

语音信息的身份和指令识别方法应用^①

姬晓飞

(沈阳航空航天大学 自动化学院, 沈阳 110136)

摘要: 针对智能机器人系统对语音交互功能的实际需求, 在 DM-B200 开放式机器人平台上, 设计并实现了基于语音信息的说话人身份识别功能和语音指令识别功能. 为了准确识别特定说话人的身份, 采用了矢量量化的识别方法, 而为了提高机器人语音控制的实时性, 语音指令识别采用了动态时间规整的方法. 应用结果表明, 所设计的语音识别系统达到 96% 以上的综合识别率, 具有实际应用价值.

关键词: 语音指令识别; 矢量量化算法; 说话人身份识别; 动态时间规整

Application of Voice-Based Speaker Identification System and Command Recognition Algorithm

Ji Xiao-Fei

(School of Automation, Shenyang Aerospace University, Shenyang 110136, China)

Abstract: For satisfying the voice-based communication requirements of intelligent robotics, based on the DM-B200 robot, the speaker identification system and command recognition system are designed. In order to improve the correct recognition rate of speaker identification, the vector quantization method is used. And the dynamic time warping algorithm is used to speed the command recognition process. The application results showed the designed systems obtained more than 96% correct recognition rate and are easy to be extended to the real production.

Key words: voice command recognition; vector quantization algorithm; speaker identification; dynamic time warping

随着智能机器人的广泛研究, 与机器人进行语音交流, 让机器人明白人类所说的语言, 既是人类长期以来梦寐以求的事情, 也是未来机器人更好的服务人类的根本需求. 与此同时, 语音识别技术也得到了迅猛发展, 支持语音识别的各种产品纷纷面世. 这也为智能机器人的语音交流功能提供了支持. 虽然具有语音交互能力的机器人的应用范围非常广泛, 但基本功能主要包括对说话人的身份识别功能和特定说话人的语音指令识别功能. 下面就这两种基本功能的发展现状做一个简单回顾.

语音识别的研究始于五十年代, 1952 年贝尔实验室的 Davis、Biddulph 和 Balashek 建立了一个与说话人相关的孤立英文数字语音识别系统. 该系统能够在说话人和麦克风距离保持不变的情况下达到 98% 的准

确率^[1,2]. 1959 年, 美国 Lincoln 实验室的 Rorgie 和 Forgie 首次应用数字计算机来识别以元音和摩擦音开头的孤立字, 从此计算机语音识别时代开始^[2]. 近年来较为常用的、综合效果较好的语音特征形式主要有线性预测倒谱系数 (Linear Prediction Cepstrum Coefficient: LPCC), 感知线性预测系数 (Preceptual Linear Predictive Cepstrum Coefficient), Mel 频率倒谱系数 (Mel Frequency Cepstrum Coefficient: MFCC) 等. 在语音识别方法研究方面, 于 60 年代进入初期阶段, 提出了动态规划 (Dynamic Programming: DP) 和线性预测分析 (Linear Predictive: LP) 技术. 到七十年代末八十年代初 LP 技术和动态规整技术 (Dynamic Time Warping: DTW) 基本成熟, 后来矢量量化 (Vector Quantization: VQ), 隐马尔可夫模型 (Hidden Markov Model: HMM),

^① 基金项目: 国家自然科学基金青年基金 (61103123)

收稿时间: 2012-09-14; 收到修改稿时间: 2012-10-28

等理论不断被提出并投入应用、加以改进,使说话人识别技术得到进一步的发展、完善^[3-5]。人工神经网络(Artificial Neural Network: ANN)近年来也被广泛的研究和应用^[4]。

语音识别用于智能机器人方面已经得到了一定的研究进展,文献[6]利用语音指令实现了机器人的远程控制,文献[7]结合语音和人脸信息实现了陪伴机器人的身份认证。

本文将依托 DM-B200 研究型全自动移动机器人平台,设计并实现说话人的身份识别功能,以及针对特定人的指令识别功能。在这两种基本功能基础上,将该移动机器人平台开发成为用于看护孤独老人的服务机器人。

