的模块,在嵌入式语音识别系统中占有非常重要的地位。语音识别根据实际需要和应用场合的不同,可以分为孤立词识别和连续语音识别、特定人识别和非特定人识别[1]。目前常用的语音识别方法有基于特定人的动态时间规正法(Dynamic Time Warping ,DTW)、基于统计模型的隐马尔可夫模型法[2](Hidden Markov Model ,HMM)、基于小波变换以及神经网络的识别法(DNN,NPN ,TDNN)[3]。隐马尔可夫模型法对非特定人连续语音有很高的识别率,目前一般都采用基于隐马尔可夫模型识别方法为基本算法,采用模式匹配的原理来实现语音识能。文中对以上几种方法的语音特征提取方法进行了比较分析,提出了MFCC+△MFCC方法来提取语音特征参数,结合双门限检测方法进行端点检测和HMM模型匹配,通过实验表明,将该方法应用到嵌入式系统中,提高了系统的鲁棒性、准确率及效率,有一定的应用研究意义。

收稿日期:2010-11-25;**修回日期**:2011-03-04
基金项目:陕西省教育专项科研基金(陕教资2008-147号)
作者简介:朱 宇(1955-),男,副教授,研究方向为嵌入式系统、计算机监测与控制。

1 语音特征提取方法

1.1 特征参数提取方法概述

语音特征提取的本质是在降低或很少降低语音分类结果性能的情况下来降低特征空间的维数[4,5]。语音识别系统的特征提取是适合于语音分类的信息特征,这些信息特征要能有效准确地区分不同的语音特

征模式,而且对相同模式的变化具有相对的稳定性。当前常用的语音特征提取方法是进行语音特征参数的提取。语音参数的选择是整个识别系统的基础,对正确的识别率有着直接影响[5]。语音特征一般包括基音周期(Pitch)、主分量分析(PCA)、独立分量分析(ICA)、线性预测系数(LPC)、美尔频率倒谱系数(MFCC[6-8])、线性预测倒谱系数(LPCC[8,9])以及线谱对系数(LSP)等等。MFCC 和 LPCC 在实际应用中最为成熟,特别是在一些真实信道噪声和频谱失真的情况下,能更好地反映人耳的听觉感知情况,因此在特征参数提取中应用的更多。

美尔频率倒谱系数(Mel Frequency Cepstral Coefficient,简称 MFCC)特征是当前语音特征提取过程中使用最广泛的语音特征之一,它是频谱上采用滤波器组的方法计算出来的,是一种能够比较充分利用人耳动态感知特性的一种特征提取参数。经过比较分析表明,MFCC 参数性能优于 LPCC 参数。

与 LPCC 相比,MFCC 优点有:

(1)语音信息大多数集中在信号低频部分,而信号高频部分容易受外界环境噪声的干扰,MFCC 将线性频标转化为 Mel 频标,它强调语音的低频信息,从而突出有利于识别的语音信息,并且屏蔽了噪声的干扰。

(2) MFCC 不依赖全极点的语音产生的模型假定,它考虑了人耳的听觉感知特性,抗噪声以及抗频谱的失真能力较强,从而提高识别系统的性能。MFCC 抗噪声能力也优于 LPCC。此外,MFCC 在各种情况下都可以使用。而 LPCC 假定所处理的信息为自回归(AR)信号,对于动态特性比较强的辅音,这个假设不严格成立。

由于传统的 MFCC 一般只反映语音特征参数的静态特性,而人耳对语音动态的特征却更为敏感,语音特征提取参数中一阶差分(△MFCC)是一种动态参数,因此提出一种 MFCC+△MFCC 相结合的方法提取语音特征参数,将此方法应用于嵌入式语音识别系统中,有较好的鲁棒性。

1.2　特征提取过程

人的听觉系统是一种特殊的非线性系统[5],基本上是一个对数的关系,它响应不同的频率信号的灵敏度是不相同的。美尔频率倒谱系数(MFCC)能充分利用人耳的这种特殊的感知特性。线性频率与 MFCC 的转换关系如图 1 所示 。

$$f_{mel} = 2595 \log_{10}(1 + \frac{f}{700}) \qquad (1)$$

MFCC 特征参数是按帧计算的,它的提取过程用

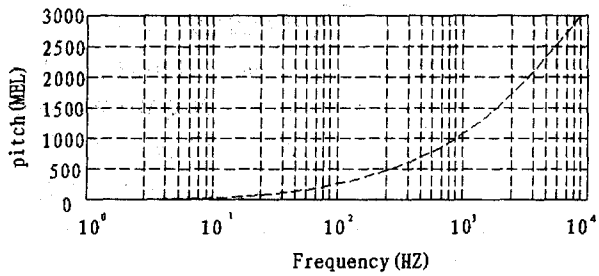

图 1　Mel 标度与频率的关系

框图表示(见图 2)。由于不同说话人声道的特征具有不同的特性,所以一般在实际的信号分析中采用预加重技术,也就是在对信号取样的过程中,加入一个一阶高通滤波器,以便于加强声道部分的语音特征,这样有助于对声道参数进行分析;由于人耳的听觉特性与三角滤波器组相似的特性,Mel 频率滤波则利用其相似性对语音信号的幅度平方谱进行平滑;对不同频段的频谱成份进行解的相关处理,使得各向量之间能相互独立,一般经过离散余弦变换(DCT)即可。

