

语音识别技术研究\*

刘德平 李欣生 刘红霞 刘武发

(郑州工业大学机电一体化研究所)

**摘 要** 介绍一种实时语音识别系统,对语音识别系统的硬件组成及软件技术进行了论述,所开发的系统具有较高的识别精度,有较为广阔的应用前景。

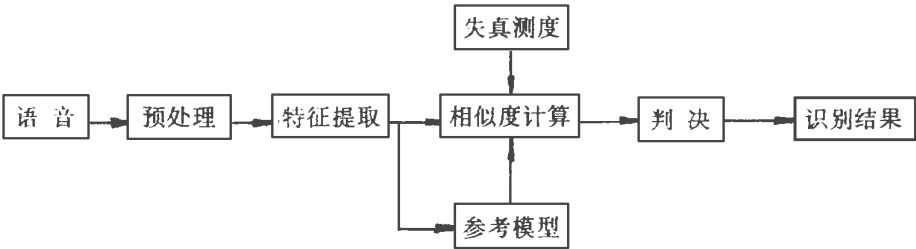
**关键词** 语音数据处理;语音基频识别;语言处理

**中图分类号** TP387

语音识别的根本目的是让机器(一般指某种计算机系统)听懂人的语音,从而根据人的意图去执行某种操作。由于语音是人类交流信息最自然和方便的手段,因而也是最理想的人机通信手段,所以语音识别的研究受到人们的极大重视。从计算机学科角度来看,语音识别可视为智能计算机的智能接口技术;从信息处理学科来看,语音识别属于语音信号处理的一个重要分支;从自动控制学科来看,语音识别属于模式识别的一个重要部分。语音识别技术有着重要的实用价值,随着计算机技术、微电子技术、信号处理、模式识别等技术的飞速发展,语音识别及理解技术向人们展示了广阔的应用前景。

1 语音识别系统概述

语音识别按其讲话人可以分为特定人和非特定人语音识别;按发音方式可以分为孤立词、连续词和连续语音识别。语音识别系统一般由以下几部分构成:



- 1.1 预处理 包括反混叠带通滤波、A/D 变换、语音端点检测等。
- 1.2 特征提取 特征提取是对输入的语音信号运用某些语音分析方法得到一个矢量序列，

\* 河南省自然科学基金资助项目(954060600)  
收稿日期:1997—09—05  
第一作者 男 1966 年 10 月生 硕士学位 讲师

用这个矢量序列去代表原始语音信号所携带的信息。常用的参数有:短时能量、短时过零率、频谱参数、共振峰参数、LPC 系数、倒谱参数等。

1.3 失真测度 失真测度是衡量相似性的基本准则,选择什么样的失真测度对识别系统的性能好坏有重要影响。常用的失真测度有:欧式距离及其多种变形、加权的倒谱距离等。

1.4 参考模型 参考模型是与词表中的词相对应的统计模型,对于模板匹配来说它是一个词对应一个或多个平均模板。

1.5 相似度计算 也称匹配计算,常用的方法有:矢量量化、动态时间规整(DTW)、隐马尔可夫模型(HMM)、神经网络等方法。

1.6 判决 指选用最佳准则及专家系统知识(构词规则、语法规则等),对各层次上的识别侯选者进行最后判决,从而得出最好的识别结果。

## 2 语音识别系统的硬件组成

汉语全音节实时识别系统要求实时实现汉语语音识别,由于一般计算机速度慢,故把数字信号处理器 TMS320C30 的高速运算能力和微机结合起来建立语音识别系统。高速语音数字信号处理板结构如下页图 1 所示。