1 语音识别系统的软硬件环境

1.1 语音识别系统实现的硬件环境

硬件环境主要分为声音采集模块和执行机构两大部分。所依托的机器人平台如图 1 所示。



图 1 DM-B200 研究型全自动移动机器人平台

声音采集模块包括一个精度较高麦克风和一台电脑,用于进行语音信号的采集,形成 WAV 格式声音文件直接应用于所开发的说话人身份识别系统以及指令识别系统中。

执行机构指 DM-B200 研究型全自动移动机器人,使用 DCT0020 为其控制驱动器,配合光电编码器反馈的高精度直流伺服电机。系统与计算机通讯口采用 USB 转 RS232。

1.2 语音识别系统实现的软件环境

(1)应用 Cyber Link Wave Editor 进行语音的录入,建立训练人模板库。系统需要录入指定机主语音 30 次

以形成指定说话人模板。

(2)应用 Matlab 程序进行身份识别系统的程序编译、测试。

(3)应用 C++进行机器人运动的编辑及身份识别系统的接入。

2 说话人身份识别系统设计

说话人身份识别系统的设计流程如图 2 所示。根据流程图,语音信息从获取到识别出结果,主要经过语音预处理,语音特征提取和识别三个过程。

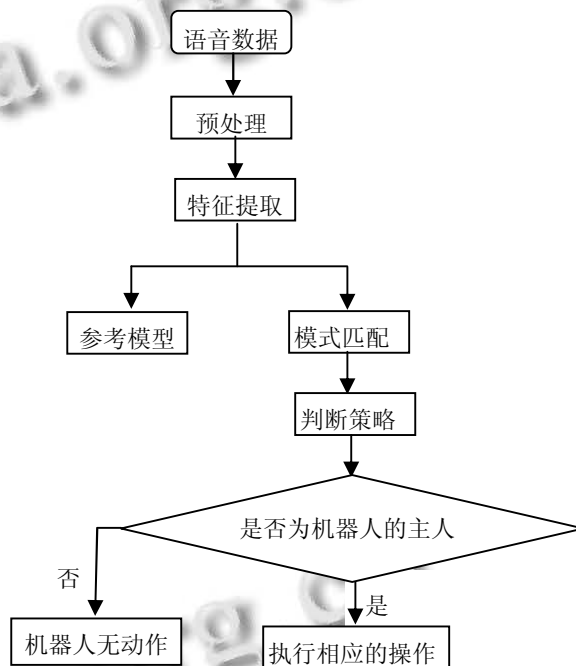


图 2 说话人身份识别系统设计流程图

2.1 语音信息的预处理

预处理过程包括预加重、分帧加窗、端点检测几个部分。预加重的作用是提升语音的高频部分,使语音信号的频谱变得平坦,以便于进行频谱分析或声道参数分析系统通过测试选定预加重系数为 0.9375。语音信号为非平稳时变信号,但可以看作短时平稳,故要将语音信号分为较短的时间间隔的语音帧。分帧处理会截断频谱和使频谱叠加,故需要对语音帧进行加窗处理,为有效减少旁瓣,通过实验选择加哈明窗。采用的端点检测方法为双门限法,即同时检测语音的短时过零率和短时能量,以截去语音段的无语音段(噪声或无声段),增加系统计算速度。用一帧信号中波形穿越零电平的次数来描述这种变化的剧烈程度,称

为过零率. 短时平均过零率可描述信号的时变频谱. 语音信号的短时能量反映了语音信号幅度随时间的变化, 短时能量的主要意义在于, 可以用来区分浊音段与清音段, 声母与韵母, 无声与有声.

2.2 语音特征提取

语音特征提取选用 Mel 倒谱系数及其一阶差分系数复合函数. Mel 倒谱系数是说话人识别中最常用的语音特征参量之一, Mel 倒谱系数较好的模拟了人耳的听觉系统, 具有识别性能高和抗噪能力强等特点. 其提出是基于以下的事实: 人耳对单个音调的感知强度近似的正比于该音调频率的对数. Mel 频率正是表述了这种语音频率与“感知频率”间对应的关系.

