

基于 ARM-Linux 的普通话发音质量评价体系^①

陈彩华¹, 龙卫兵²

¹(湖南三一工业职业技术学院, 长沙 410129)

²(湖南工程职业技术学院, 长沙 410114)

摘要: 从当前普通话测试的现状与需求出发, 对基于 ARM-Linux 的语音评价系统进行了深入研究, 提出了基于嵌入式技术的普通话发音质量评价方案。其中, 系统硬件以 S3C2410X 处理器和 UDA134TS 为主要部分, 软件则主要研究基于语音特征比较的普通话评价方法。本设计方案能有效实现普通话测试系统从 PC 平台到嵌入式平台的移植。

关键词: 普通话发音质量评价; S3C2410X; Linux; UDA1341TS

Chinese Pronunciation Quality Evaluation System Based on ARM and Linux

CHEN Cai-Hua¹, LONG Wei-Bing²

¹(Hunan SANY Polytechnic College, Changsha 410129, China)

²(Hunan Engineering Polytechnic, Changsha 410114, China)

Abstract: In view of the current situation and demands of the Chinese test, this paper made a thorough study of the pronunciation evaluation system based on the ARM-Linux, and put forward a design of Chinese pronunciation quality evaluation system based on the embedded technology. S3C2410X and UDA1341TS are used as the main part of the system. Through the research of speech characteristic comparison, the design easily realized the transplant of Chinese test system from PC to embedded system.

Key words: Chinese pronunciation quality evaluation; S3C2410X; Linux; UDA1341TS

1 引言

随着各种消费电子产品对低成本、高稳定性片上语音识别系统需求的增加, 语音识别系统已从实验室的 PC 平台向各种嵌入式设备转移。在普通话日益推广的新形式下, 计算机辅助普通话培训、测试工作也取得了跨越式发展, 但目前所推行的普通话水平智能测试系统仍局限于 PC 平台, 而且测试成本高。针对当前计算机辅助测试系统降低成本的需要, 本文旨在设计一个基于 ARM9 的嵌入式普通话评测系统, 用于普通话的发音质量评分, 该系统以 ARM 芯片 S3C2410X 和嵌入式操作系统 Linux 为基础, 配合其他功能模块, 能够独立完成普通话发音质量的评分工作。系统采用 S3C2410X 作为控制芯片与处理芯片, 设计了基于语音特征比较的语音评分算法, 使得该嵌入式

系统的语音识别更加方便快捷, 并具有一定的通用性。

2 普通话发音质量评价原理

目前, 国内的普通话水平测试方式主要有人工测试与计算机辅助测试两种, 分别采用主观评价与客观评价两种发音质量评价体系^[1], 其评价原理如图 1 所示。

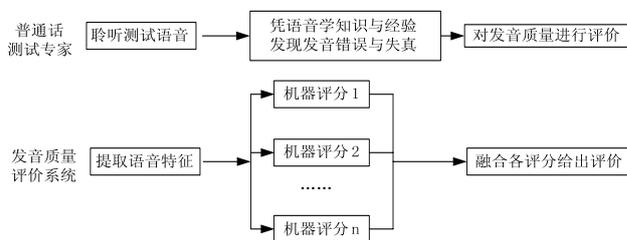


图 1 普通话发音质量的主/客观评价过程

① 基金项目: 国家语委普通话培训测试科研规划课题(PY09003)

收稿时间: 2010-12-20; 收到修改稿时间: 2011-04-01

主观评价主要以人为主体来评价普通话发音质量，首先由普通话测试专家聆听测试语音，然后专家们凭自己的语音测试经验，检测测试语音与预先约定的标准发音之间的差异，最后根据被测试语音的失真程度给出一个总体评价。主观评价方法虽然能真实反应普通话的发音质量，符合人对语音质量的整体感觉，但工作量大，不易重复，加上语音评价不但与语音学、语言学、信号处理等学科相关，而且还与心理学、生理学等学科有密切关系^[2]，因而主观评价方法的评价结果与测试专家有很大的关系，这进一步凸显了主观评价的主观性缺陷。