2.1 TMS320C30 高速信号处理板充分考虑了语音信号处理的特点,又兼顾了其它数字信号处理场合的应用。

2.2 处理板设计成 IBM~PC 的一块插件板,适合 ISA 总线,它与微机构成主从系统,微机为主机,处理板为从机构成语音识别的硬件环境。

2.3 处理板以 TMS320C30 芯片为主处理器,具有 33.3MFLOPS 和 16.7MIPS 的处理能力,该芯片是美国 Texas instruments(TI)公司的浮点数据信号处理器,它是全 32 位芯片,其指令周期为 60ns,特别适合于数字信号处理及其它数值计算密集型应用,能够满足语音识别系统的实时性要求。

2.4 处理板上配有 512K 字节 0 等待高速存储器与主机 IBMPC486 构成 128 \* 32 位双寻址存储器

2.5 主机与(TMS320C30)之间配置了 8k 字节的双口 RAM 和相互中断电路,这样就大大增加了主/从机之间的通讯能力。

2.6 该板还配置了一片可编程音频信号接口芯片 TLC32044,它带有防混输入滤波器、14 位 A/D、D/A 转换器和一个重构滤波器,其采样速率为 7.2KHz——19.2KHz 可调。

2.7 处理板上有语音信号输入前置放大器和输入功效,前者可将话筒信号无失真的放大到  $\pm 3V$  内,功放可直接驱动 2W 的喇叭,这大大方便了语音信号的处理。

2.8 板上译码逻辑由一片 GAL16V8 和 4 片 GAL20V8 组成,这减少了处理机的面积同时也增加了可靠性和灵活性。

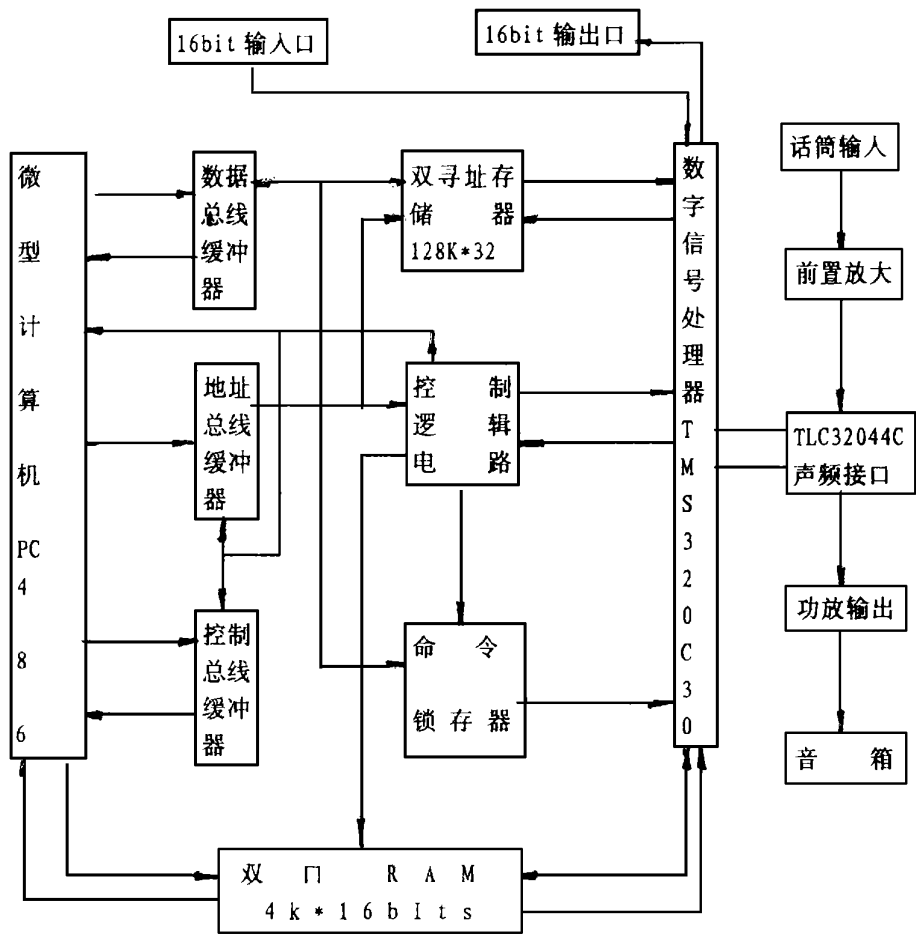


图 1 数字信号处理板结构

3 软件算法概述

我们所开发的系统,其识别过程分成四个部分:预处理,特征提取,匹配识别,决策输出,过程如下页图 2 所示.