在 Mel 频域内, 人对音调的感知度为线性关系. 其中 Mel 倒谱系数较好的描述了语音的静态特性, 而其一阶差分系数能够较好的重现语音参数的动态特性. 通过对比不同阶数特征提取的效果可知, 当所选 MFCC 参数阶数过小时, 说话人之间因方言、语调以及声道构成的差异所带来的差别过于明显, 噪声所造成的干扰也十分严重, 使系统的准确性变差. 而当阶数过高时, MFCC 参数过于复杂, 这大大增加了系统的计算量, 过大的计算量几乎不能提升系统的准确性, 却造成了系统的实时性变低, 所以选定的 MFCC 阶数为 20 阶. 系统在这一阶数时较为稳定、准确、实时性较高.

2.3 基于 VQ 的语音识别方法

矢量量化(VQ)方法是 70 年代后期发展起来的一种数据压缩和编码技术, 其理论基础是信息论中的率-畸变理论. 对于一定的量化速率 R (以量化信号和原信号之间的误差均方值和原信号均方值之比来衡量) 是一定的. 矢量量化与标量量化相比, 其畸变值能够更加接近于率-畸变理论所给出的畸变下限值, 即矢量量化能够在 R 一定的情况下给出更小畸变, 或者在相同畸变下给出更高的量化效率(即 R 值更低). 矢量量化的研究目的在于针对特定的信息源和矢量维数, 找到一种最有效的矢量量化器, 使其能够在 R 一定的情况下给出最低畸变. VQ 算法在限定说话者使用较小词汇时, 效果相当好, 应用 VQ 算法进行处理既能达到识别率要求, 也能减少计算量. 其中最佳码本设计方法选用 LBG 算法, 其基本思想是每次叠代时都使用最小距离准则对训练样本进行重新分类, 使每次叠代后的总的量化失真减小.

相对于 VQ 方法, 动态时间规整(DTW)由于其的时间相关性只能分辨指令, 在分辨说话人身份上效果很不好. 而一个基于隐马尔可夫模型(HMM)的语音识别系统在说话者使用较小词汇时, 几乎没有带来性能的提高却增加了计算的复杂度, 因而使计算时间增长^[4,5]. 因此本文将对话人身份识别功能使用 VQ 方法进行识别.

3 语音指令识别系统设计

根据语音指令识别的特点, 所设计识别系统结构如图 3 所示, 主要包括声音采集、指令识别和执行机构三个部分以及一个参考特征库.

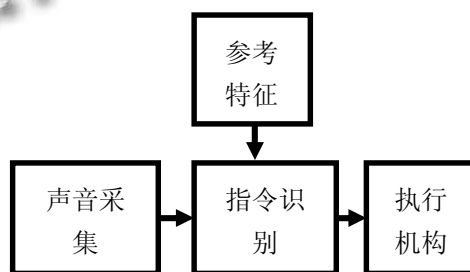


图 3 语音指令识别系统设计流程图

声音采集模块负责接收外界的声音, 这一部分功能与身份识别功能共用相同的硬件设备. 同时也需要与身份识别类似的语音预处理过程.

指令识别模块负责提取声音的有效区段、提取特征并通过一定的算法将特征与参考特征库进行比较, 得到识别结果后将其发送给执行机构. 所采用的语音特征是线性预测倒谱系数(LPCC). 采用的识别方法为动态时间规整(DTW)^[5].

执行机构从指令识别模块接收识别结果, 将识别结果转换为硬件可以识别的控制流, 完成相应动作. 具体设备及设置将在第 4 节给出.