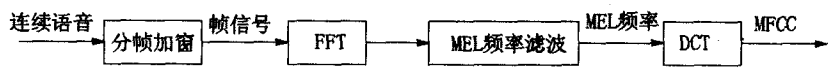

图 2　MFCC 计算流程图

(1)语音分帧。分帧可以用连续方法,也可以用交叠的方法。为使语音帧与帧之间能平滑过渡,保持应有的连贯性,根据语音信号特征的短时平稳特性,采用交叠分段的方法将语音信号分成若干帧进行处理,每帧长度约为 10ms ~ 30ms,即每语音帧的帧尾与下一帧的帧头重叠[10]。

(2)语音加窗。用有限长度的可移动的窗口进行加权的方法,也就是选用一定的窗函数来乘以语音信号,便可形成加窗的语音信号。语音加窗主要是为了减小语音帧的截断效应,减小帧的起止点处的不连续,降低帧的两端坡度,使其平滑过渡。语音分帧窗函数 $w(n)$ 的选择很重要,为使其能更好地反映语音信号的变化特性,一般在预处理中,大多数选用合适的汉明窗(Hamming)来进行语音分帧的处理。设窗函数为 $w(n)$,语音帧信号为 $x(n)$,加窗后的信号为 $y(n)$,则:

$$y(n) = x(n) \ *3 \ w(n) \qquad (2)$$

其中,$0 \leqslant n \leqslant N - 1$;N 为每帧取样的点数,取值为 410 点。

汉明(Hamming) 窗 $w(n)$ 表达式如下:

$$w(n) = 0.54 - 0.46\cos\frac{2\pi n}{N-1} \quad 0 \leqslant n \leqslant N - 1 \qquad (3)$$

(3)快速傅里叶变换(FFT)。采用快速傅里叶变换是为了高效地使语音帧从时域变换到频域。

(4)Mel 频率滤波。快速傅里叶变换后得到一组

离散频谱，采用三角形滤波器对离散频谱进行滤波处理，得到一组系数，$m_1, m_2, \cdots, m_i$。其中滤波器组中每个三角形滤波器在 Mel 频率轴上以及 Mel 频率标度上都呈等间隔分布。p 为滤波器组的个数，由语音信号的截止频率决定，本系统中 p 取值为 24，所有的滤波器总体上覆盖是从 0Hz 到 Nyquist 频率，即采样率的二分之一。m_i 的公式计算如下：

$$m_i = \sum_{k=0}^{N-1} \ln(|X(k)| \cdot Hi(k)), \quad i = 1, 2, \cdots, p \tag{4}$$

$$\begin{cases} 0, & k < f[i-1] \text{ 或 } k > f[i+1] \\ \dfrac{2(k - f[i-1])}{(f[i+1] - f[i-1])(f[i] - f[i-1])}, & f[i-1] \leqslant k \leqslant f[i] \\ \dfrac{2(f[i+1] - k)}{(f[i+1] - f[i-1])(f[i+1] - f[i])}, & f[i] \leqslant k \leqslant f[i+1] \end{cases} \tag{5}$$

公式中 $f[i]$ 为三角形滤波器的中心频率，满足：

$$\text{Mel}(f[i+1]) - \text{Mel}(f[i]) = \text{Mel}(f[i]) - \text{Mel}(f[i-1]) \tag{6}$$

(5) 离散余弦变换(DCT)。把从上一步获得的 Mel 频谱变换到时域后，结果就是 MFCC。由于 MFCC 都是实数，可以使用 DCT 将其变换到时域。MFCC 的计算公式如下：

$$C_{\text{Mel}}(i) = \sqrt{\frac{2}{p}} \sum_{j=1}^{p} m_j [\cos(j - 0.5)] \frac{\pi i}{P} \tag{7}$$