客观评价系统则采用机器来对发音质量进行自动评价，系统首先提取测试语音与参考语音各自的特征值，然后比较二者在各个特征值方面的差异，根据差异值给出各方面的机器评分，最后对各种机器评分值进行融合，最终得出所测试的语音分数值。相对主观评价系统而言，客观评价系统由于能提供统一的评价标准，因而能有效去除主观评价系统中的主观性，同时还能极大的提高语音评价的效率。

本文研究的具有参考语音标准的普通话发音质量评价系统主要基于语音识别与语音韵律分析，采用梅尔倒频谱参数表征声音的内容，基频轨迹用来反映声音的韵律变化，通过对比测试语音和参考语音之间的内容与韵律差异，得出反映内容与韵律的机器评分，两者融合后便得到了最终的客观分数。普通话语音质量评价原理如图 2 所示。

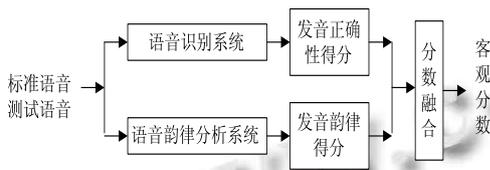


图 2 普通话发音质量评价原理

3 普通话发音质量评价体系硬件设计

本文研究的普通话发音质量评价系统主要用于便携式语言学习与测试系统，因此系统硬件主要采用低成本、低功耗、高性能的 32 位 ARM 微处理器 S3C2410X 作为系统主控制器，整个系统的硬件结构如图 3 所示。

S3C2410X 微处理器使用 ARM920T 核，工作频率为 203MHz，带 MMU 先进体系结构，支持 Linux，WinCE，EPOC32 等嵌入式操作系统，集成了外部存储控制器、LCD 控制器、PS2 接口，4 通道的 UART、

2 个 USB 主机接口、117 个通用 I/O、24 个外部中断、芯片内置 PLL^[3]。系统的外围电路主要包括存储器、数据采集模块、LDC 显示模块等。存储器包括 SDRAM 与 FLASH ROM，SDRAM 主要作为应用程序和数据的运行空间，Flash ROM 则用来存储 BIOS、嵌入式操作系统及运算所需要的各种数据。

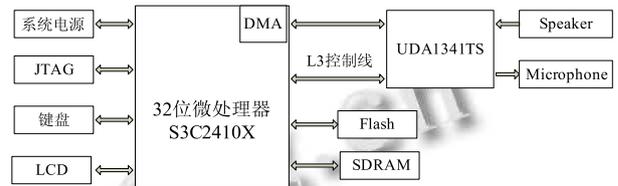


图 3 系统硬件结构

整个系统采用 ARM-Linux 架构，系统的核心部分由 S3C2410X 和 UDA1341TS 构成。由于 S3C2410X 内置 IIS 总线接口，加上 UDA1341TS 支持 IIS 总线数据格式，因此 S3C2410X 的 IIS 总线信号与 UDA1341TS 的 IIS 信号可直接相连，通过 DMA 传输模式，可完成语音的录音与放音。语音信号通过 UDA1341TS 放大、滤波、A/D 转换后，由 DMA 控制器从 FIFO 缓冲区送入 DMA 缓冲区，供 S3C2410X 微处理器处理。

4 普通话发音质量评价体系软件设计

操作系统是整个普通话发音质量评价体系的核心，本文选取开源的、功能可裁减的 Linux 作为嵌入式操作系统。整个系统能够完成语音数据的预处理、语音特征提取、发音质量的专家模拟评价等功能。

4.1 语音数据预处理

语音数据预处理的主要任务是为语音特征提取准备可计算的数字语音信号，其处理流程如图 4 所示。首先，系统根据预先确定的采样频率，采集由开发板上的麦克风传入的语音信号，然后通过 A/D 转换电路将采集的模拟语音信号转换成数字信号；接着，系统应用端点检测技术来确定语音中的音素、音节以及词等信息在语音中的起始点与终止点，以进一步排除原语音信号中的无声段，获取有效的语音段；然后，系统将语音信号通过一个低通滤波器进行高频噪声滤波处理；为了模拟人耳的高频提升特性，最后系统还需要将语音信息作预强调处理。预处理好的语音信息可在需要时直接读入使用，这在很大程度上降低了系统的运算量，同时也提高了系统的实时性。