预处理包括音频信号的预放大,低通滤波,模/数转换等.语音信号经话筒进入,由前置放大器放大为-3~+3 的电信号,经过 4.5KHz 的低通抗混叠滤波,做采样率为 10KHz、精度为 14 位的 A/D 转换,然后通过串行口进入 TMS320C30 处理器.端点检测中利用的是过零率和能量两参数来进行判决,声韵分割部分利用了语音的帧能量和帧间距离等一系列参数,其性能有明显改善。

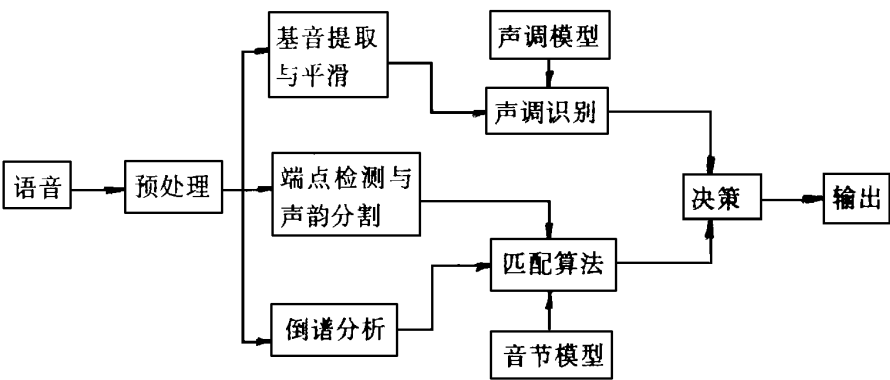


图 2 软件算法识别过程

综合性能和复杂度等多方考虑,选用中心削波自相关函数法提取基音周期,利用中心削波进行谱平整,抑制信号其它方面带来的干扰,突出了周期性。采用中值平滑对提取的基音数据进行处理,利用基音的相对稳定性来消除零点及奇异点。

模/数转换后的语音数据经过预加重进行高频提升后分帧,一帧 256 个点,帧间位移 100 点。对每帧加 Hamming 窗后,求出其自相关系数,然后用 Levinson-Durbin 算法求出 12 阶 LPC 参数,进而求出倒谱参数。

本系统中,语音信号预处理,特征参数提取,模板匹配等一系列工作全在 TMS320C30 高速数字信号处理板上完成。其中预处理、特征参数提取是实时完成的,PC 机仅起一个系统管理作用,系统起动时对处理板加载程序,识别后显示识别结果。

随着计算机技术、模式识别技术、神经网络等相关技术的发展,语音识别技术也必将走向成熟、进入实际应用。更自然、更高难度的语音识别系统上相信会在移动计算、工业、军事、医疗、出版及文字处理等领域广泛应用,其前景异常广阔。

参考文献

1 樊来耀. TMS320C25. C30 数字信号处理器;西安电子科技大学出版社  
2 刘丰. 语音处理新技术研究.[学位论文];西安电子科技大学出版社  
3 樊来耀. 32 位数字信号处理器软件硬件;西安电子科技大学出版社

Research on Rpeech Recognition Technology

Liu Deping Li Xinsheng Liu Hongxia Liu Wufa  
(Zhengzhou University of Technology)

**Abstract** A real-time speech recognition system is introduced, the composition of hardware and software technology is discussed. The system has higher recognition accuracy and wider application prospect.

**Keywords** voice data processing; language fundamental frequency recognition; speech processing