4 实验验证

在鸥鹏科技的 DM-B200 研究型全自主移动机器人平台上, 将设计的说话人身份识别和语音指令识别系统进行技术移植. 利用 PC 机的运算速度和 MATLAB 软件平台强大的语音处理函数, 将声音采集、预处理等环节直接使用 PC 机实现, 语音识别程序则通过 Matlab 中的 MCC 命令将 m 文件以动态链接库(DLL)文件的方式进行编译. 之后机器人平台则通过

VC++引用生成的头文件完成对识别功能的调用. 之间的通讯可以直接调用机器人平台配套的动态链接库实现(DCTMotionCtrl.dll).

机器人平台上设计的操作界面如图 5 所示.



图 5 语音识别系统的操作界面

通讯功能通过调用 DCT*相关函数完成通讯口 (COM)设定、与下位机的通断、下位机驱动器的准备检测和结束检测以及主程序的退出功能. 语音识别部分则调用 MAT*相关函数对麦克风或文件中提供的声音源进行识别分析判断, 并将判断结果分别以文字和按钮的方式在下方显示. 按钮部分则将当前识别状态变成灰色无效, 其余 4 个命令按钮有效, 用于进行手动控制.

另外, Matlab 中将 m 文件编译为 DLL 使用 MCC 命令, 其方法如下:

以功能函数 getLpcc.m 为例:

```
mcc -W cpplib:libgetLpcc -T
```

```
lib getLpcc.m -c
```

该命令实现将 getLpcc.m 文件以 C++的 lib 形式生成动态链接库(DLL)文件, 文件名为 libgetLpcc.dll.

在以上的测试平台上, 在相同的背景噪声下, 对 30 个说话人的 5 秒钟语音进行录制, 预处理后形成说话人身份识别的训练数据库, 之后进行了测试, 测试结果如表 1 所示, 正确识别率达到 96.7%.

表 1 实际测试说话人身份识别率结果

说话人身份 识别正确	说话人身份 识别错误	说话人身份 无法识别	正确识 别率
29	0	1	96.7%

同时针对设计的语音指令识别功能, 录制了 50 个

人的 6 种语音指令(停止, 前进, 后退, 左转, 右转, 摇摆), 形成训练数据库, 之后也进行了测试, 测试结果如表 2 所示, 在模板库内的语音指令, 正确识别率达 96%, 库外的语音指令, 正确识别率为 88%.

表 2 实际测试语音指令识别率结果

	语音指令 识别正确	语音指令 识别错误	语音指令 无法识别	正确识 别率
模板库内 说话人	48	0	3	96%
模板库外 说话人	44	2	4	88%

通过以上的实际应用结果可以看出, 所设计方法可以得到较高的正确识别率, 这也为该系统的实际应用和推广提供了依据.

5 结论

在 DM-B200 机器人上设计并实现了说话人身份识别功能和语音指令识别功能, 实际的测试结果表明, 所设计的两个语音识别系统都能够达到 96% 以上的正确识别率. 这两项功能的成功实现为下一步将该机器人改造为服务机器人提供了数据支持, 同时也为不同语音识别方法提供了一个实际的测试平台.

参考文献

- 张震,王化清.语音信号特征提取中 Mel 倒谱系数 MFCC 的改进算法.计算机工程与应用,2008,44(22):55-54,58.
- 刘玄和,宋庭新.语音识别与控制应用技术.北京:科学出版社,2008.
- 王秀萍.基于 LBG 和 DTW 的模板训练算法.计算机工程与应用,2005,26:85-87.
- 李继连.动态时间弯折模式匹配技术(DTW)的算法改进研究.科技信息,2006,10:83-84.
- Kim C, Seo KD. Robust DTW-based Recognition Algorithm for Hand-held Consumer Devices. IEEE Transactions on Cosumer Electronics, 2005,51(2):699-709.
- 李书贞,施玉霞.基于语音指令的远程控制机器人系统设计与实现.微型电脑应用,2008,24(11):1-3.
- Cao JT, Kubota N. The visual-audio integrated recognition method for user authentication system of pattern robots. International Journal of Human Robot, 2011,8:691-698.