图 4 语音数据预处理流程

4.2 语音基频轨迹曲线提取

普通话具有包含轻声在内的五种声调，是一种典型的有调语音，要对普通话的发音质量作出正确评价，就必须对普通话的声调发音准确程度作出判断。在语音信息中，基频参数是反映语音的语调韵律的一个重要参数，在现有的基于语音波形估计基音周期的方法中，平均幅度差法（AMDF）的特点是不涉及到乘法和除法，比较适合于资源相对匮乏的嵌入式系统平台，因此本文主要采用 AMDF 法来求取语音的基频参数，语音基频轨迹曲线求取的流程如图 5 所示。

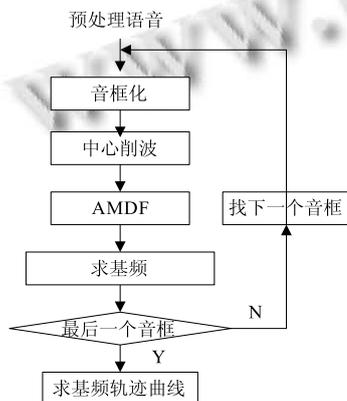


图 5 基频轨迹曲线提取流程

考虑到语音信号的时变特性，系统首先对经过预处理的数字语音信号进行音框化处理，音框大小取为 256 点（32ms），同时让相邻两音框之间重叠 128 点，以防止音框之间的剧烈变动。在系统对音框化后的语音信号进行基音周期估计之前，必须先对语音信号进行中心削波非线性处理，以消除共振峰的影响，在中心削波后，系统只需要对语音信号进行 AMDF 处理，即可求取所处理帧的基频，在求取各语音帧的基频后，就可得到整个语音信号的基频轨迹曲线。

3.3 梅尔倒频谱参数提取

语音特征参数的选择和提取是构建系统的关键，所选取的特征参数不仅要求计算方便，而且要求有高效的计算方法，以确保语音识别的实时实现。目前常用的语音识别参数主要有 3 种^[4]：线性预测倒谱系数（LPCC）、Mel 频标倒谱系统（MFCC）和感知线性预

测系数（PLPC）。在三种语音参数中，梅尔倒频谱参数（MFCC）是利用 Mel 频率与实际频率之间的非线性关系计算得到的频谱特征，能很好的模拟人耳的特殊感知特性，而且识别性能好，同时还可以方便的在 16 位定点 DSP 上实时实现，因此本设计主要选用 39 维 MFCC 特征参数来表征语音内容的正确性，MFCC 特征参数的提取流程如图 6 所示。系统在读入经过预处理的数字语音之后，先进行取音框操作，然后对各音框乘上汉明窗，以补偿音框边缘所造成的信号不连续现象，接着对各音框做快速傅利叶变换，求取各音框的频谱，在提取 MFCC 特征的同时，借助一组在 Mel 频标上均匀分布的 20 个三角滤波器即可求出各频带的输出对数频谱，可进一步减少系统的运算量。

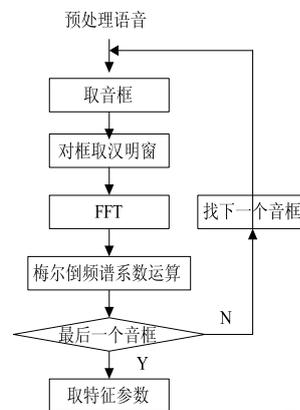


图 6 梅尔倒频谱参数提取流程

3.4 发音质量评价算法

为了使系统能够在存储资源、运行速度都受到限制的嵌入式平台上更好的模拟语言专家的评分，本文设计了一款改进的基于特征比较的普通话发音质量评价算法，只需要一次动态时间规整（DTW）操作就可结合两种特征参数进行模式对比。

首先系统利用 MFCC 参数对待评分语音与标准语音进行 DTW 非线性校正，既保证了两段语音内容相似位置的对应性，同时还可以获得一条误差最小的校正路径以及反映两段语音内容相似度的 DTW 距离；然后系统以校正路径为基础，比较待评分语音基频变化轨迹与标准语音的基频变化轨迹，通过计算两者相似内容对应位置的基频点差距与基频点变化量差距，来获取两段语音在语调上的相似度。