(6) 一阶差分 MFCC(△MFCC)。假设当前所获得的特征参数是 P 维，差分参数的计算采用公式(8)：

$$d\text{CeP}_i(n) = \frac{1}{\sqrt{\sum_{i=-k}^{k} i^2}} \sum_{i=1}^{k} k(\text{CeP}_{i-k}(n) - \text{CeP}_{i+k}(n))$$

$$n = 1, 2, \cdots, P \tag{8}$$

其中 dCeP 表示动态特征，CeP 表示倒谱，K 是求差分的帧的范围，k 为常数，通常为 2，由公式(7)计算得到的差分参数为一阶差分△MFCC。

2 实验设计

2.1 硬件设计

本系统主要以 SAMSUNG (三星) 公司的 ARMS3C2410 芯片为核心，外围是麦克风(MIC)输入模块电路、功放和喇叭输出模块电路、LED 显示电路和通信模块电路。电源由外部 5V 电源供电。S3C2410 处理器是 SAMSUNG 公司基于 ARM 公司的 ARM920T 处理器核，采用 0.18μm 制造工艺 32 位微控制器。该处理器拥有独立的 16kB 指令 Cache 和 16kB 数据 Cache，集成了 ROM、RAM、A/D，和多个外设，如通用 I/O 口，定时器、串行口，MMU，支持 TFT 的 LCD 控制器，NAND 闪存控制器等。A/D 转换器采用了 PHILIPS 公司的音频数字信号编译码器 UDA1341TS，它可以将模拟信号与数字信号进行相互转换。输入输出部分由话筒、扬声器、LCD 液晶显示器、键盘和电路组成。系统硬件框图如图 3 所示。

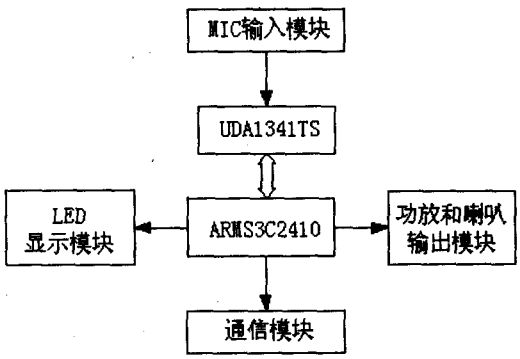

图 3 系统硬件框图

MIC 输入模块、喇叭输出模块电路完成语音的录入、提示语音以及识别结果的输出等功能。ARMS3C2410 存储所有的程序及有关语音提示。LED 显示模块作为辅助识别结果的判断工具，如果识别结果正确则 LED 亮。通信模块将过零率计算、噪声能量、模型库的训练等大量语音数据处理都上传给 PC，由 PC 完成。

2.2 软件设计

系统的软件设计平台是基于 Linux 系统平台的设计，分为控制软件设计和算法设计。由于 Linux 系统的源代码开放，所以已广泛应用于嵌入式的众多领域。控制软件设计是系统整体软件的设计。系统中的算法采用隐马尔可夫算法(HMM)来实现语音模型的训练及识别。

本系统的核心是软件控制系统的模块化设计。流程图如图 4 所示。主要包括语音识别的系统初始化、语音训练、语音识别。首先，进行语音训练，从训练语音中提取出随时间变化的语音特征序列、建立语音参考模型并存储；然后进行模式匹配，再次输入语音，从待识别语音中提取语音特征，将同一语音特征与已训练好并存储的语音参考模型进行匹配和比较，如果识别不成功则提示错误的原因，之后再次进行匹配和比较，直至识别最佳语音结果，成功后即可执行用户的命令。

隐马尔可夫模型(HMM)是一种强有力的概率机器学习过程，已被成功应用于语音识别[10]。它是一个二重的严格的马尔可夫随机过程[11]。本系统遍历模型 HMM 的状态数取 4。HMM 为有限状态的随机过程[9]，从状态 s_i 到状态 s_j 的转移概率为 $a_{ij} = p(s_i | s_j)$；对应状态 s_i，语音特征 x(即随机向量)的概率密度函数为 $p(x | s_i)$。由给定模型 M 产生具有 N 帧语音特征

向量 x_1, x_2, …, x_N 的似然值为:

$$p(x_1,x_2,\cdots,x_N \mid M) = \sum_{s_1,s_2,\cdots,s_N \in s,} \prod_{i=1}^{N} p(x_i \mid s_i)p(x_i \mid s_{i-1}) \quad (9)$$