对给定语音的各语音帧分别进行基频轨迹提取与

（下转第 94 页）

常操作的用户。以上方法只是检测异常用户，即用户在单一会话中的异常行为检测，对于那些多会话的异常行为是无能为力的。而大多数系统使用者本身的操作行为，必然存在很多有趣的行为模式，如熟练用户与不熟练用户的操作模式一定是不同的，经常偷懒的使用者和积极的工作者在实用系统中也必然存在巨大的差异，什么样的使用者会使系统出现更多的异常。下一阶段将在现有的模式上建立系统使用者的行为模型，当挖掘出用户的使用模式后，可以使用 svm 等方法对以上描述类用户的使用模式建立一个模型，然后使用这个模型区分出不同类型的用户，准确地预测非法用户的入侵。

表 4 测试结果

数据集	异常类型	异常事件 (个)	检测异常事件 (个)	检测率	误报率
抽取 6386 条样本, 异常 90 条, 误报 158 条	DOS	18	15	81.3953%	158/(6386-90)=25.138
	Intrusion Attempt	20	16	81.25%	
	Penetration	45	31	68.5185%	
	R2L	7	5	70.5882%	

(上接第 97 页)

梅尔倒频谱参数提取，即可获得对应语音的基频特征向量与 MFCC 特征向量，设标准语音的 MFCC 特征向量为 $M_1=[m_1(1), m_2(2), \dots, m_1(T)]$ ，基频特征向量为 $P_1=[p_1(1), p_2(2), \dots, p_1(T)]$ (T 为标准语音长度)；待评价语音的 MFCC 特征向量为 $M_2=[m_2(1), m_2(2), \dots, m_2(S)]$ ，基频特征向量为 $P_2=[p_2(1), p_2(2), \dots, p_2(S)]$ (S 为待评价语音长度)，系统只需进行一次 DTW 操作，就可按以下公式求取基频变化相似度 P 以及 MFCC 特征相似度 M 。

$$C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = DTW(M_1, M_2), (C \text{ 是特征比较矩阵})$$

$$\begin{pmatrix} P \\ M \end{pmatrix} = \begin{pmatrix} P_1 & P_2 \\ M_1 & M_2 \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$$

由于各语音特征参数间存在着关联性，可根据评分的侧重点不同，在机评分计算公式中引入各特征参数的权值，实现机评分与专家评分之间的最佳映射。

$$Scores(P, M) = k_1P + k_2M + k_3PM$$

参考文献

- Hawkins D. Identification of Outliers. London: Chapman and Hall, 1980.
- 柴平璋,程时端.入侵检测技术分析.计算机工程与应用, 2003,14:164-166.
- 肖国强,肖铁.一种从 WEB 日志中挖掘访问模式的新算法.华中科技大学学报(自然科学版),2004,32(5):70-72.
- H Z, Xu X, Deng S. FP-outlier: frequent pattern based outlier detection. ComSIS, 2005,2(1):103-118.
- Han J, Kamber M.范明,孟小峰,译.数据挖掘概念与技术.北京:机械工业出版社,2001.3-22.
- Knorr E, Roymond NG. Algorithms for mining distance-based outlier in large databases. Proc. of the VLDB Conf. New York: 1998: 390-405.
- 崔贵勋.基于密度的离群数据挖掘算法研究[硕士学位论文].重庆:重庆大学,2007.
- 徐翔,刘建伟,罗雄.离群点挖掘研究.计算机应用研究,2009, 26(1):34-40.

4 结语

本文在功能强大的 ARMS3C2410X 硬件平台上，构建了一个嵌入式平台的普通话发音质量评价系统，分别从语音的准确性与朗读的韵律两个方面对普通话的发音质量进行评价，在保证评分质量的情况下，对基于特征比较的语音评价算法进行了改进，大大降低了系统实现平台的硬件资源配置要求，对研制嵌入式语音识别片上系统具有很好的参考价值。

参考文献

- 易克出,田斌.语音信号处理.北京:国防工业出版社, 2000.16-22,331-335.
- 杜普选,马庆龙.实时 DSP 技术及浮点处理器的应用.第 2 版.北京:清华大学出版社,北京交通大学出版社,2007: 86-90.
- 陈彩华,龙卫兵,刘彬.基于 ARM-Linux 的家用网络平台设计与实现.计算机测量与控制,2010,(9):2176-2177,2193.
- 韩纪庆,张磊,郑铁然.语音信号处理.北京:清华大学出版社, 2004.133-135.