其中,P 为相对应的状态的概率,$S = \{1,2,3,4\}$ 表示状态的集合。

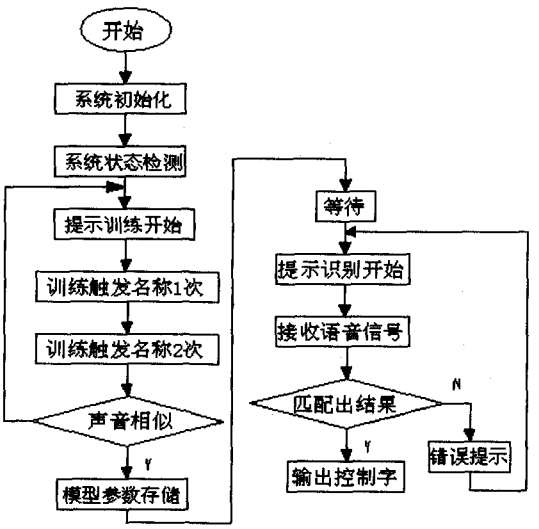

图 4　系统程序流程图

3　实验分析

文中在实验室条件下,使用麦克风单声道来录制采集 10 人说话的语音数据。采样率为 16kHz,每个说话人中 10 个语音段成为训练样本集,其中 6 个语音段作为训练测试集。对语音信号进行特征提取时,选取语音帧长为 410 个采样点,帧移为 160 个采样点,并且对语音帧进行预加重和加汉明窗处理,预加重系数为 0.97。语音信号经过语音训练、HMM 模型匹配识别后,分别提取 LPCC、MFCC、MFCC +△MFCC 的特征参数,由图 5 可见,该 10 人语音识别的识别率为:LPCC 为 88.52%、MFCC 为 91.56%、MFCC +△MFCC 为 92.54%。比较得知识别率以 MFCC +△MFCC 特征提取为最高。由此可知,MFCC +△MFCC 相结合的方法能有效地适用于语音特征参数的提取。

4　结束语

文中介绍了一种基于动态特征参数的 MFCC+△MFCC 相结合的语音特征提取方法,并将此方法应用于嵌入式语音识别系统,与传统的 MFCC 方法比较,

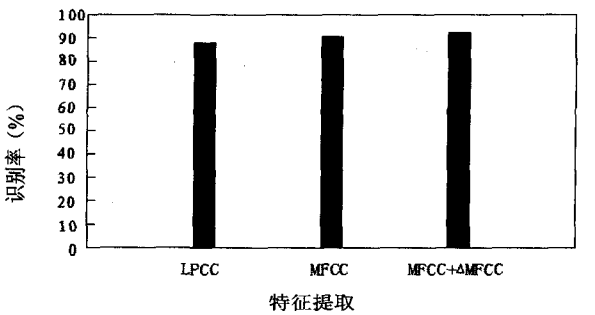

图 5　实验结果分析图

识别性能有较大提高,此方法在语音识别领域中有很好的应用前景。系统硬件结构上采用以 ARMS3C2410 为核心的处理器,加快了系统的处理速度[12],相对提高了语音识别的实时性。

参考文献:

[1] 徐　敏,邹　莹,魏洪兴.一种嵌入式语音识别控制模块的设计与实现[J].厦门理工学院学报,2008,16(4):44-47.

[2] 王坤卿.HMM 模型在语音识别研究中的应用[J].电脑知识与技术,2008,14(7):1966-1968.

[3] 杨　毅,杨　宇,余达太.语音增强及其消噪能力研究[J].微电子学与计算机,2006,23(7):203-208.

[4] 魏　星,周　萍.语音识别系统及其特征参数的提取研究[J].计算机与现代化,2009(9):228-243.

[5] 李弼程,邵美珍,黄　洁.模式识别原理与应用[M].西安:西安电子科技大学出版社,2008:72-98.

[6] Fakhr W,Salam A A,Hamdy N. Enhancement of mismatched conditions in speaker recognition for multimedia applications [C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. [s. l.]:[s. n.],2004.

[7] Rabiner L,Juang Biing-Hwang. Fundamentals of Speech Recognition [M]. [s. l.]:Prentice Hall, 1992.

[8] 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.

[9] 张军英.说话人识别的现代方法与技术[M].西安:西北大学出版社,1994.

[10] Rabiner L E. A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition[J]. Proceedings of The IEEE,1989,77(2):257-286.

[11] 韩　普,姜　杰.HMM 在自然语言处理领域中的应用研究[J].计算机技术与发展,2010,20(2):246-252.

[12] 周立功.ARM 嵌入式系统基础教程[M].北京:北京航天航空大学出版社,2005.

(上接第 174 页)

[9] Richard D,Edward J,Timothy R. Adding Attributes to Role-Based Access Control[J]. IEEE Computer Society,2010(6):79-81.

[10] 陆庭辉,文贵华.B/S 结构下的用户访问控制方法[J].计算机工程与设计,2010,31(7):1433-1436.

[11] 杨　云,刘　军.Web 安全设计之道[M].北京:人民邮电出版社,2009.

[12] 李天平.NET 深入体验与实战精要[M].北京:电子工业出版社,2009.